Journal of proteome research

Article

pubs.acs.org/jpr

Terms of Use

Identification of Trans-Golgi Network Proteins in *Arabidopsis thaliana* Root Tissue

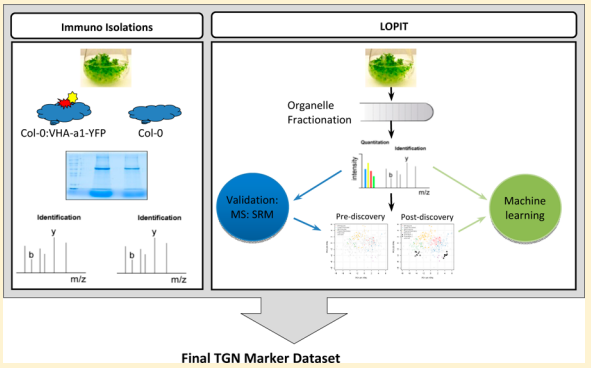
Arnoud J. Groen,[†] Gloria Sancho-Andrés,[‡] Lisa M. Breckels,[†] Laurent Gatto,[†] Fernando Aniento,[‡] and Kathryn S. Lilley*,[†]

[†]Cambridge Centre for Proteomics, Cambridge Systems Biology Centre, Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom

[‡]Departamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Valencia, Valencia, Spain

Supporting Information

ABSTRACT: Knowledge of protein subcellular localization assists in the elucidation of protein function and understanding of different biological mechanisms that occur at discrete subcellular niches. Organelle-centric proteomics enables localization of thousands of proteins simultaneously. Although such techniques have successfully allowed organelle protein catalogues to be achieved, they rely on the purification or significant enrichment of the organelle of interest, which is not achievable for many organelles. Incomplete separation of organelles leads to false discoveries, with erroneous assignments. Proteomics methods that measure the distribution patterns of specific organelle markers along density gradients are able to assign proteins of unknown localization based on comigration with known organelle markers, without the need for organelle purification. These methods are greatly enhanced when coupled to sophisticated computational tools. Here we apply and compare multiple approaches to establish a high-confidence data set of *Arabidopsis* root tissue trans-Golgi network (TGN) proteins. The method employed involves immunoisolations of the TGN, coupled to probability-based organelle proteomics techniques. Specifically, the technique known as LOPIT (localization of organelle protein by isotope tagging), couples density centrifugation with quantitative mass-spectometry-based proteomics using isobaric labeling and targeted methods with semisupervised machine learning methods. We demonstrate that while the immunoisolation method gives rise to a significant data set, the approach is unable to distinguish cargo proteins and persistent contaminants from full-time residents of the TGN. The LOPIT approach, however, returns information about many subcellular niches simultaneously and the steady-state location of proteins. Importantly, therefore, it is able to dissect proteins present in more than one organelle and cargo proteins en route to other cellular destinations from proteins whose steady-state location favors the TGN. Using this approach, we present a robust list of *Arabidopsis* TGN proteins.



KEYWORDS: *trans-Golgi network, LOPIT, Arabidopsis thaliana, immunoisolation, phenoDisco, machine learning, organelle proteomics*

INTRODUCTION

Membrane trafficking is vital for all eukaryotes. The proteins that are involved in the processes underlying membrane trafficking have therefore received much attention. In the field of organelle proteomics, one of the main goals is to identify the protein composition of the subcellular compartments to gain a better understanding of protein trafficking pathways. Organelle protein identification is complicated by the fact that due to the dynamic nature of membrane trafficking, it is generally hard to distinguish cargo proteins that are en route to their final cellular destination from full-time endomembrane residents who carry out their function at a given location. This makes obtaining a list of reliable marker proteins for endomembrane compartments somewhat challenging.

In both plant and animal cells, biosynthetic protein trafficking is initiated at the endoplasmic reticulum (ER), from which newly synthesized membrane or soluble cargo proteins are transported through the Golgi apparatus to the trans-Golgi network (TGN) for sorting. Proteins destined to be secreted are further trafficked to the plasma membrane (PM), whereas proteins that traffic to the vacuole in the case of plant cells, pass through the multi vesicular body (MVB)/prevacuolar comparment (PVC)/late endosome (LE), as reviewed in ref 1. Furthermore, PM proteins can undergo endocytosis. In plants, it has been shown that for many proteins including PIN proteins and the endocytic tracer FM4−64, their endocytosis is

Received: August 18, 2013
Published: December 17, 2013

clathrin-dependent.[2−4] A similar clathrin-dependent process also exists in animal cells. All endocytosed plant PM proteins are first delivered to the TGN, from where they are recycled to the PM via recycling endosomes or else they are delivered via PVCs/MVBs/LEs to the lytic vacuole for degradation.[1] Many plant PM proteins undergo constitutive endocytosis and recycling, including the auxin-efflux carriers PIN1 and PIN2,[5] the brassinosteroid receptor BRI1,[6] and the boron transporter BOR1.[7] In contrast, other PM receptors are only internalized upon ligand binding, such as the pattern recognition receptor FLS2.[8]

The plant TGN is thus a highly dynamic organelle that constitutes a major sorting center for both the biosynthetic and endocytic pathways.[9] It functions as an early endosome by receiving endocytosed cargo from the PM[10,11] and also as a sorting station for biosynthetic cargo coming from the Golgi and destined either to the PM/cell wall/cell plate (glycoproteins, cell wall polysaccharides)[9,12] or to the vacuole.[1,13] Given that many proteins transit through the TGN, it is challenging to distinguish cargo proteins from TGN residents that have a steady-state position in the TGN and carry out their functions within this organelle. Knowledge of which proteins fall into this latter category aids our understanding of the sorting events taking place within the TGN.

Although there is much overlap between plant and animal cells in the molecular machinery involved in membrane trafficking, direct comparison between animal proteins and their plant homologues is foolhardy because, in many cases, they are present in different compartments.[1] The GTPase Rab5, for example, a classical marker of early endosomes in animal cells,[14,15] has three homologues in *Arabidopsis*. Although RHA1/RAB-F2A and ARA7/RAB-F2b colocalize at the PVCs/MVBs,[16] plant-specific ARA6/RAN-F1 locates in different endosome populations, although its localization overlaps to some extent.[17] The localization of animal TGN proteins thus cannot simply be extrapolated to their plants homologues.

Many approaches ranging from low- to high-throughput have been undertaken to determine the subcellular location of membrane proteins. Although 2D gel-based proteomics analysis has been utilized in the past to determine catalogues of organelle proteins, for example, the nuclear proteome in mouse liver[18] and *Arabidopsis thaliana* mitochondria,[19] they are not compatible with integral membrane proteins because of solubility issues during isoelectric focusing.[20] Nongel approaches circumvent the problem of a bias toward soluble proteins in organelle proteomics studies.

To date, many such studies have relied on organelle purification. Methods to achieve purification include free flow electrophoresis (FFE), where organelles are separated based on surface charge and immunoisolation of specific vesicle populations expressing a surface marker for which antibody reagents are available. FFE has been used effectively to produce enriched Golgi fractions from *Arabidopsis*.[21] Immunoisolation of membrane fractions has been successfully employed, for example, in a recent study where a SYP61 compartment was immunoisolated from *Arabidopsis* by targeting the TGN marker protein SYP61. In this study, 147 proteins were found to be associated with this compartment.[22] Such methods, however, are not able to distinguish true residents from trafficking cargo proteins and also may carry a high level of false discoveries without the use of carefully crafted controls. Moreover, performance of a high number of biological replicates may not distinguish cargo and contaminants from true residents in an immunoisolation as proteins in both categories are likely to persist through multiple experiments. Furthermore, any method that results in the analysis of a single compartment leads to a binary, present or not present, answer. Such approaches are not well-suited, therefore, to demonstrate subtle changes in protein localization that occur via trafficking or due to a change in protein localization upon stimuli, which is becoming a necessity to chart system-wide dynamic changes in subcellular protein localization in response to perturbation[23]

Gradient-based quantitative proteomics techniques have been developed, including protein correlation profiling (PCP[24]) and localizaton of organelle proteins by isotope tagging (LOPIT[25]), to be able to distinguish between true residents, shared proteins, and trafficking proteins. Both PCP and LOPIT are based on the principle developed by Christian de Duve, whereupon separation by continuous equilibrium density centrifugation an organelle will have a specific distribution pattern along that gradient and proteins of unknown localization can be assigned to organelles by comparing their distribution patterns with those of proteins of known localization.[26] LOPIT has been successfully applied to *Arabidopsis thaliana* callus,[25,27,28] the DT40 lymphocyte cell lines[29] and *Drosophila melanogaster*.[30] Recently, LOPIT has been used to characterize novel enzymes involved in the biosynthesis of complex polysaccharides, glycoproteins, and glycolipids in the Golgi apparatus.[31]

The matching of proteins with unknown localization to specific organelle distribution patterns in the LOPIT workflow initially used multivariate methods such as partial least-squares discriminant analysis (PLS-DA)[25] and machine learning methods such as support vector machine.[32] The success of using such methods to assign proteins to organelles is, however, highly dependent on the presence of well-described organelle clusters within the data, the degree of separation achieved between organelle clusters across the gradient employed, and the number of reliable known marker proteins within the data set that can be used in classifier creation.

In the present study we apply (1) LOPIT, (2) computational modeling, and (3) selected reaction monitoring (SRM) analysis to characterize the TGN membrane proteome in *Arabidopsis* roots. By employing this combinatorial approach, we find 5 membrane proteins already assigned to the TGN in previous studies and importantly, 25 novel TGN membrane proteins. Using the protein distribution profiles, a semisupervised novelty detection algorithm[33] is applied prior to any protein classification to first identify a distinct TGN cluster in the data without giving the algorithm any a priori information of the existence of the TGN in the data sets generated. This preliminary computational analysis identifies a number of TGN candidates that then serve as input for the main supervised machine learning classification, in which we employ the *K*-nearest neighbor algorithm to classify TGN proteins. We also demonstrate the application of SRM, a targeted proteomics approach, to determine the distribution profiles of TGN marker proteins. The employment of SRM analysis enables validation of the LOPIT quantitation of TGN markers using isobaric tagging to exemplify that the assignments are based on reliable quantitation data. Finally, we compare our data to TGN proteome catalogues by immunoisolation and demonstrate that the approach presented here is capable of distinguishing TGN residents from cargo proteins whose steady-state location in cells is not in this organelle. This approach will aid visualization of the dynamic trafficking processes in cells and promises to

provide a method to interrogate dynamic events in cellular trafficking upon perturbation.

MATERIAL AND METHODS

Plant Material

Arabidopsis thaliana Col-0 wild-type plants and transgenic lines expressing VHA:a1-GFP (Dettmer et al. 2006) were grown in 500 mL flasks containing 100 mL of Murashige and Skoog (MS) liquid medium (2.2 g/L), and 10 g/L sucrose, 0.15 g/L MES, adjusted to pH 5.7 with KOH, for 10 days under 16 h light/8 h dark at 115 rpm and 25 °C. Roots were separated from the green parts and ground with a mortar and pestle with ice-cold homogenization buffer (HB, 0.17 M 8% sucrose; 1 mM EDTA; 20 mM HEPES pH 7.5; 20 mM KCl; 1 mM DTT; 0.2% protease inhibitor cocktail sigma) (5 mL/g dry root tissue). Debris was pelleted by centrifugation at 800*g* for 5 min, and the supernatant was further centrifuged at 1500*g* for 5 min to obtain a postnuclear supernatant (PNS). 4.5 mL of PNS was loaded on top of a 1 mL 42% sucrose cushion and centrifuged at 150 000*g* for 1 h at 4 °C in an MLS-50 rotor (Beckman). Root membranes from Col-0 wild-type plants or transgenic lines expressing VHA:a1-GFP were collected from the 42%/8% sucrose interface and used for immunoisolation or for gradient fractionation.

Immunoisolation

A monoclonal GFP antibody (MA1, Pierce) was coupled to sheep antimouse magnetic beads (Dynabeads, Invitrogen) by incubation in a rotary shaker (8 rpm) for 16 h at 4 °C in PBS containing 2 mg/mL BSA. Beads were washed three times with PBS-BSA and two times with immunoisolation buffer (IB: PBS pH 7.4; 2 mM EDTA; 5% BSA). For immunoisolation, 100 μL beads (4×10^8 beads/mL) were incubated overnight at 4 °C on a rotary shaker with 1 to 2 mg root membranes (Col-0 or VHA:a1-GFP) in 1 mL of IB. Beads were washed two times with IB and three times with HB and extracted with lysis buffer (0.05 M Tris-HCl pH 7.5; 0.5% Triton X-100; 1 mM EDTA; 1 mM PMSF; 0.15 M NaCl) for Western blot analysis and proteomic analysis. For proteomics analysis, the immunoisolates were loaded on a SDS-PAGE gel and stained with Coomassie. Gel lanes were sliced into seven fractions for subsequent in-gel digestion. The Coomassie stain was removed from the gel fraction by washing three times with 80 μL of 50% acetonitrile (ACN)/50 mM ammonium hydrogen carbonate, reduced with 10 mM DTT for 1 h, alkylated with 5 mM iodoacetamide for 45 min, and again washed three times with 50% ACN/50 mM ammonium hydrogen carbonate. A final wash with 100% ACN was carried out before digestion overnight at 37 °C with 60 μL of 0.005 μg/μL of trypsin.

Gradient Fractionation and Protein Extraction

One mL root membranes (~0.5 mg), obtained as previously described, were loaded on top of a 4.5 mL continuous gradient made by using 20 and 47% sucrose solutions, and centrifuged for 3 h at 150 000*g* in an MLS-50 rotor (Beckman). After centrifugation, 0.5 mL fractions were collected from the bottom of the gradient, sucrose concentration was measured, and the fractions were analyzed by Western blotting and used for proteomic analysis. The intensities of the bands obtained were quantified using the Quantity One software (Bio-Rad Laboratories).

Antibodies used for Western blot analysis were the following: BiP,[34] Sec21,[35] SYP21 and SYP51 (both kind gifts from Prof. David Robinson), ARF1,[35] VSR1,[36] H^+PPase, Sec7-GNOM,[37] RabA4,[38] and PIN1[39].

Carbonate Wash

100 μL of ice-cold 162.5 mM $NaCO_3$ was added to the 0.5 mL gradient fractions and left incubating for 30 min on ice. Fractions were centrifuged at 100 000*g* for 15 min in a Beckman benchtop ultracentrifuge. Supernatant was removed and kept for further analysis. Pellets were washed twice with ice cold HPLC grade water and centrifuged at 100 000*g* for 10 min in both cases. Pellets were then resolubilized in 40 μL of 8 M urea and 0.1% SDS in 50 mM TEAB (pH 8) and sonicated for 3 × 30 s for full recovery. Protein concentration was estimated by BCA assay (Invitrogen).

Digestion and iTRAQ Labeling

For LOPIT 1, fractions 2, 4, 6, and 8−10 (4 × 100 μg) were taken for iTRAQ 4-plex labeling. For LOPIT 2, two iTRAQ labelings were performed. From the 15 fractions, 2, 4, 6, and 8−11 (LOPIT 2A, 4 × 68 μg) and 3, 4, 5, and 7 (LOPIT 2B, 4 × 63 μg) were iTRAQ 4-plex labeled (ABSciex). After protein estimation, each fraction was taken and reduced using TCEP (ABSciex) and alkylated using MMTS (ABSciex). The sample was subsequently diluted 10 times with 50 mM TEAB to bring down the concentration of urea to 0.8 M. Trypsinisation was performed using 2.5 μg of trypsin (TRESZQ, Worthington, Lorne Laboratories Limited) per 100 μg and incubated for 1 h at RT. Another batch of 2.5 μg trypsin was added and incubated overnight. Samples were freeze-dried and stored at −20 °C or labeled immediately. For the labeling, fractions were dissolved in 25 μL of 1 M TEAB pH 8.0 and 75 μL ethanol, and either the 114, 115, 116, or 117 iTRAQ label (ABSciex) was added. The peptide/iTRAQ mix was incubated for 1 h at RT, after which 100 μL of H_2O was added to quench the reaction. After 15 min, the samples were pooled and lyophilized to dryness.

RP Chromatography and MS/MS

Two-dimensional peptide separation was achieved by a combination of high and low pH reverse-phase chromatography. A UPLC reverse-phase column (Waters, BEH C18, 2.1 × 150 mm, 1.7 μm) was utilized during the first dimension of separation. 20 mM NH_4-formate in HPLC water (pH10) was used as hydrophilic mobile phase and ammonium formate in HPLC water/80% ACN was used as the organic mobile phase. Eighteen 2 min fractions were collected per gradient (75 min gradient; 0−10 min: 0% buffer B, 10−60 min: 0→35% buffer B, 60−67 min: 100% buffer B, 67−75 min: 0% buffer B) and were freeze-dried overnight. The freeze-dried pellets were dissolved in 25 μL of HPLC water/3% ACN/0.1% formic acid prior MS analysis. From the total of 25 μL, 1 μL was taken for nano-LC ESI−MS/MS analysis. The analysis was performed on an Orbitrap Velos (Thermo) coupled to a nanoAcquity LC (Waters). Samples were trapped (Waters, C18, 180 μm × 20 mm), loaded on a RP column (Waters, BEH130, C18, 75 μm × 150 mm, 1,7 μm) with a flow rate of 300 nL/min (buffer A: HPLC H_2O, 0.1% formic acid, buffer B: 100% ACN, 0.1% formic acid, 120 min gradient; 0−100 min: 3→35% buffer B, 101−106 min: 35%→85% buffer B, 107−120 min: 3% buffer B). Data were acquired in a top-10 data-dependent acquisition (DDA) in HCD collision mode with a 0.5 Da precursor ion selection window and a 30 000 resolution.

dx.doi.org/10.1021/pr4008464 | J. Proteome Res. 2014, 13, 763−776

Data Processing

For all tandem mass spectrometry experiments, msConvert[40] was used to create the .mzXML files from .raw Thermo Orbitrap files. The iSPY software (in-house software and ref 41) was used to create .mgf files that were imported in Mascot (Matrix Science, London, U.K., version 2.3.2) for peptide identification (TAIR 8 nonredundant protein database (27 ,234 sequences)). The search was run using the following settings: carbamidomethyl, iTRAQ (4plex) K, iTRAQ (4plex) N-term as fixed modifications; oxidation on methionine (M) residues and iTRAQ (4plex) Y as variable modifications; 25 ppm of peptide tolerance, 0.8 Da of MS/MS tolerance; max of 2 missed cleavages, a peptide charge of +2, +3, or +4; and selection of decoy database. Mascot .dat output files were imported in iSPY and run through percolator for improved identification.[42] In the case of LOPIT, the peptides in the iSPY .tsv output files were imported into R statistical programming environment (http://www.r-project.org) and processed using the *MSnbase* infrastructure.[43] Only unique peptides identified by spectra with posterior error probabilities smaller than 0.01 were retained. Also, peptides with a cumulative iTRAQ reporter ion intensity of less than 10 000 ions were discarded. Peptides were merged into proteins and the iTRAQ reporter ion intensities were normalized to six ratios (114/115, 114/116, 114/117, 115/116, 115/117, 116/117), and then each protein abundance was further normalized across its six ratios by sum. The iTRAQ reporter ion intensities of peptides of the same protein were averaged in an intensity-dependent manner.

For the identification of proteins in the immunoisolations, only proteins with a final protein error probability less than 0.01 were included in the final data set.

Machine Learning and Multivariate Data Analysis

The Bioconductor[44] package MSnbase (Gatto and Lilley, 2012, version 1.9) and *pRoloc* (http://bioconductor.org/packages/devel/bioc/html/pRoloc.html, version 1.1) for the R statistical programming language (R Core Team, 2013, version 3.1) were used for handling of the quantitative proteomics data and the protein-localization prediction.

The assignment of proteins to the TGN compartment was a two-step process that involved a first initial application of the *phenoDisco* novelty detection algorithm (Breckels et al., 2013) to identify and confirm the existence of a distinct TGN cluster and second a supervised machine learning classification using the *K*-nearest neighbor (*k*-NN) algorithm for final protein localization assignment.

The *phenoDisco* algorithm in the *pRoloc* package is a semisupervised novelty detection algorithm that is able to identify novel clusters of which the algorithm has no prior knowledge, which represent putative subcellular niches in quantitative organelle proteomics data. Here we applied the *phenoDisco* algorithm prior to protein localization assignment for two reasons: (1) to test the existence of a well-define TGN structure within the data and (2) to extract a training set of markers to be used to train a supervised protein localization classifier for protein localization assignment. The *phenoDisco* algorithm and its application including the specific parameters used to run this analysis are described in detail in the supporting methods in the Supporting Information.

Following identification of a phenotype cluster that represented the TGN, additional TGN markers were able to be extracted and used as input-labeled training examples for three independent *k*-NN supervised machine learning classification experiments (one for LOPIT 1, 2A, and 2B). The *k*-NN algorithm is an established instance-based learning method based on the notion that the instances within a data set generally exist in close proximity to other instances with similar properties. In *k*-NN, an instance, that is, a protein, is classified by a majority vote of its *k* neighbors; if the neighboring proteins are labeled with one of the classes in the training set, then the value of the label of an unclassified instance can be determined by observing the class of its nearest neighbors. The relative distance between instances is determined using a distance metric (here Euclidean). As previously mentioned, to classify proteins to subcellular compartments using supervised machine learning requires a set of labeled training examples, that is, protein profiles of known localization, to train a classifier. This involves optimization of classifier parameters. Proteins of unknown localization to a specific subcellular niche can then be matched to a specific subcellular location using this training data. For employment of the *k*-NN approach, the value of *k* for each data set needs to be optimized. This was done using stratified cross-validation as implemented in the *pRoloc* software. (Specific implementation is described in the supporting methods in the Supporting Information.) The optimal value of *k*, to be used in the final classification, for each individual data set was three.

Because our primary goal was to identify new TGN localized proteins the classification problem was modeled as a binary '*TGN vs other*' experiment in which one class contained solely TGN markers and the class "other" class grouped non-TGN markers from other organelles, these specifically included markers from the plasma membrane (PM), mitochondrion (MT), chloroplast (CL), vacuole (V), endoplasmic reticulum (ER), and Golgi apparatus (GA). The input TGN marker set for protein assignment was generated from curation of the *phenoDisco* output, and information in the Uniprot database and the literature and consisted of VHA-a1 (At2g28520), RabA2b (At1g07410), YIP1 (At4g30260), Ran1 (At5g44790), RanBP1A (At1g07140), SYP41, 42, 43, and 61 (At5g26980, At4g02195, At3g05710, AT1G28490), chloride channel protein CLC-d (At5g26240), Scamp1 and 2 (At2g20840, At1g03550), GAUT10 (At2g20807), putative uncharacterized protein At4g21700 (At4g21700), and ECHIDNA (At1g09330). Using the optimized value of $k = 3$, the three independent *k*-NN experiments for the LOPIT 1, 2A, and 2B data sets were conducted, and unknown proteins were assigned to the TGN by majority vote.

Selected Reaction Monitoring

SRM is a method that allows identification and quantitation of targeted peptides of interest only with high specificity and sensitivity. Here we developed SRM assays for the confirmation of the quantitation of TGN markers and for TGN marker proteins missing from the LOPIT data sets previously described or present in low abundance and thus not quantifiable. The top-three or -four most intense fragment ions, as measured by the initial DDA analysis on the Orbitrap Velos when present, plus the four iTRAQ reporter ions were included in the method (seven to eight transitions). The measured peptide was considered to be the target peptide, and the quantitation data were subsequently used when three or four transitions (iTRAQ reporter ions not included) could be measured at the same chromatographic time point. SRMs were measured on a Quattro Premier QQQ (Waters) coupled to a nanoAcuity LC (Waters) (buffer A: HPLC H_2O, 0.1% FA; buffer B 100%

dx.doi.org/10.1021/pr4008464 | *J. Proteome Res.* 2014, 13, 763−776

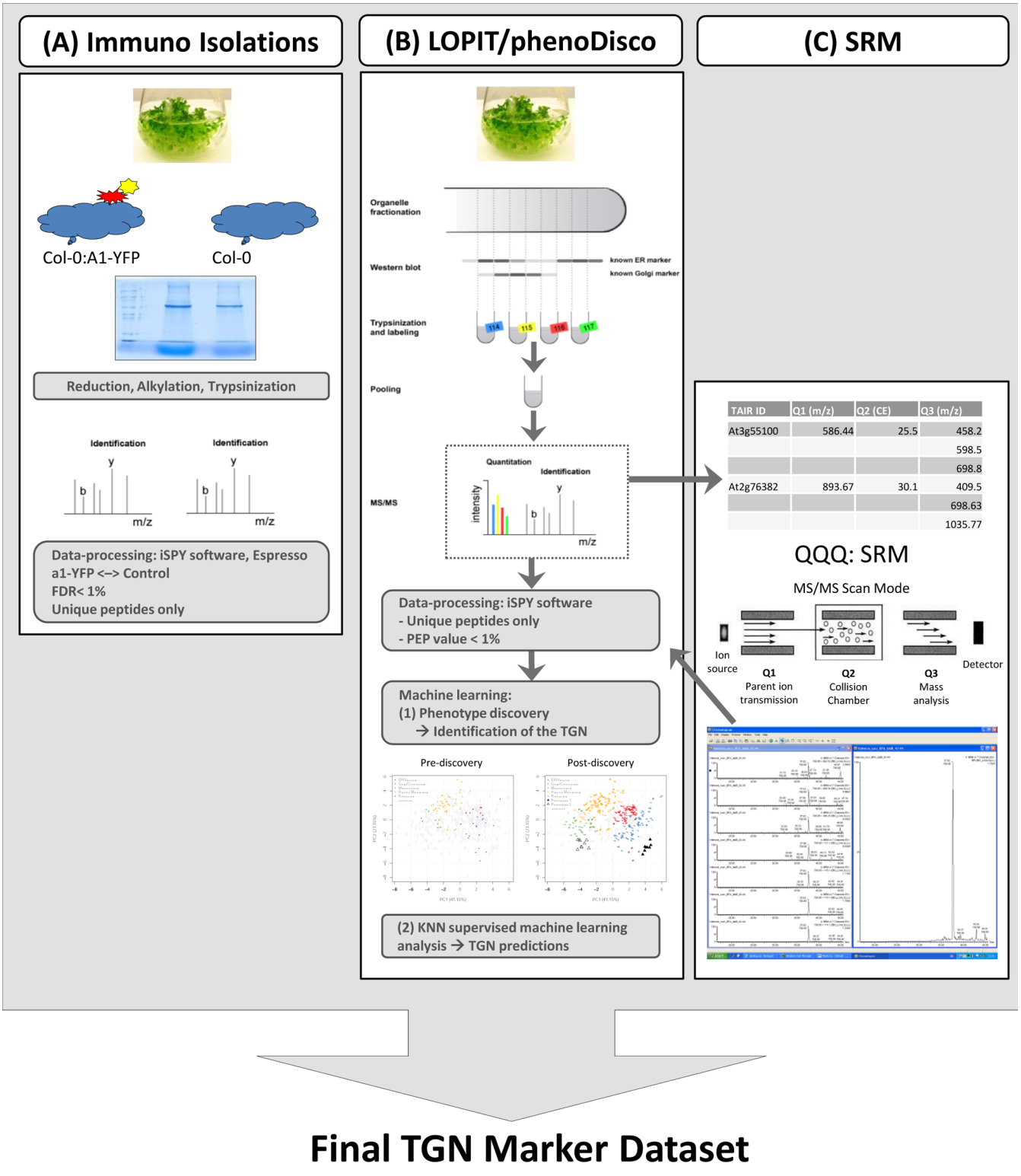


Figure 1. Workflow overview: The workflow consists of three parts: (a) immunoisolations, (b) gradient-based experiments combined with machine learning analysis on which LOPIT is based, and (c) single reaction monitoring (SRM) experiments targeted at selected proteins in the gradient experiments. Central to the analysis are the three probability-based iTRAQ-labeled gradient experiments, which form the basis of LOPIT and are analyzed by machine learning methods. To increase the number of TGN proteins that are needed to perform a reliable supervised classification method, we applied the semisupervised machine learning method *phenoDisco*. The outcome showed a cluster that included all three known TGN markers and was therefore identified as the TGN cluster. Four proteins in this cluster were considered "guilty by association" and added to the TGN marker set that was subsequently used for the three independent supervised *k*-NN machine learning experiments. SRMs were performed to test the quality of the markers and putative markers or add quantitation data in the case of missing values to gain confidence in the quantitation data and hence in the LOPIT and *phenoDisco* results. To take a protein into account as a possible TGN marker protein, a protein had to be labeled as a TGN protein in at least two out of three of the machine learning experiments. If a protein appeared in only one or two of the LOPIT data sets, then the protein must have been assigned TGN by the machine learning analysis in that single case or in both cases.

dx.doi.org/10.1021/pr4008464 | *J. Proteome Res.* 2014, 13, 763−776

ACN, 0.1% FA; 60 min gradient; 0−40 min: 3→40% buffer B, 40−45 min: 40%→100% buffer B, 45−60 min: 3% buffer B). Settings of the QQQ during for SRM were: low mass/high mass resolution (LM/HM) resolution 10, LM/HM 2 resolution 15, entrance and exit voltage: 1 V, multiplier 650 V, ion energy 1: 0.5, ion energy 2: 1.0, gas pressure: 8.63×10^{-3}, collision gas flow: 0.5. The voltage that was applied for the collision-induced dissociation was parent-ion-specific depending on mass and charge. The quantitation of the peptides was performed by calculating the area of the iTRAQ reporter ions using TargetLynx (Waters, version 4.1) and normalizing the areas to one.

RESULTS

The overall experimental strategies employed to characterize the *Arabidopsis* root TGN membrane proteome are shown in Figure 1.

Identification of Proteins in VHA-a1 Immunoisolated Fractions

Putative TGN membranes were immunoisolated using root membranes from transgenic *Arabidopsis thaliana* plants expressing VHA:a1-GFP, a TGN-localized protein (Dettmer et al. 2006), by employing a GFP antibody. Immunoisolated fractions were analyzed by Western blotting with antibodies raised against GFP to detect VHA:a1-GFP, BiP (ER marker), SYP21 (PVC marker), H^+PPase (tonoplast marker), and VSR1 (TGN/PVC) to detect any contamination from other organelles. As a control, the immunoisolation protocol was performed from wild-type (Col-0) root membranes. As shown in Figure 2A, the GFP antibody recognized a protein of the expected molecular weight (130 kDa) in a total membrane fraction from VHA:a1-GFP roots and also in the immunoisolated fractions but not in membranes from wild-type (Col-0) root. Immunoisolated fractions did not contain detectable amounts of BiP (ER), SYP21 (PVC), or the H^+-pyrophosphatase (tonoplast), suggesting low amounts of contaminant organelles. Interestingly, immunoisolated fractions also contained the vacuolar sorting receptor VSR1, which in *Arabidopsis thaliana* roots has been shown to localize both to the PVC and the TGN.[45,46,9] Four immunoisolation experiments of the VHA-a1 positive membrane fractions were performed and generated 194, 496, 224, and 308 proteins with a type-1 error cut off of 1.0×10^{-2} (1%),respectively. The VHA-a1 protein was not detected in three of four negative controls and was only present in minor amounts in the fourth immunoisolation (type 1 error 4.23×10^{-4} compared with 6.34×10^{-149}), which suggests that the proteins detected in the VHA-a1 immunoisolated fractions were specifically associated with the VHA-a1 positive membranes.

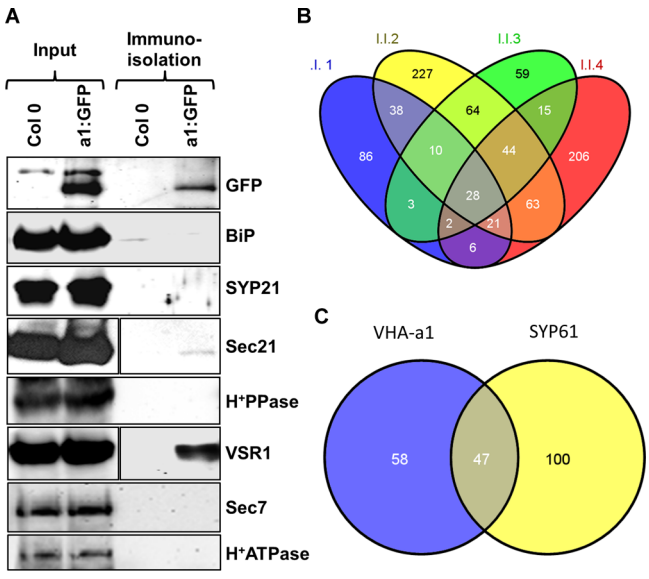


Figure 2. Immunoisolations. (A) Western Blot analysis of the immunoisolated fractions using antibodies against GFP, to detect the bait protein VHA-a1-GFP (TGN), and markers of ER (BiP), PVC (SYP21), Golgi (Sec21), TGN/PVC (VSR1), tonoplast (H+PPase), recycling endosomes (Sec7), or plasma membrane (H+ATPase). (B) Venn diagram showing the number of proteins that were found in each immunoisolation and the overlap (I.I. = immunoisolation). (C) Venn diagram showing the amount of proteins found in the VHA-a1-GFP immunoisolations and the SYP61 immunoisolations as performed by Drakakaki et al (2012) and their overlap.

Twenty-eight proteins were present in all 4 immunoisolations, and 77 were present in 3 out of 4 immunoisolated fractions (Figure 2B, Supplemental Table 1 in the Supporting Information). Several TGN marker proteins were among these including VPS45 (At1g77140), the syntaxins SYP42, SYP43, and SYP61 (At4g02195, At3g05710, At1g28490), YIP1 (At4g30260), and Ran1 (At5g44790). In addition, 36 proteins were identified that were at least associated, although not necessarily uniquely, with the TGN according to Uniprot, showing that the immunoisolation specifically enriched for the TGN. These included the vacuolar sorting receptor VSR7 (At4g20110), the cation-chloride co-transporter CCC1 (At1g30450), chloride channel protein CLC-d (At5g26240), vesicle tethering components, including the transport protein particle complexes (TRAPPs) (At5g54750 and At5g11040), some of which have been shown to localize to the TGN in yeast,[47] and the GTPase RabD1 (At3g11730), which has been shown to colocalize with RabD2A on Golgi and TGN and to cluster upon BFA treatment.[48]

No distinction could be made here between proteins that had multiple locations and proteins that reside exclusively in the TGN, and indeed the list of identified proteins included 31 proteins for which there is no evidence that they are located in the TGN. These 31 proteins could equally represent novel TGN marker proteins, false positives from contaminating membrane fractions, and proteins that were in transit to another organelle. Examples of these were the phospholipid-transporting ATPase 1, which is reputed to reside at the PM (At5g04930),[49] a probable methyltransferase PMT18 located in the ER (At1g33170), and cobra-like protein 8 located at the PM-cell wall interface (At3g16860).[49,50] Strikingly, 48 proteins had an association with the Golgi, which demonstrates the close relationship between the TGN and the Golgi but at the same time exemplifies the challenge of interpretation of protein data sets coming from immunoisolation experiments.

This list of putative TGN proteins consisted of both membrane proteins and nonmembrane proteins. Transmembrane hidden Markov modeling (TMHMM Server v. 2.0) showed at least one membrane domain in 77/105 proteins, and according to the Uniprot database, 46/105 proteins have at least one membrane domain

In 2012, a list of proteins was published after immunoisolation of a SYP61-containing vesicle fraction.[22] SYP61 has been annotated as a TGN protein and was included in our training set as such. The SYP61 compartment immunoisolation

dx.doi.org/10.1021/pr4008464 | J. Proteome Res. 2014, 13, 763−776

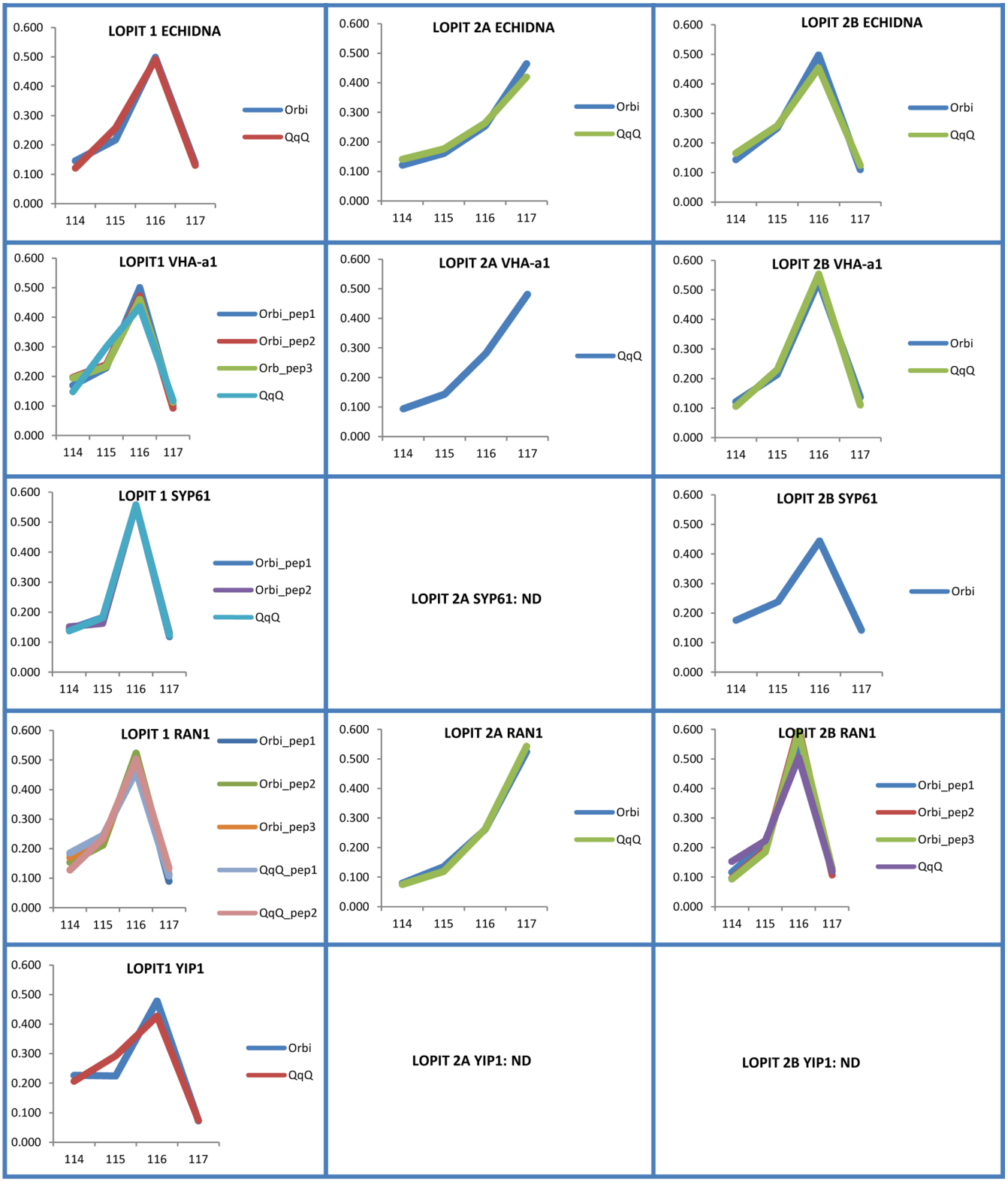


Figure 3. Results and comparison of SRM analysis and DDA data of the TGN marker proteins used in this study (ECHIDNA, Ran1, VHA-a1, SYP61, and YIP1) in all three experiments where detectable. Each line represents the iTRAQ distribution pattern of one proteotypic peptide which was acquired by either the Orbitrap (Orbi) or the Triple Quadrupole (QqQ) mass spectrometers. These distribution patterns show a comparable TGN iTRAQ distribution pattern in the same LOPIT data sets and a comparable iTRAQ distribution pattern between both methods emphasizing the confidence of the iTRAQ quantitation of the TGN markers upon which novel TGN assignments are based. The vertical axis represents the normalized-to-one intensity of the reporter ions (that are shown on the horizontal axis). The peptides were identified by measuring three or four typical fragment ions (not being the iTRAQ fragment ions), which equates to seven to eight transitions being measured in total (including the iTRAQ reporter ions) (N.D.: not detected).

dx.doi.org/10.1021/pr4008464 | J. Proteome Res. 2014, 13, 763−776

generated 147 hits, of which 47 were also found in at least three out of four of the immunoisolations reported here (Figure 2C). These included VPS45 (At1g77140), SYP42 and 43 (At4g02195, AT3G05710), YIP1 (At4g30260), Ran1 (At5g44790), and the two bait proteins SYP61 and VHA-a1 (At1g28490, At2g28520). Examples of additional proteins that were in common between the immunoisolations described here and the SYP61 proteome were the secretory carrier-associated membrane protein 3 (SCAMP3, At1g61250), YIP1-like protein (At3g05280), CESA1 (At4g32410), cation chloride cotransporter 1 (CCC1, At1g30450), and callose synthase 9 (At3g07160).

In the Drakakaki data set, 60% of the proteins were predicted or known to be associated with the endomembrane system (Drakakaki et al., 2012). The other 40% had either unknown localization or may have represented contaminating proteins or cargo being transported through SYP61 positive vesicles. For example, Callose synthase 12 (At4g03550) is associated with the PM and Golgi apparatus, and the PM-type H(+)-ATPase (At2g18960) is a PM marker and indicated as such in the Uniprot database. SYP71 (AT3G09740), a marker for the ER, was also found in the Drakakaki immunoisolations.[22]

Both of these studies exemplify the problem that arises with proteomic studies that are solely based on immunoisolation experiments. Although immunoisolations clearly enrich for TGN proteins, they generate a black and white result, as a result of which no distinction can be made between true marker proteins, proteins that are shared by other organelles, and proteins that are in transit or cargo proteins.

To better answer this question, we next compared the results of the immunoisolations with data collected upon utilization of the LOPIT technique, a method that is able to discriminate between these cases by assigning probabilities to protein localization.[27,51,31]

LOPIT Analysis of *Arabidopsis* Roots

The LOPIT data reported here represented the first such experiment performed on a tissue containing multiple cell types in plants. We noted that the quality of organelle separation was sufficient for this study, albeit less profound than that observed with *Arabidopsis* callus tissue, which is largely a source of homogeneous cells.[25] Previous studies in either a homogeneous cell line (DT-40[29]) or a whole organism (*Drosophila* embryos[30]), demonstrated similar observations where data from heterogeneous sources represent an average subcellular location for proteins within all of the different cell types present.

The setup of the LOPIT experiments was primarily targeted to membrane proteins for the following reasons: (1) membrane proteins are the main focus of interest for this study because of their pivotal role in signaling and trafficking; (2) it simplifies the complexity of the sample, which leads to more membrane protein identifications by mass spectrometry. For these reasons, all LOPIT experiments discussed here underwent a carbonate wash that removed nonmembrane proteins, as described in the Materials and Methods section.

Two LOPIT experiments on *Arabidopsis* roots were performed. The first consisted of a single 4-plex iTRAQ labeling (LOPIT 1), and the second consisted of a double 4-plex iTRAQ labeling (LOPIT 2A and 2B). After ultracentrifugation of the membrane fraction on a continuous sucrose gradient, Western blots of the different fractions were performed to check for the separation of organelles along the gradient. The TGN, Golgi apparatus, and the endoplasmic reticulum exhibited unique distributions (Supplemental Figure 1A,B in the Supporting Information). After first dimension separation of peptides by high-pH reverse-phase chromatography, 18 fractions were measured using LC−MS/MS, which resulted in the identification and quantitation of 1340 proteins for LOPIT 1 and 936 and 706 proteins for LOPIT 2A and 2B, respectively (Supplemental Table 3A−C in the Supporting Information). The TGN cluster was clearly separate from the other organelles in all three data sets, as shown in the Supplemental Figures 2A−C in the Supporting Information. Further description of proteins present in the TGN cluster is given below.

Selected Reaction Monitoring

SRM was performed for three reasons. First, SRM was carried out as a confirmation and validation of the TGN marker proteins used in LOPIT. Second, there were missing values of TGN markers from the LOPIT data set because in data-dependent tandem mass spectrometry only the most abundant peptide ions are selected for fragmentation and subsequent identification. Finally, SRM analysis was used to validate the quantitation and hence assignment of novel TGN proteins. Validation was deemed to be important because inaccurate quantitation will lead to incorrect clustering and hence incorrect protein assignments. A total of 51 SRM assays were performed that were selected for tryptic peptides derived from proteins associated with the TGN in this study, which included the TGN markers VHA-a1, SYP61, Ran1, YIP1, and ECHIDNA and potential novel TGN proteins (Supplemental Table 4 in the Supporting Information). All TGN marker proteins were confirmed by an SRM assay with the exception of SYP61 in LOPIT 2B. SRM assays confirmed the presence and quantification of the putative TGN proteins At1g61670.1 and At5g60640.1 in LOPIT 2A and At1g64200.1 and At5g08540.1 in LOPIT 2B and the TGN marker protein VHA-a1 in LOPIT 2A

Figure 3 shows the normalized iTRAQ reporter ion distribution patterns of VHA-a1, SYP61, Ran1, ECHIDNA, and YIP1 for all three data sets. It demonstrates consistency between the LOPIT data as collected using data-dependent acquisition and the SRM data. The two data sets are thus validatory of one another. Figure 3 also shows that the TGN distribution patterns of these marker proteins in each data set are comparable. For consistency, the data-dependent iTRAQ quantitation data was used as default, and only in the case of missing values was the data obtained by SRM superimposed onto the final data set.

Identification of the TGN Cluster Using a Semisupervised Learning Phenotype Discovery Approach

The reliability of novel assignments to organelles by supervised classification methods increases by the number of markers that are already known for the organelle of interest. This posed a challenge for the identification of novel TGN proteins because of the lack of sufficient TGN markers and the relatively low abundance of the TGN in samples without enrichment. In LOPIT 1, only six TGN marker proteins were found and in LOPIT 2A and 2B only three and four TGN marker proteins were identified, respectively. However, these markers showed clear separate clusters (Supplemental Figure 2 in the Supporting Information), demonstrating our success in adjusting the gradient conditions to achieve separation of the TGN from other organelles, which was necessary as the better the

dx.doi.org/10.1021/pr4008464 | J. Proteome Res. 2014, 13, 763−776

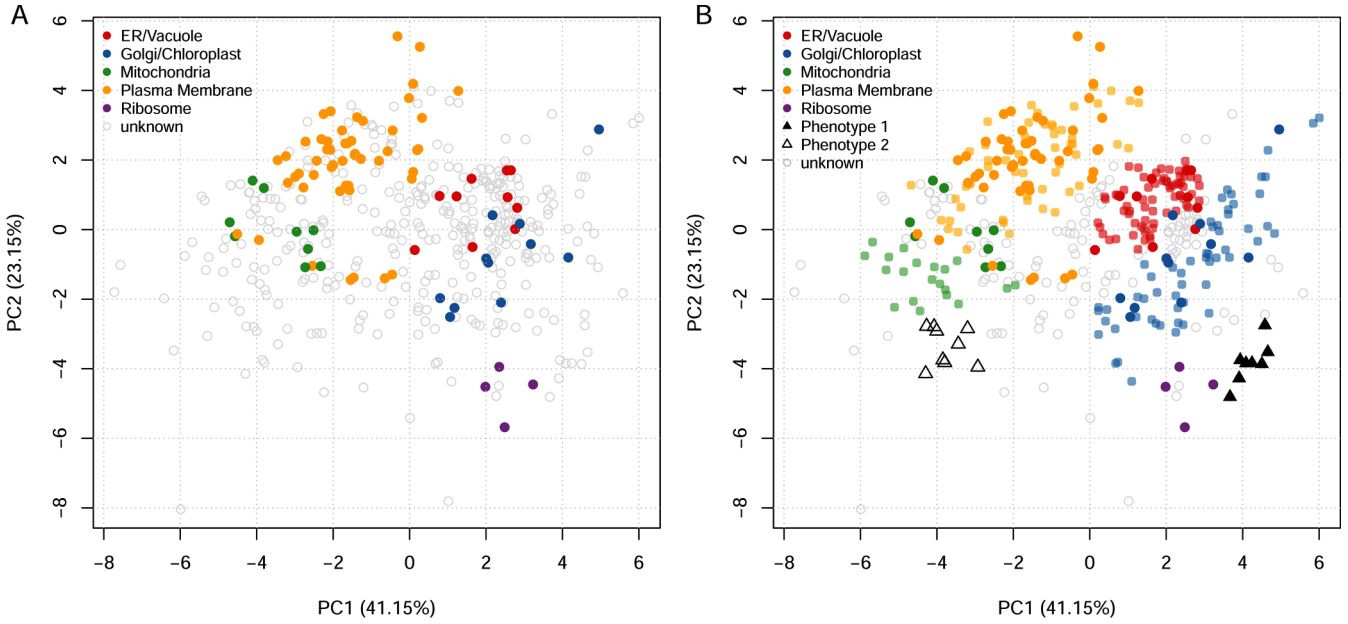


Figure 4. Principal components plot showing the *phenoDisco* results. As discussed in the supplementary methods, all three data sets were concatenated for the *phenoDisco* analysis to increase the number of quantitative values per protein to increase organellar resolution. The algorithm was given prior marker knowledge of the mitochondria, the Plasma Membrane, the ribosomes, the Golgi/Chloroplast and the vacuole/ER (A). Two clusters were discovered from the *phenoDisco* analysis of which the first one contains the TGN markers VHA-a1, Ran1 and ECHIDNA (black filled triangles) (B).

separation of the organelle of interest, the more reliable the protein assignments will be.

A semisupervised organelle discovery approach *phenoDisco* (Breckels et al., 2013) was applied to increase the number of TGN markers. The *phenoDisco* algorithm employs a semi-supervised novelty detection schema to identify additional organelle clusters of gradient profiles, beyond those identified solely by annotation. Here we employed *phenoDisco* to determine whether a TGN cluster could be identified without giving the algorithm any prior knowledge of its existence, that is, any labeled TGN markers in the training data. Using this approach, we identified two new phenotypes; phenotype 1, a cluster of predominantly TGN localized proteins, and phenotype 2, a small cluster containing cytoskeletal localized proteins (Figure 4). The TGN cluster (phenotype 1) was confirmed by the presence of the TGN marker proteins, Ran 1, VHA-a1, and ECHIDNA. Phenotype 1 consists of an additional five proteins, At4g12650, an endomembrane p70 protein, and the uncharacterized protein At1g52780, both found before to be associated with the TGN,[33] the lung seven transmembrane receptor family protein (At1g61670), and an uncharacterized protein, At5g18520 (Table 1).[52] The additional four TGN proteins previously mentioned were added to the existing TGN training set, thus increasing the training data set size for the application of supervised machine learning protein localization prediction experiments. Another protein in this cluster was the protein At3g26520, but it has been reported as a tonoplast protein in the literature, and thus it was decided not to add this protein to the TGN training set for the sake of stringency. The complete *phenoDisco* results are listed in Supplementary Table 2 in the Supporting Information.

TGN Protein Assignment Using the Supervised Classification Method *k*-NN

A supervised *K*-nearest neighbor (*k*-NN) classification was conducted in a binary '*TGN vs other*' fashion to identify TGN localized proteins. PCA plots showing the results from the *k*-NN experiments on all LOPIT data sets are shown in the Supplemental Figures 2A–C in the Supporting Information. Following the addition of the four new TGN markers found from the phenotype discovery analysis, the TGN markers set available for the supervised *k*-NN classification increased to a total of 10 TGN markers present in LOPIT 1 and to 7 and 8 TGN markers in LOPIT 2A and B, respectively. To achieve a good generalization when employing such machine learning approaches, one requires as many examples on which to train as possible.

After *k*-NN analysis, 55 proteins were assigned TGN in LOPIT 1, 22 in LOPIT 2A, and 28 proteins were assigned TGN in LOPIT 2B (Supplemental Table 5 in the Supporting Information). For a protein to be taken into account as a possible TGN marker protein, it had to be labeled as a TGN protein in at least two out of three of the machine learning experiments. If a protein appeared in only one or two of the LOPIT data sets, then the protein had to be assigned as TGN by the machine learning analysis in that single case or both cases. Forty-eight proteins fulfill these criteria (Supplemental

Table 1. *phenoDisco* Results[a]

	protein ID	protein description
1	At5g44790	Ran1
2	At2g28520	VHA-a1
3	At1g09330	ECHIDNA
4	At3g26520	aquaporin TIP1-2
5	At4g12650	endomembrane family protein 70
6	At5g18520	put. lung 7 transmemb. receptor
7	At1g52780	put. uncharac. protein
8	At1g61670	lung 7 transmemb. receptor family protein

[a]Proteins that were identified by the *phenoDisco* algorithm to belong to the same cluster that includes the TGN marker proteins Ran1, VHA-a1, and ECHIDNA.

dx.doi.org/10.1021/pr4008464 | J. Proteome Res. 2014, 13, 763–776

Table 5 in the Supporting Information). Because the TGN cluster overlapped with the ribosome cluster, our assignments contain many ribosomes. We have left these out of the final TGN marker set. Also, our aim was to obtain the TGN membrane proteome, and for this reason our final TGN data set consists only of proteins that had at least one transmembrane domain, as predicted by TMHMM. This led to a final TGN putative membrane protein list that consists of 30 proteins (Table 2). This list includes probable methyltransferases 12 and 26 (At5g06050, At5g64030), the vacuolar sorting receptors VSR1, 3, and 6 (At3g52850, At2g14740, At1g30900), and the proton pump VHA-g1 (At3g01390).

Table 2. Final List of TGN Markers[a]

	protein ID	protein name
1	At4g30260	YIP1
2	At5g44790	Ran1
3	At2g28520	VHA-a1
4	At1g28490	SYP61
5	At1g09330	ECHIDNA
6	At1g52780	put. uncharac. protein
7	At4g12650	endomembrane family protein 70
8	At1g30450	cation-chloride cotransporter 1
9	At3g21190	*O*-fucosyltransferase family protein
10	At3g26520	aquaporin TIP1−2
11	At2g14740	vacuolar-sorting receptor 3 (BP80A/VSR3)
12	At3g52850	BP80B/VSR1
13	At5g64030	probable methyltransferase PMT26
14	At1g13900	probable inactive purple acid phosphatase 2
15	At1g61670	lung 7 transmemb. receptor family protein
16	At1g08700	presenilin-like protein
17	At3g01390	V-type proton ATPase subunit G1 (VHA-G1)
18	At5g18520	put. lung 7 transmemb. receptor
19	At1g51630	*O*-fucosyltransferase family protein
20	At3g54300	vesicle-ass. memb. protein 727 (Vamp727)
21	At3g58460	uncharacterized protein
22	At1g30900	vacuolar-sorting receptor 6
23	At3g08630	put. uncharac. protein
24	At4g22750	probable *S*-acyltransferase
25	At3g04080	apyrase 1
26	At2g46890	oxidoreductase
27	At5g23040	uncharacterized protein
28	At1g12240	acid beta-fructofuranosidase 4
29	At5g06050	probable methyltransferase PMT12
30	At1g56340	calreticulin-1

[a]If a protein appeared in only one or two of the LOPIT data sets, the protein must have been assigned TGN by the machine learning analysis to be included in the final stringent TGN list. If a protein appeared in all three data sets, then it was assigned to the final TGN list if it was labelled as a TGN protein in at least two out of three of the machine learning experiments. This list includes membrane proteins only. Proteins in red are the original TGN markers used and identified in the analysis.

DISCUSSION

In this study, we have combined the data acquired using three techniques, SRM, LOPIT, and computational modeling, to create a stringent and robust TGN membrane protein data set, and we have compared these results with immunoisolation data.

For the first time, a LOPIT data set is presented on *Arabidopsis thaliana* root tissue. Because of its biological complexity and practical difficulties, this tissue is challenging to work with, but it will also generate more biologically representative data than Callus cell suspension tissue. Furthermore, biologically complex tissue will be subject to organelle proteomics questions in the future, perhaps also involving perturbation. This manuscript shows that LOPIT is able to generate quality data even when the biological questions are demanding.

In 2012, Drakakaki and colleagues published a data set of TGN proteins based on the immunoisolation of a SYP61-positive compartment.[22] SYP61 is a known TGN marker and is also present in the training set of proteins we applied in the study presented here. The VHA-a1 immunoisolated fractions described in this study show a 32% overlap (47/147) with the published SYP61 compartment proteome, taking into account only those proteins that were at least detected in three out of four VHA-a1 immunoisolations. Some of these have already been shown to locate to the TGN, for instance, SYP42 and SYP43 or VTI12.[53−55] The relatively poor overlap deftly exemplifies the problem with this approach. As already described by Drakakaki and coworkers, false positives are present in the data set. For example, SYP121 (At3g11820), SYP71 (At3g09740), and Ara2 (At1g06400) found in the immunoisolation data sets are all PM proteins, as described in literature, but are present in the final data set. The proteins previously mentioned were identified in the LOPIT data set and were distinctively and differently localized on the PCA plot (Supplemental Table 6 and Supplemental Figure 3 in the Supporting Information). This neatly demonstrates the strength of LOPIT analysis that determines the steady-state positions of proteins and hence is able to distinguish cargo and contaminants from full-time residents of organelles. The strength of a multidisciplinary approach may be further exemplified by the following two examples. The glucan synthase-like protein ATGSL5 (At4g03550), which was used as a PM marker in our training set, clusters with the PM, is considered to be a PM protein in the literature,[56] but was found in three out of four VHA-a1 immunoisolation experiments and is also present in the Drakakaki data set. SRMs were performed to measure the VHA-a3 (At1g64200), which was assigned TGN in LOPIT 1 but was missing from LOPIT 2A. After adding SRM data for this protein, it was no longer assigned to the TGN in LOPIT 2B. Hence this protein was not considered as a TGN protein.

Not all proteins that were identified by immunoisolations but were missing in LOPIT or assigned "other" should be considered false positives. First, some of these proteins may also be TGN residents, which are shared by other compartments or proteins that traffic through the TGN or cargo proteins. However, one of the aims of this study was to present a TGN protein list consisting of proteins whose steady-state position is within the TGN. For example, VTi12 (At1g26670) that was found in the immunoisolations is involved in vesicle docking of transport vesicles in the TGN but has also been reported to localize to the Golgi and PVC.[55] The protein was identified in LOPIT 2A and 2B but was not assigned as TGN. However, this does not mean it is a false positive; it clearly has a function in the TGN and is a resident of the TGN, but it is not included in our final data set because its steady-state location does not coincide with the majority of the TGN marker set. Second, TGN proteins found in both the VHA-a1 and SYP61 immunoisolations that are not identified in the LOPIT experiments presented here, for example, the syntaxin SYP43

(At3g05710, TGN[57]), could be low abundance proteins and were therefore not identified. Third, the approach for LOPIT and *phenoDisco* performed here was selective for membrane proteins due to a carbonate wash, and hence nonmembrane proteins were discarded during the process. The immunoisolations, however, contained both membrane and nonmembrane proteins (Supplementary Table 1 in the Supporting Information). This may be the reason why only few proteins of the Rab family were found in the data sets (not assigned TGN) but were present in the immunoisolations.

The creation and application of a robust training set is key to the accuracy of protein subcellular assignments using the LOPIT data analysis pipeline. The training set applied in this LOPIT analysis was very stringent, which decreased the number of TGN markers available for use but concomitantly decreased the chance of assigning proteins to the TGN erroneously. TGN proteins with low confidence in the training set can result in false assignments, which may perpetuate new low-confidence assignments in subsequent LOPIT experiments when these themselves are applied within new iterations of training data. This effect is even more profound when small training sets are used, as is the case for the TGN, where an incorrect marker protein may significantly bias assignments. Conversely, more reliable TGN markers in the training set lead to additional and more reliable identifications of TGN markers in subsequent data sets.

The final stringent list of 30 TGN membrane proteins including the TGN training set marker proteins share 10 proteins with the study of Drakakaki et al.[22] These 10 proteins include the proteins that constitute our training set: VHA-a1, SYP61, YIP1, Ran1, and ECHIDNA. In addition, the overlapping list includes VSR3 (At2g14740), the uncharacterized protein At1g52780, the cation chloride cotransporter 1 (CCC1, At1g30450), the *S*-adenosyl-L-methionine-dependent methyltransferase PMT26 (At5g64030), and the putatative lung 7 transmembrane receptor (At5g18520).

The proteins that were not present in the data set of Drakakaki but present in our final high stringency list include the endomembrane family protein 70 (EMP70) (At4g12650), a yeast homologue-related protein that has been shown to localize to early endosomal compartments and to be required for endosomal sorting.[58] Furthermore, the list includes the SNARE VAMP27 (Vesicle-associated membrane protein 727, At3g54300), which has been shown to localize on subpopulations of FM4-64-stained endosomes but not at the PM.[17b,59] We also find presenilin-1 (At1g08700), a protein similar to animal presenilin, a component of the gamma-secretase complex, which may function in endosomes and the TGN in animal cells,[60,61] two *O*-fucosyltransferase family proteins (At3g21190, At1g51630), and the vacuolar sorting receptors VSR1 and 6 (At3g52850, At1g30900). The presence of VSRs in the final list is consistent with the presence of VSR1 in the immunoisolated fractions (Figure 2A) and with their role in the transport of cargo proteins at the TGN for vacuolar transport via the PVC.[62] The list includes PAP2 (probable inactive purple acide phosphatase 2 (At1g13900), calreticulin (At1g56340), apyrase (At3g04080), oxidoreductase (At2g46890), acid beta-fructofuranosidase 4 (At1g12240), the uncharacterized proteins At3g08630, At5g23040, and At3g58460, VHA-g1 (At3g01390), which is described to have multiple localizations[63,64] but in this study was shown to have a steady-state position in the TGN, the lung 7 transmembrane receptor family protein (At1g61670), Tip1−2 (At3g26520) (present on multiple localization according to Uniprot but again the steady in this data set is TGN), and the DHHC-type zinc finger family protein (At4g22750).

Here we have shown that LOPIT offers opportunities to study organelle residency because LOPIT is a gradient-based technique that generates probabilities for protein localization instead of generating a black and white list, which is paramount to the analysis strategy we have employed in this study. It returns information about the steady-state protein localization in subcellular compartments simultaneously and is able to dissect this from proteins present in more than one organelle and cargo proteins en route to other cellular destinations. The phenotype discovery algorithm (*phenoDisco*) identifies organelles and subcompartments present in the data without any prior knowledge of their existence, adding value to the existence of the TGN cluster as well as identifying proteins that can be used in further supervised machine learning analysis. The SRM experiments add confidence to the training set and the identified proteins.

The approach is a "self-learning" system that is improved by training such that old and new data sets can continuously be interrogated, serving as a data set memory. The multidisciplinary-probability-based approach offers the possibility to look at subtle dynamic changes in protein localization upon perturbation of the system by ligands, gene knockout, disease state, or differences in protein localization in developmental stages. In these cases, it is even harder to distinguish between cargo, proteins that have multiple localization, false positives, true residents, and, additionally, proteins that change localization as a result of the perturbation. The more data sets available, the more reliable the outcome will be and the more subtle the changes in relocalization that can be monitored.

This is a novel approach in organelle proteomics and will assist in the assignment of proteins to smaller, more dynamic organelles. It will also assist when more information is required about the trafficking of many proteins in relation to their function instead of assigning a list of proteins to well-characterized organelles. The ultimate goal for the future will be to have a systems overview of the spatial and temporal dynamics of proteins and to elucidate cellular mechanisms to gain biological insight impacting our understanding of disease.[23]

■ ASSOCIATED CONTENT

Supporting Information

Western blots showing the distribution of protein markers along the gradient from LOPIT experiments 1 and 2 A,B, respectively. Principal components plots showing the results from the *k*-NN classification of unknown proteins to the TGN or "other", before and after analysis in LOPIT 1, LOPIT 2A, and LOPIT 2B experiments. Comparison of LOPIT 1 with SYP61 and VHA-a1 immunoisolation experiments. Immunoisolation results. PhenoDisco input and results for all three LOPIT data sets. Total list of proteins identified by LOPIT in LOPIT 1, LOPIT 2A, and LOPIT 2B with the protein Q-values, *k*-NN assignments, and normalized iTRAQ distribution data. Results of all SRM analyses in all three LOPIT experiments. LOPIT data after *k*-NN analysis. Comparison of LOPIT 1 with SYP61 and VHA-a1 immunoisolation experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel:+441223760255. Fax:+441223333345. E-mail: k.s.lilley@bioc.cam.ac.uk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The transgenic line expressing VHA:a1-GFP was kindly provided by Karin Schumacher (University of Heidelberg). We thank Michael Deery for assistance with mass spectrometric analysis and Laurent Gatto for useful discussions about data analysis and the application of machine learning approaches. A.J.G., G.S.-A., and F.A. were supported by an ERA-PG Consortium award (BBSRC grant BB/E024777/), and additionally A.J.G. was supported by a generous gift from King Abdullah University for Science and Technology, Saudi Arabia. L.M.B. was primarily funded by BBSRC grant BB/KK00137X/1. G.S.A. was a recipient of a fellowship from Universitat de Valencia (V Segles Program). F.A. was supported by the Ministerio de Ciencia e Innovación (grant no BFU2009_07039) and Generalitat Valenciana (ISIC/2013/004).

REFERENCES

(1) Park, M.; Jürgens, G. Membrane traffic and fusion at post-Golgi compartments. *Front. Plant Sci.* **2012**, *2*, 111.

(2) Dhonukshe, P.; Aniento, F.; Hwang, I.; Robinson, D. G.; Mravec, J.; Stierhof, Y.-D.; Friml, J. Clathrin-Mediated Constitutive Endocytosis of PIN Auxin Efflux Carriers in Arabidopsis. *Curr. Biol.* **2007**, *17* (6), 520−527.

(3) Robert, S.; Kleine-Vehn, J.; Barbez, E.; Sauer, M.; Paciorek, T.; Baster, P.; Vanneste, S.; Zhang, J.; Simon, S.; ovanová, M.; Hayashi, K.; Dhonukshe, P.; Yang, Z.; Bednarek, S. Y.; Jones, A. M.; Luschnig, C.; Aniento, F.; Zažímalová, E.; Friml, J. ABP1Mediates Auxin Inhibition of Clathrin-Dependent Endocytosis in Arabidopsis. *Cell* **2010**, *143* (1), 111−121.

(4) Kitakura, S.; Vanneste, S.; Robert, S.; Löfke, C.; Teichmann, T.; Tanaka, H.; Friml, J. Clathrin Mediates Endocytosis and Polar Distribution of PIN Auxin Transporters in Arabidopsis. *Plant Cell* **2011**, *23* (5), 1920−1931.

(5) Kleine-Vehn, J.; Friml, J. Polar Targeting and Endocytic Recycling in Auxin-Dependent Plant Development. *Annu. Rev. Cell Dev. Biol.* **2008**, *24* (1), 447−473.

(6) Geldner, N.; Jurgens, G. Endocytosis in signalling and development. *Curr. Opin. Plant Biol.* **2006**, *9* (6), 589−594.

(7) Takano, J.; Tanaka, M.; Toyoda, A.; Miwa, K.; Kasai, K.; Fuji, K.; Onouchi, H.; Naito, S.; Fujiwara, T. Polar localization and degradation of Arabidopsis boron transporters through distinct trafficking pathways. *Proc. Nat. Acad. Sci.* **2010**, *107* (11), 5220−5225.

(8) Robatzek, S.; Chinchilla, D.; Boller, T. , Ligand-induced endocytosis of the pattern recognition receptor FLS2 in Arabidopsis. *Genes Dev.* **2006**, *20* (5), 537−542.

(9) Viotti, C.; Bubeck, J.; Stierhof, Y.-D.; Krebs, M.; Langhans, M.; van den Berg, W.; van Dongen, W.; Richter, S.; Geldner, N.; Takano, J.; Jürgens, G.; de Vries, S. C.; Robinson, D. G.; Schumacher, K. Endocytic and Secretory Traffic in Arabidopsis Merge in the Trans-Golgi Network/Early Endosome, an Independent and Highly Dynamic Organelle. *Plant Cell* **2010**, *22* (4), 1344−1357.

(10) Dettmer, J.; Hong-Hermesdorf, A.; Stierhof, Y. D.; Schumacher, K. Vacuolar H+-ATPase activity is required for Endocytic and secretory trafficking in Arabidopsis. *Plant Cell* **2006**, *18* (3), 715−730.

(11) Lam, S. K.; Siu, C. L.; Hillmer, S.; Jang, S.; An, G.; Robinson, D. G.; Jiang, L. Rice SCAMP1 Defines Clathrin-Coated, trans-Golgi-Located Tubular-Vesicular Structures as an Early Endosome in Tobacco BY-2 Cells. *Plant Cell* **2007**, *19* (1), 296−319.

(12) (a) Kang, B.-H.; Nielsen, E.; Preuss, M. L.; Mastronarde, D.; Staehelin, L. A. Electron Tomography of RabA4b- and PI-4Kβ1-Labeled Trans Golgi Network Compartments in Arabidopsis. *Traffic* **2011**, *12* (3), 313−329. (b) Chow, C.-M.; Neto, H. l.; Foucart, C.; Moore, I. Rab-A2 and Rab-A3 GTPases Define a trans-Golgi Endosomal Membrane Domain in Arabidopsis That Contributes Substantially to the Cell Plate. *Plant Cell* **2008**, *20* (1), 101−123.

(13) Otegui, M. S.; Herder, R.; Schulze, J.; Jung, R.; Staehelin, L. A. The Proteolytic Processing of Seed Storage Proteins in Arabidopsis Embryo Cells Starts in the Multivesicular Bodies. *Plant Cell* **2006**, *18* (10), 2567−2581.

(14) Zerial, M.; McBride, H. Rab proteins as membrane organizers. *Nat. Rev. Mol. Cell Biol.* **2001**, *2* (2), 107−117.

(15) Zeigerer, A.; Gilleron, J.; Bogorad, R. L.; Marsico, G.; Nonaka, H.; Seifert, S.; Epstein-Barash, H.; Kuchimanchi, S.; Peng, C. G.; Ruda, V. M.; Conte-Zerial, P. D.; Hengstler, J. G.; Kalaidzidis, Y.; Koteliansky, V.; Zerial, M. Rab5 is necessary for the biogenesis of the endolysosomal system in vivo. *Nature* **2012**, *485* (7399), 465−470.

(16) (a) Lee, M. C. S.; Miller, E. A.; Goldberg, J.; Orci, L.; Schekman, R. Bi-directional protein transport between the ER and Golgi. *Annu. Rev. Cell Dev. Biol.* **2004**, *20*, 87−123. (b) Reichardt, I.; Stierhof, Y.-D.; Mayer, U.; Richter, S.; Schwarz, H.; Schumacher, K.; Jürgens, G. Plant Cytokinesis Requires De Novo Secretory Trafficking but Not Endocytosis. *Curr. Biol.* **2007**, *17* (23), 2047−2053.

(17) (a) Ueda, T.; Yamaguchi, M.; Uchimiya, H.; Nakano, A. Ara6, a plant-unique novel type Rab GTPase, functions in the endocytic pathway of Arabidopsis thaliana. *EMBO J.* **2001**, *20* (17), 4730−4741. (b) Ueda, T.; Uemura, T.; Sato, M. H.; Nakano, A. Functional differentiation of endosomes in Arabidopsis cells. *Plant J.* **2004**, *40* (5), 783−789.

(18) Jung, E.; Hoogland, C.; Chiappe, D.; Sanchez, J. C.; Hochstrasser, D. F. The establishment of a human liver nuclei two-dimensional electrophoresis reference map. *Electrophoresis* **2000**, *21* (16), 3483−3487.

(19) Taylor, N. L.; Heazlewood, J. L.; Millar, A. H. The Arabidopsis thaliana 2-D gel mitochondrial proteome: Refining the value of reference maps for assessing protein abundance, contaminants and post-translational modifications. *Proteomics* **2011**, *11* (9), 1720−1733.

(20) Lilley, K. S.; Dupree, P. Methods of quantitative proteomics and their application to plant organelle characterization. *J. Exp. Bot.* **2006**, *57* (7), 1493−1499.

(21) Parsons, H. T.; Christiansen, K.; Knierim, B.; Carroll, A.; Ito, J.; Batth, T. S.; Smith-Moritz, A. M.; Morrison, S.; McInerney, P.; Hadi, M. Z.; Auer, M.; Mukhopadhyay, A.; Petzold, C. J.; Scheller, H. V.; Loqué, D.; Heazlewood, J. L. Isolation and Proteomic Characterization of the Arabidopsis Golgi Defines Functional and Novel Components Involved in Plant Cell Wall Biosynthesis. *Plant Physiol.* **2012**, *159* (1), 12−26.

(22) Drakakaki, G.; van de Ven, W.; Pan, S.; Miao, Y.; Wang, J.; Keinath, N. F.; Weatherly, B.; Jiang, L.; Schumacher, K.; Hicks, G.; Raikhel, N. Isolation and proteomic analysis of the SYP61 compartment reveal its role in exocytic trafficking in Arabidopsis. *Cell Res.* **2012**, *22* (2), 413−424.

(23) Vidal, M.; Chan, D.; Gerstein, M.; Mann, M.; Omenn, G.; Tagle, D.; Sechi, S.; Participants, W. The human proteome - a scientific opportunity for transforming diagnostics, therapeutics, and healthcare. *Clin. Proteomics* **2012**, *9* (1), 6.

(24) (a) Foster, L. J.; de Hoog, C. L.; Zhang, Y. L.; Zhang, Y.; Xie, X. H.; Mootha, V. K.; Mann, M. A mammalian organelle map by protein correlation profiling. *Cell* **2006**, *125* (1), 187−199. (b) Wiese, S.; Gronemeyer, T.; Ofman, R.; Kunze, M.; Grou, C. P.; Almeida, J. A.; Eisenacher, M.; Stephan, C.; Hayen, H.; Schollenberger, L.; Korosec, T.; Waterham, H. R.; Schliebs, W.; Erdmann, R.; Berger, J.; Meyer, H. E.; Just, W.; Azevedo, J. E.; Wanders, R. J. A.; Warscheid, B. Proteomics characterization of mouse kidney Peroxisomes by tandem mass spectrometry and protein correlation profiling. *Mol. Cell. Proteomics* **2007**, *6* (12), 2045−2057.

(25) Dunkley, T. P. J.; Hester, S.; Shadforth, I. P.; Runions, J.; Weimar, T.; Hanton, S. L.; Griffin, J. L.; Bessant, C.; Brandizzi, F.; Hawes, C.; Watson, R. B.; Dupree, P.; Lilley, K. S. Mapping the Arabidopsis organelle proteome. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (17), 6518−6523.
(26) de Duve, C. Tissue fractionation. *J. Cell Biol.* **1971**, *50*, 20D−55D.
(27) Dunkley, T. P. J.; Watson, R.; Griffin, J. L.; Dupree, P.; Lilley, K. S. Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell. Proteomics* **2004**, *3* (11), 1128−1134.
(28) Sadowski, P. G.; Groen, A. J.; Dupree, P.; Lilley, K. S. Sub-cellular localization of membrane proteins. *Proteomics* **2008**, *8* (19), 3991−4011.
(29) Hall, S. L.; Hester, S.; Griffin, J. L.; Lilley, K. S.; Jackson, A. P. The Organelle Proteome of the DT40 Lymphocyte Cell Line. *Mol. Cell. Proteomics* **2009**, *8* (6), 1295−1305.
(30) Tan, D. J. L.; Dvinge, H.; Christoforou, A.; Bertone, P.; Martinez Arias, A.; Lilley, K. S. Mapping Organelle Proteins and Protein Complexes in Drosophila melanogaster. *J. Proteome Res.* **2009**, *8* (6), 2667−2678.
(31) Nikolovski, N.; Rubtsov, D.; Segura, M. P.; Miles, G. P.; Stevens, T. J.; Dunkley, T. P. J.; Munro, S.; Lilley, K. S.; Dupree, P. Putative Glycosyltransferases and Other Plant Golgi Apparatus Proteins Are Revealed by LOPIT Proteomics. *Plant Physiol.* **2012**, *160* (2), 1037−1051.
(32) Trotter, M. W. B.; Sadowski, P. G.; Dunkley, T. P. J.; Groen, A. J.; Lilley, K. S. Improved sub-cellular resolution via simultaneous analysis of organelle proteomics data across varied experimental conditions. *Proteomics* **2010**, *10* (23), 4213−4219.
(33) Breckels, L. M.; Gatto, L.; Christoforou, A.; Groen, A. J.; Lilley, K. S.; Trotter, M. W. B. The effect of organelle discovery upon sub-cellular protein localisation. *J. Proteomics* **2013**, *88*, 129−140.
(34) Pedrazzini, E.; Giovinazzo, G.; Bielli, A.; de Virgilio, M.; Frigerio, L.; Pesca, M.; Faoro, F.; Bollini, R.; Ceriotti, A.; Vitale, A. Protein quality control along the route to the plant vacuole. *Plant Cell* **1997**, *9* (10), 1869−80.
(35) Pimpl, P.; Movafeghi, A.; Coughlan, S.; Denecke, J. r.; Hillmer, S.; Robinson, D. G. In Situ Localization and in Vitro Induction of Plant COPI-Coated Vesicles. *Plant Cell* **2000**, *12* (11), 2219−2235.
(36) Li, J.; Wen, J.; Lease, K. A.; Doke, J. T.; Tax, F. E.; Walker, J. C. BAK1, an Arabidopsis LRR Receptor-like Protein Kinase, Interacts with BRI1 and Modulates Brassinosteroid Signaling. *Cell* **2002**, *110* (2), 213−222.
(37) Steinmann, T.; Geldner, N.; Grebe, M.; Mangold, S.; Jackson, C. L.; Paris, S.; Gälweiler, L.; Palme, K.; Jürgens, G. Coordinated Polar Localization of Auxin Efflux Carrier PIN1 by GNOM ARF GEF. *Science* **1999**, *286* (5438), 316−318.
(38) Preuss, M. L.; Serna, J.; Falbel, T. G.; Bednarek, S. Y.; Nielsen, E. The Arabidopsis Rab GTPase RabA4b Localizes to the Tips of Growing Root Hair Cells. *Plant Cell* **2004**, *16* (6), 1589−1603.
(39) Gälweiler, L.; Guan, C.; Mueller, A.; Wisman, E.; Mendgen, K.; Yephremov, A.; Palme, K. Regulation of Polar Auxin Transport by AtPIN1 in Arabidopsis Vascular Tissue. *Science* **1998**, *282* (5397), 2226−2230.
(40) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; aMoulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918−920.
(41) Gutteridge, A.; Pir, P.; Castrillo, J.; Charles, P.; Lilley, K.; Oliver, S. Nutrient control of eukaryote cell growth: a systems biology study in yeast. *BMC Biol.* **2010**, *8* (1), 68.
(42) Brosch, M.; Yu, L.; Hubbard, T.; Choudhary, J. Accurate and Sensitive Peptide Identification with Mascot Percolator. *J. Proteome Res.* **2009**, *8* (6), 3176−3181.
(43) Gatto, L.; Lilley, K. S. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* **2012**, *28* (2), 288−289.
(44) Gentleman, R.; Carey, V.; Bates, D.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; Hornik, K.; Hothorn, T.; Huber, W.; Iacus, S.; Irizarry, R.; Leisch, F.; Li, C.; Maechler, M.; Rossini, A.; Sawitzki, G.; Smith, C.; Smyth, G.; Tierney, L.; Yang, J.; Zhang, J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **2004**, *5* (10), R80.
(45) Niemes, S.; Labs, M.; Scheuring, D.; Krueger, F.; Langhans, M.; Jesenofsky, B.; Robinson, D. G.; Pimpl, P. Sorting of plant vacuolar proteins is initiated in the ER. *Plant J.* **2010**, *62* (4), 601−614.
(46) Stierhof, Y.-D.; El Kasmi, F. Strategies to improve the antigenicity, ultrastructure preservation and visibility of trafficking compartments in Arabidopsis tissue. *Eur. J. Cell Biol.* **2010**, *89* (2−3), 285−297.
(47) Thellmann, M.; Rybak, K.; Thiele, K.; Wanner, G.; Assaad, F. F. Tethering Factors Required for Cytokinesis in Arabidopsis. *Plant Physiol.* **2010**, *154* (2), 720−732.
(48) Pinheiro, H.; Samalova, M.; Geldner, N.; Chory, J.; Martinez, A.; Moore, I. Genetic evidence that the higher plant Rab-D1 and Rab-D2 GTPases exhibit distinct but overlapping interactions in the early secretory pathway. *J. Cell Sci.* **2009**, *122* (20), 3749−3758.
(49) Gomes, E.; Jakobsen, M. K.; Axelsen, K. B.; Geisler, M.; Palmgren, M. G. Chilling Tolerance in Arabidopsis Involves ALA1, a Member of a New Family of Putative Aminophospholipid Translocases. *Plant Cell* **2000**, *12* (12), 2441−2453.
(50) Roudier, F.; Schindelman, G.; DeSalle, R.; Benfey, P. N. The COBRA Family of Putative GPI-Anchored Proteins in Arabidopsis. A New Fellowship in Expansion. *Plant Physiol.* **2002**, *130* (2), 538−548.
(51) Sadowski, P. G.; Dunkley, T. P.; S., I.; Dupree, P.; Bessant, C.; Griffin, J. L.; Lilley, K. S. Quantitative proteomic approach to study subcellular localization of membrane proteins. *Nat. Protoc.* **2006**, *1* (4), 1778−89.
(52) Gattolin, S.; Sorieul, M.; Hunter, P.; Khonsari, R.; Frigerio, L. In vivo imaging of the tonoplast intrinsic protein family in Arabidopsis roots. *BMC Plant Biol.* **2009**, *9* (1), 133.
(53) Sanderfoot, A. A.; Kovaleva, V.; Zheng, H.; Raikhel, N. V. The t-SNARE AtVAM3p Resides on the Prevacuolar Compartment in Arabidopsis Root Cells. *Plant Physiol.* **1999**, *121* (3), 929−938.
(54) Chen, Y.; Shin, Y.-K.; Bassham, D. C. YKT6 is a Core Constituent of Membrane Fusion Machineries at the Arabidopsis trans-Golgi Network. *J. Mol. Biol.* **2005**, *350* (1), 92−101.
(55) Zheng, H.; von Mollard, G. F.; Kovaleva, V.; Stevens, T. H.; Raikhel, N. V. The Plant Vesicle-associated SNARE AtVTI1a Likely Mediates Vesicle Transport from the Trans-Golgi Network to the Prevacuolar Compartment. *Mol. Biol. Cell* **1999**, *10* (7), 2251−2264.
(56) Jacobs, A. K.; Lipka, V.; Burton, R. A.; Panstruga, R.; Strizhov, N.; Schulze-Lefert, P.; Fincher, G. B. An Arabidopsis Callose Synthase, GSL5, Is Required for Wound and Papillary Callose Formation. *Plant Cell* **2003**, *15* (11), 2503−2513.
(57) Uemura, T.; Ueda, T.; Ohniwa, R. L.; Nakano, A.; Takeyasu, K.; Sato, M. H. Systematic Analysis of SNARE Molecules in Arabidopsis: Dissection of the post-Golgi Network in Plant Cells. *Cell Struct. Funct.* **2004**, *29* (2), 49−65.
(58) Aguilar, P. S.; Frohlich, F.; Rehman, M.; Shales, M.; Ulitsky, I.; Olivera-Couto, A.; Braberg, H.; Shamir, R.; Walter, P.; Mann, M.; Ejsing, C. S.; Krogan, N. J.; Walther, T. C. A plasma-membrane E-MAP reveals links of the eisosome with sphingolipid metabolism and endosomal trafficking. *Nat. Struct. Mol. Biol.* **2010**, *17* (7), 901−908.
(59) Ebine, K.; Miyakawa, N.; Fujimoto, M.; Uemura, T.; Nakano, A.; Ueda, T. Endosomal trafficking pathway regulated by ARA6, a RAB5 GTPase unique to plants. *Small GTPases* **2012**, *3* (1), 23−27.
(60) Baulac, S.; LaVoie, M. J.; Kimberly, W. T.; Strahle, J.; Wolfe, M. S.; Selkoe, D. J.; Xia, W. Functional gamma-secretase complex

assembly in Golgi/trans-Golgi network: interactions among presenilin, nicastrin, Aph1, Pen-2, and gamma-secretase substrates. *Neurobiol. Dis.* **2003**, *14* (2), 194−204.
(61) Vetrivel, K. S.; Cheng, H.; Lin, W.; Sakurai, T.; Li, T.; Nukina, N.; Wong, P. C.; Xu, H.; Thinakaran, G. Association of Gamma-Secretase with Lipid Rafts in Post-Golgi and Endosome Membranes. *J. Biol. Chem.* **2004**, *279* (43), 44945−44954.
(62) Suen, P. K.; Shen, J.; Sun, S. S. M.; Jiang, L. Expression and characterization of two functional vacuolar sorting receptor (VSR) proteins, BP-80 and AtVSR4 from culture media of transgenic tobacco BY-2 cells. *Plant Sci.* **2010**, *179* (1−2), 68−76.
(63) Rouquie, D.; Tournaire-Roux, C.; Szponarski, W.; Rossignol, M.; Doumas, P. Cloning of the V-ATPase subunit G in plant: functional expression and sub-cellular localization. *FEBS Lett.* **1998**, *437* (3), 287−292.
(64) Nelson, N.; Perzov, N.; Cohen, A.; Hagai, K.; Padler, V.; Nelson, H. The cellular biology of proton-motive force generation by V-ATPases. *J. Exp. Biol.* **2000**, *203* (1), 89−95.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on January 17, 2014 with a truncated Supporting Information file. The corrected version with the full Supporting Information file was reposted on January 22, 2014