The Plant Journal (2006) **46**, 708–721 doi: 10.1111/j.1365-313X.2006.02724.x

TECHNICAL ADVANCE

Membrane-associated transcripts in Arabidopsis; their isolation and characterization by DNA microarray analysis and bioinformatics

Mark de Jong, Bas van Breukelen[†], Floyd R. Wittink[‡], Frank L.H. Menke, Peter J. Weisbeek and Guido Van den Ackerveken[*]
Molecular Genetics Group, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received 30 September 2005; revised 17 January 2006; accepted 30 January 2006.
[*]For correspondence (fax +31 30 253 2837; e-mail g.vandenackerveken@bio.uu.nl).
[†]Present address: Department of Biomolecular Mass Spectrometry, Faculty of Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands.
[‡]Present address: MicroArray Department, Swammerdam Institute for Life Sciences, University of Amsterdam, Kruislaan 318, 1098 SM, Amsterdam, The Netherlands.

Summary

Membrane-associated, integral membrane and secreted proteins are of key importance in many cellular processes. For most of the 28 952 predicted proteins in Arabidopsis, the actual subcellular localization has not been demonstrated experimentally. So far, their potential membrane-association has been deduced from algorithms that predict transmembrane domains and signal peptides. However, the comprehensiveness and accuracy of these algorithms is still limited. The majority of membrane-associated and secreted proteins is synthesized on membrane-bound polysomes. Therefore, the isolation and characterization of mRNA associated with membrane-bound polysomes offers an experimental tool for the genome-wide identification of these proteins. Here we describe an efficient method to isolate mRNA from membrane-bound polysomes and report on the validation of the method to enrich for transcripts encoding membrane-associated and secreted proteins. The sensitivity and reproducibility of the isolation method was investigated by DNA microarray analysis. Pearson correlations between transcript levels obtained from three replicate isolations showed that the method is highly reproducible. A significant enrichment for mRNAs encoding proteins containing predicted transmembrane domains and signal peptides was observed in the membrane-bound polysomal fraction. In this fraction, 301 transcripts were classified by gene ontologies as 'cellular component unknown', and potentially encode previously unrecognized secreted or membrane-associated proteins.

Keywords: *Arabidopsis thaliana*, CATMA, membrane-bound polysomes, microarray, mRNA fractionation, mRNA localization.

Introduction

Membrane-associated, integral membrane and secreted proteins are important in many cellular processes, e.g. cell wall biosynthesis, nutrient uptake, ion flux, signal transduction, cell–cell interactions and defence. Many of these proteins contain an N-terminal secretion signal or signal anchor and are synthesized on the endoplasmic reticulum (ER). The translation of encoding mRNAs starts in the cytoplasm, but as soon as the newly synthesized N-terminal signal peptide (SP) or signal anchor is recognized, translation is halted (Liao *et al.*, 1997; Woolhead *et al.*, 2004). The signal recognition particle (SRP), a ribonucleoprotein complex, subsequently conducts the translating ribosome to the ER membrane. Proteins are then co-translationally translocated into the ER lumen via protein channels, the so-called translocons, or are integrated into the ER membrane (reviewed by Siegel, 1997; Walter and Johnson, 1994). Proteins can remain in the ER or be transported from the ER, often via the Golgi complex, to their final destination.

Membrane association of proteins does not only occur through co-translational insertion into membranes, but

can also take place through interactions with membrane-bound proteins or lipids, glycosylphosphatidylinositol-anchoring or lipid modifications. Furthermore, nuclear-encoded proteins that are targeted to organelles, e.g. chloroplasts and mitochondria, are in most cases translated in the cytosol (reviewed by Jarvis and Robinson, 2004; Rapaport, 2003). Active post-translational targeting processes mediate the import of transit sequence-containing proteins into these organelles.

The experimental demonstration of membrane-association or secretion of proteins is a major challenge, and has so far only been shown for a limited number of proteins. The subcellular localization of the 28 952 putative proteins encoded in the genome of *Arabidopsis thaliana* has been predicted by a variety of algorithms. However, algorithms have their limitations with respect to reliability and comprehensiveness. Large-scale proteomic and immunolocalization studies will provide conclusive evidence on the subcellular localization of most Arabidopsis proteins, but the current technical and biological limitations hamper these analyses on a large scale.

In contrast, the genome-wide analysis of transcripts is common practice nowadays, allowing the analysis of mRNAs that associate with membranes via polysomes (ribosomes bound to mRNA). As a large majority of membrane-associated and secreted proteins are synthesized on these polysomes, analysis of the associated transcripts offers a powerful experimental tool for their genome-wide identification.

A genome-wide analysis of membrane-associated and secreted proteins from yeast and mammalian cells has been described by Diehn *et al.* (2000) using polysomes isolated from membrane and soluble fractions. Microarray analysis of the membrane-associated fraction from yeast cells revealed a maximum enrichment of 11.3 for transcripts encoding proteins containing transmembrane domains (TMDs) and signal peptides (SPs). Interestingly, 285 yeast transcripts were identified encoding proteins not predicted to be membrane-localized, suggesting alternative signals for mRNA membrane association.

Several methods exist for the fractionation of membrane-bound polysomes (Davies and Abe, 1995; Diehn *et al.*, 2000; Larkins, 1986; Mechler, 1987; Stoltenburg *et al.*, 1995; Taliercio and Ray, 2001), often tailored for specific organisms. In this paper, we describe a robust method to isolate intact RNA from membrane-bound polysomes of Arabidopsis seedlings. The relative abundance of mRNA was visualized on CATMAs (complete Arabidopsis transcriptome microarrays) containing gene-specific tags (GSTs; Allemeersch *et al.*, 2005; Hilson *et al.*, 2004). The enrichment of transcripts was verified by prediction algorithms and gene ontologies. Our results indicate that the method allows a significant enrichment of predicted membrane-associated and secreted proteins.

Results

Fractionation of polysomes

For Arabidopsis, there was no optimized protocol available for the isolation of membrane-bound polysomes (MBP). A first step in the construction of a dedicated protocol was the choice of an extraction buffer meeting the major challenges of MBP fractionation. These include maintaining the association of polysomes with membranes, avoiding RNA degradation, preventing artificial polysome aggregation and averting release of ribosomes following polypeptide completion (run-off). This was achieved with a high-salt extraction buffer (pH 8.5), containing an RNase inhibitor able to perform its function at this high pH, and sucrose as an osmotic stabilizer. In addition, it contained a balanced mix of Mg^{2+} (70 mM) and EGTA (50 mM). Excess Mg^{2+} prevents ribosome dissociation and EGTA preferably chelates Ca^{2+}, thereby diminishing the role of Ca^{2+} as a cofactor in RNases. The reducing agent DTT was included to stabilize polysomes, and cycloheximide to prevent ribosome run-off (Larkins, 1986).

The fractionation of membrane-bound and free polysomes was performed using liquid nitrogen-pulverized tissue of 2-week-old above-ground parts of Arabidopsis seedlings. One volume of powdered tissue was slowly added to a half volume of buffer to extract the polysomes. Cellular debris, nuclei, mitochondria and chloroplasts were removed by filtering and several low-speed centrifugation steps. Medium-speed centrifugation of the obtained post-mitochondrial supernatant yielded a membrane pellet containing the crude microsomal fraction, whereas the supernatant contained the crude cytosolic fraction. The membrane pellet was dissolved in extraction buffer containing 1% v/v Triton X-100 to solubilize the membranes, thereby releasing the polysomes. From the supernatant and the solubilized membrane fraction, polysomes were further purified by centrifugation through two discontinuous layers of buffered sucrose (Figure 1), and recovered as a pellet.

The technical reproducibility of the procedure was investigated on three replicate polysomal fractionations of materials taken from one batch of above-ground parts of 2-week-old seedlings. RNA from resuspended polysomal pellets was purified and the ratio between the 18S and 28S ribosomal bands was compared with total RNA from Arabidopsis (Figure 2). The 28S/18S ratios from the MBP and FP fractions were comparable in all three isolations as well as to the ratio observed in total RNA, indicating that ribosomal RNA was kept intact during the isolation procedure. In addition, the three bands representing chloroplast ribosomal RNA, which were present in total RNA, were almost completely absent in the MBP and free polysomal (FP) RNA samples.

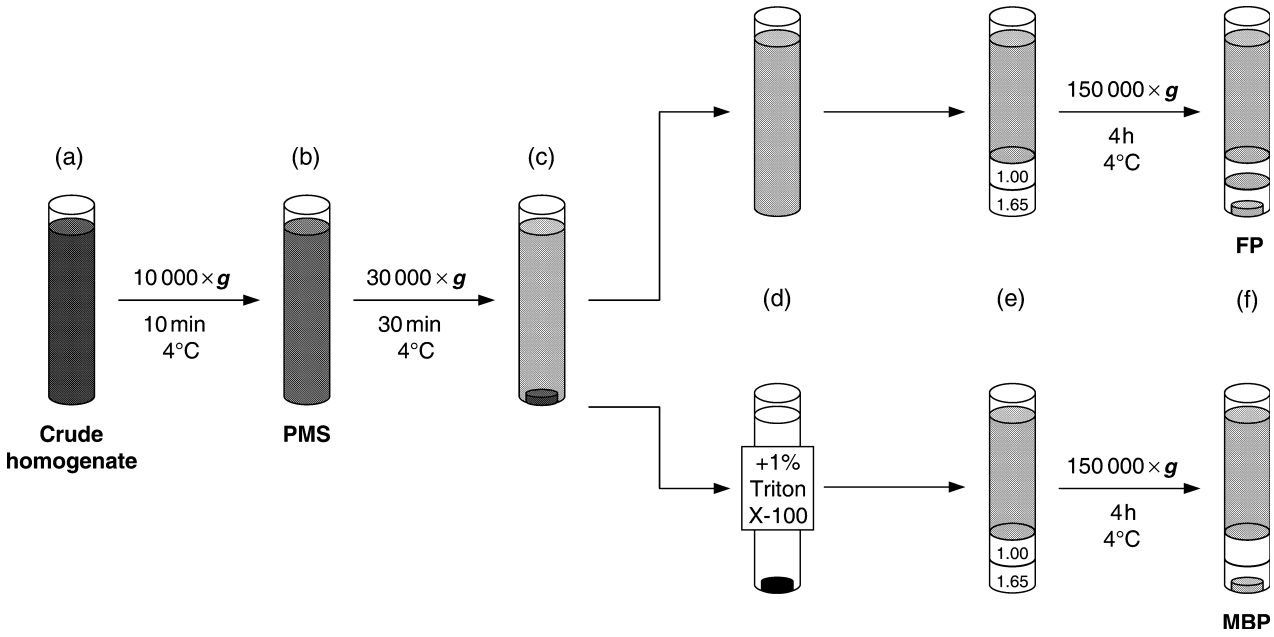

Figure 1. Fractionation of membrane-bound and free polysomes.
(a) One volume of liquid nitrogen-pulverized tissue was slowly added to a half volume of extraction buffer and allowed to thaw on ice. The slurry was strained through two layers of Miracloth, (Calbiochem, San Diego, CA, USA) by centrifugation at 4°C (1 min at 100 ***g*** and 5 min at 1000 ***g***).
(b) The crude homogenate was centrifuged at 10 000 ***g*** for 10 min at 4°C to remove additional cellular debris, nuclei, chloroplasts and mitochondria, yielding the post-mitochondrial supernatant (PMS).
(c) Membranes were pelleted by centrifugation at 30 000 ***g*** for 30 min at 4°C and (d) solubilized in 1% v/v Triton X-100. The solubilized membrane fraction, containing membrane-bound polysomes (MBP), and the cytoplasmic fraction, containing free polysomes (FP), were centrifuged at 30 000 ***g*** for 30 min at 4°C to pellet residual membranes.
(e) Both crude fractions were loaded onto separate discontinuous sucrose layers (1.00 and 1.65 M) and centrifuged at 150 000 ***g*** for 4 h at 4°C.
(f) The interphase between 1.00 and 1.65 M sucrose in the FP gradient contained residual membranes which were discarded. The gradients were carefully aspired and the polysomal pellets were resuspended in extraction buffer.

We inspected the integrity of the isolated polysomes and compared it with the profile obtained from total polysomes. The polysomal pellets were resuspended and loaded onto sucrose gradients (20–60% w/v), and after ultracentrifugation the absorbance was measured from bottom to top of the gradients (Kawaguchi *et al.*, 2004; Figure 3). All three profiles revealed that the polysomal fractions consist of heavy complexes that were recovered from the lower part of the gradients. Interestingly, none of the gradients showed significant absorbance in the top part, indicating that the polysomal fractions were deprived of free RNA and monosomes and that the polysomal complexes were stable in the employed buffers. The profiles obtained from total and free polysomes are very similar in shape, which was expected because only membranes were separated from the 'free' fraction whereas the rest of the procedure was identical. Interestingly, the MBP profile appears slightly shifted towards the lower region of the gradient when compared with the FP and total polysomal profiles. This phenomenon could be due to generally larger membrane-associated transcripts when compared with 'free' transcripts, giving them a higher capacity of ribosome loading. Alternatively, the membrane-associated polysomal complexes could contain additional features making them heavier than 'free' polysomal complexes.

Overall, the analysis of the 28S/18S ratio of isolated RNA from polysomal pellets, the intactness of polysomal complexes and deprivation of free RNA and monosomes shows that our method is successful in isolating free and membrane-bound polysomes.

mRNA profiling

The relative abundance of transcripts in the MBP and FP fraction of each isolation was analysed by hybridization of labelled cDNA on CATMA version 2 arrays. mRNA from the fractionated polysomes was amplified, and modified cDNA was synthesized by the incorporation of 5-(3-aminoallyl)-dUTP which was subsequently labelled with either Cy3 or Cy5 mono-reactive dye. Each isolation of MBP and FP was hybridized on four CATMA arrays including a dye swap on two arrays. From the scanned arrays, spot intensities were determined and subsequently normalized. Fitted least-squares averages of *M*-values [$\log_2$(MBP/FP)] of four slides for each fractionation were calculated and applied to determine the Pearson coefficient of correlation between the

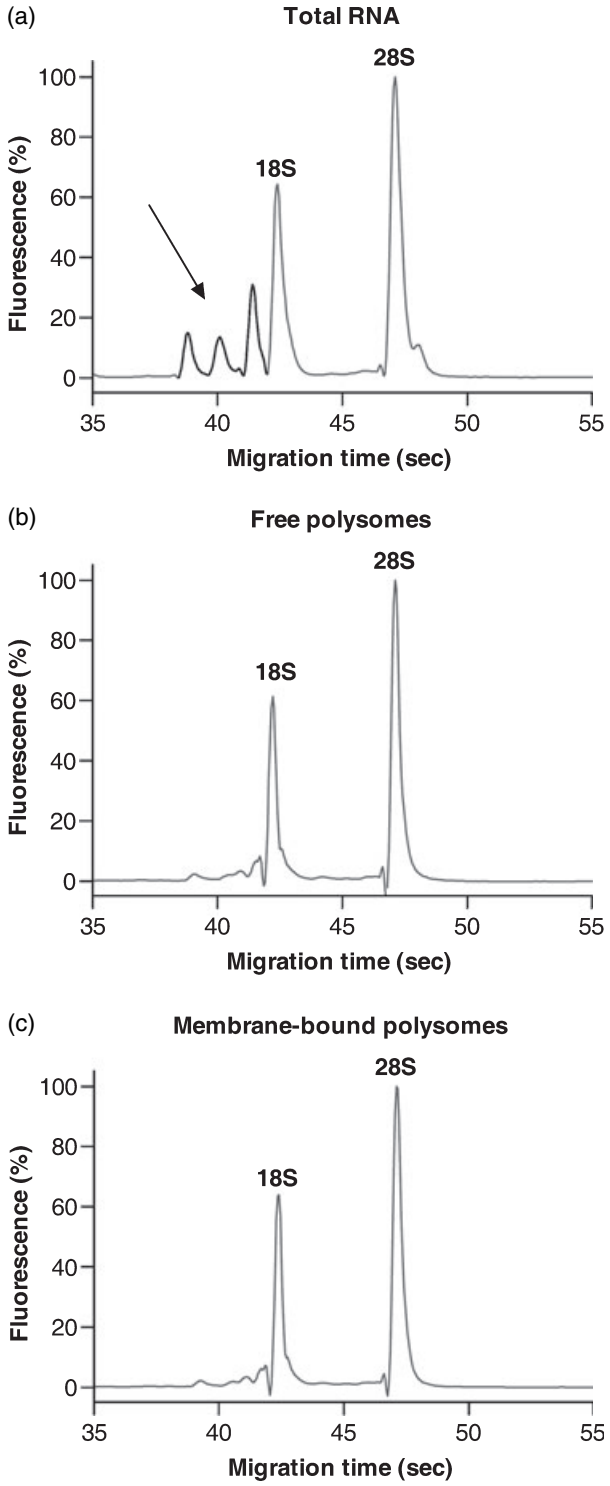

Figure 2. RNA profiles of MBP and FP compared with total RNA.
(a) Total RNA from above-ground parts of seedlings.
(b) RNA from a representative FP isolation.
(c) RNA from a representative MBP isolation. Notice that the 28S/18S ratio in the FP and the MBP fractions is comparable with the ratio observed in total RNA, indicating that no degradation occurred during the isolation. The three ribosomal RNA peaks from the chloroplasts present in total RNA (indicated by the arrow) were efficiently removed from the polysomal fractions.

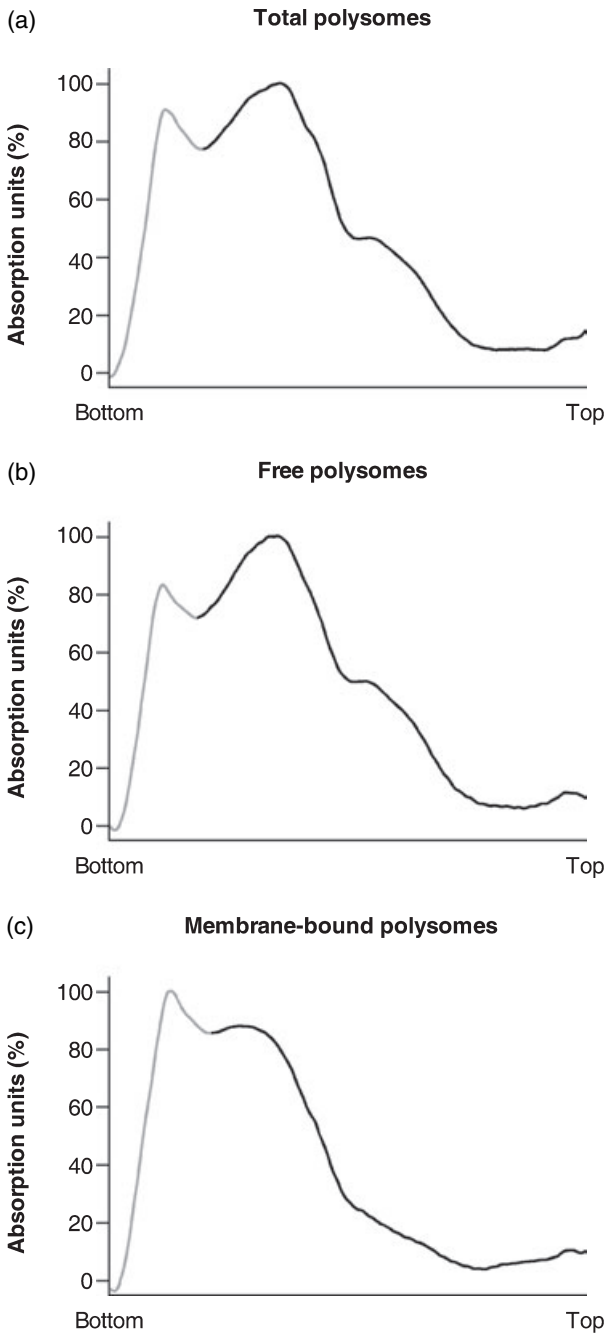

Figure 3. Integrity of isolated polysomes
Total polysomes, MBP and FP were isolated and loaded on sucrose density gradients (20–60% w/v). Gradients were subjected to centrifugation at 275 000 ***g*** for 90 min at 4°C and absorbance was measured from bottom to top at 260 and 320 nm. The 320 nm profile was subtracted from the 260 nm profile and background was corrected. The plots are displayed as relative percentage of the highest observed value for each size distribution. All three profiles show that the polysomal complexes are mainly recovered near the lower third part of the gradients. No significant absorption was observed at the top one-third section, indicating that the purified polysomal pellets consist mainly of heavy complexes. The absorbance values indicated by the grey lines show the accumulation of heavy complexes near the bottom of the gradients.

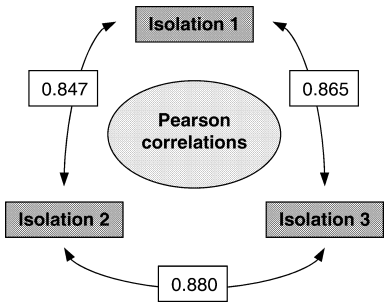

Figure 4. Pearson correlations between three polysomal isolations.
The correlations between the three isolations were calculated from the fitted *M*-values (least-squares averages) of four microarrays for each isolation.

three experiments (Figure 4). A correlation of approximately 0.86 between the fractionations was found, indicating that the method provides a high level of reproducibility.

The high Pearson correlations between the three isolations allowed the combined analysis of all 12 arrays into one reliable dataset. The acquired data were corrected for multiple testing errors using the false discovery rate (FDR) approach, yielding *q*-values (Storey and Tibshirani, 2003). The dataset was chosen such that the number of false discoveries was as low as 0.001, yielding the approximate finding of less than five false positives in the selected dataset of 4714 transcripts for the MBP and FP fraction. Final datasets were selected based on *q*-values equal to or lower than 0.001 and a cut-off of 1.25 times enrichment. The number of significantly enriched transcripts was 3270 for the MBP fraction ($M > 0$) and 1325 for the FP fraction ($M < 0$; Figure 5) from a total of 24 411 GSTs on the CATMA array.

The microarray data was verified by quantitative PCR (qPCR) on three highly enriched transcripts from either the MBP or the FP fraction (Figure 5). RNA from the three MBP and FP isolations was precisely quantified and equal amounts were used as input for the cDNA synthesis. Each transcript was measured in triplicate in all fractions. This allowed the precise determination of the cycle thresholds (C_T), which were compared in a histogram to visualize the enrichment of the three MBP and FP isolations (Figure 6). A common reference, e.g. *ACTIN*, was not included as it is unlikely that such a transcript would be equally present in the MBP and FP fractions. The qPCR data clearly indicate that the enrichment is equal between the three MBP and FP isolations, supporting the calculated Pearson correlations. The differences in C_T values between the MBP and FP

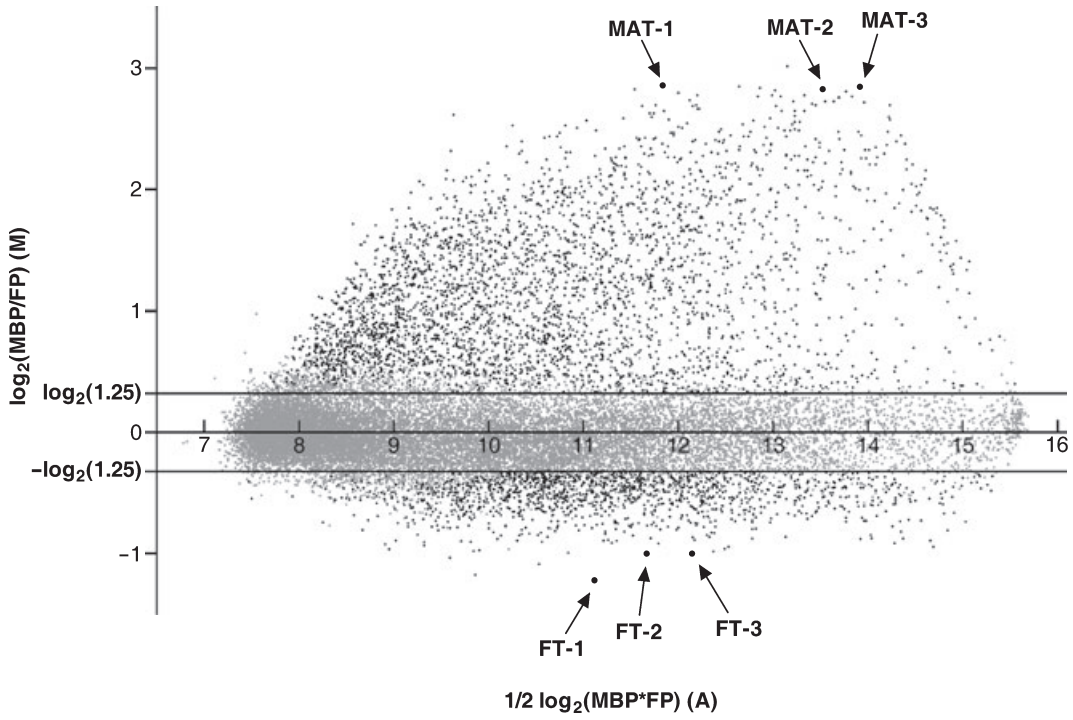

Figure 5. M.A plot of MBP versus FP.
One combined data set was created of 12 arrays representing the relative abundance of (MBP transcripts ($M > 0$) and FP transcripts ($M < 0$) in a total of 24 411 GSTs. Selections were collectively based on significance ($q \leq 0.001$) and relative abundance (≥1.25-fold enrichment for MBP and FP respectively). The black spots represent the selected transcripts in each fraction (3270 for MBP and 1325 for FP). Transcripts were selected from the top region of the graph [membrane-associated transcript, MAT-1 (At1g09750, eukaryotic aspartyl protease, homologous to chloroplast nucleoid DNA-binding protein from *Nicotiana tabacum*), MAT-2 (At5g23660, nodulin MtN3 family protein) and MAT-3 (At1g49750, leucine-rich repeat family protein)] and the bottom region [free transcript, FT-1 (At2g22122, expressed protein), FT-2 (At3g07910, expressed protein) and FT-3 (At5g12140, cysteine protease inhibitor)] for the verification of enrichment in each fractionation by qPCR.

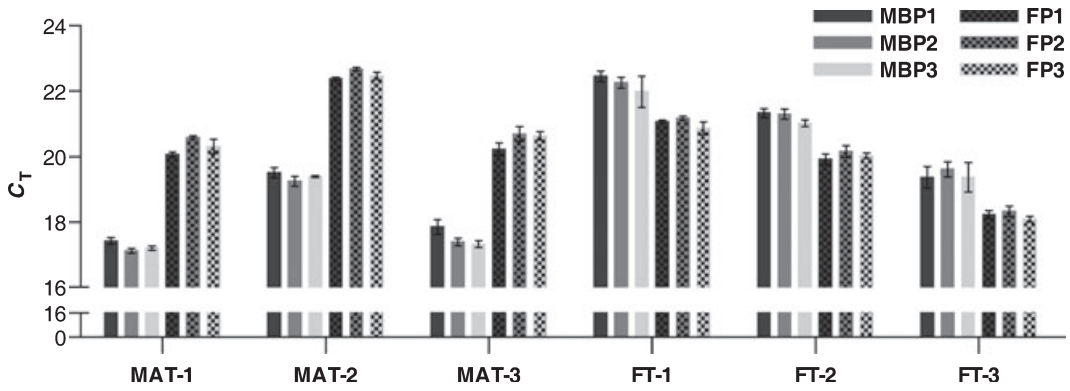

Figure 6. qPCR analysis of six selected transcripts in MBP and FP RNA samples.
Three highly enriched transcripts from the MBP fraction and three from the FP fraction were selected based on the data from the M–A plot. qPCRs were performed in triplicate on cDNA made from normalized input RNA of three fractionations (3 × MBP and 3 × FP). The cycle thresholds (C_T), with calculated error bars representing the 95% confidence interval, clearly indicate that the membrane-associated transcripts (MATs) are enriched in the MBP fractions and the free transcripts (FTs) in the FP fractions. Note that higher bars represent lower abundance of transcripts. The difference in C_T values between the MBP and FP isolations shows that the MBP fractions are approximately three PCR cycles (approximately eightfold) enriched and the FP fractions one PCR cycle (approximately twofold) enriched for the selected transcripts, which is in agreement with the values observed in the M–A plot (Figure 5).

isolations is in good agreement with the data visualized in the M.A plot (Figure 5), suggesting that normalization of the microarray data did not create artefacts. In addition, it shows that amplification of mRNA did not alter the relative abundance of the selected transcripts.

Bioinformatics analysis of membrane-associated transcripts

Data conversion. CATMA arrays contain GSTs designed on gene models predicted by the eukaryotic gene finder EuGène (Schiex *et al.*, 2001) for the Arabidopsis genome sequence of January 2001 (TIGR, The Institute for Genome Research, ftp://ftp.tigr.org/pub/data/a_thaliana/ath1/). However, EuGène occasionally predicts two genes where TIGR only predicts one (Pavy *et al.*, 1999). This gave several instances were two CATMA GSTs were designed for a single gene. In addition, the EuGène and TIGR annotations contain multiple instances of non-overlapping predictions. Because the TIGR annotation is more commonly used than EuGène, we converted CATMA-IDs (identification numbers) to AGI (Arabidopsis Genome Initiative)-IDs. To prevent double counting due to this conversion, we only utilized transcripts with a strict one-to-one relationship between CATMA IDs and AGIIDs; all others were omitted from further analysis. This one-to-one relationship between TIGR and EuGène gives a better likelihood that a certain annotation is correct. Finally, this set of AGI IDs was checked against the TIGR 5.0 annotation of January 2004, and if not present AGI IDs were discarded. As the discarded CATMA IDs (5702 in total) were solely selected based on their incompatibility with the TIGR 5.0 annotation, the remaining 18 709 AGI IDs remain representative for the Arabidopsis genome.

The obtained unambiguous AGI IDs were employed to analyse the cellular component (i.e. localization) of the proteins with gene ontologies. In addition, they were also converted to protein-IDs to assess the prediction of TMDs and targeting of the encoded proteins. Each prediction was given a weighted score to correct for double counting of different gene models (protein-IDs). This weighting method led to a total weight of the chosen datasets corresponding exactly to the number of unambiguous AGI-IDs, thereby allowing proper calculations of chi-squared values and the corresponding *P*-values. The flowchart for the conversion of CATMA IDs to AGI IDs and subsequently protein-IDs is given in Figure 7.

Prediction algorithms. Proteins corresponding to genes in the combined dataset of 12 arrays were analysed with prediction algorithms for SPs and TMDs and categorized based on gene ontologies to classify putative membrane-associated and secreted proteins.

The amino acid sequences of the protein-IDs in the MBP, FP and CATMA selections were investigated by TARGETP (Emanuelsson *et al.*, 2000). TARGETP is one of the prediction algorithms of choice because it discriminates between proteins targeted to the mitochondrion, the chloroplast and the secretory pathway. The relative abundances of protein-IDs in these predicted classes were calculated for the selections and are represented in a cumulative histogram (Figure 8a). The histogram shows that the MBP fraction is strongly enriched for secretory pathway proteins, whereas occurrence of mitochondrion- and chloroplast-targeted proteins is diminished. In contrast, the FP fraction is reduced for secretory pathway proteins and slightly enriched for chloroplast- and mitochondrion-targeted proteins. Interestingly, there are still predicted chloroplast and mitochondrion proteins present in the MBP set.

We employed a moving average analysis to determine the proper cut-off where the percentage of observed predicted

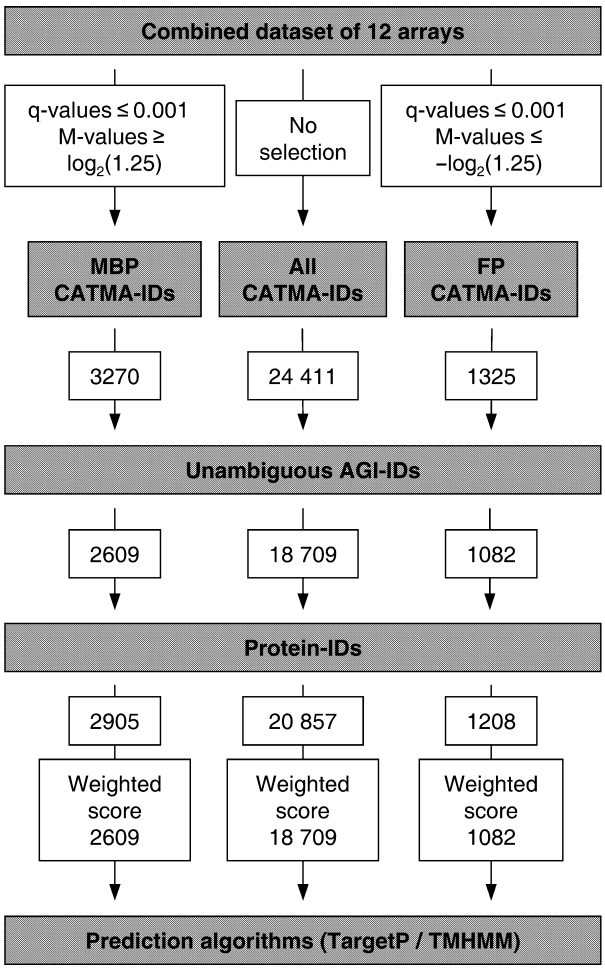

Figure 7. Flow chart for the data conversion of CATMA, MBP and FP datasets.
The combined dataset of 12 arrays was employed to select MBP, FP and all transcripts present on the array. Selections for the MBP and FP fractions were made as in the M–A plot (Figure 5). CATMA-IDs for each selection were converted to unambiguous AGI-IDs and protein-IDs (gene models). The higher number of protein-IDs is due to recurrence of multiple gene models. Therefore, protein-IDs were given a weighted score to correct for double counting.

features was higher than chance. A moving average analysis is used to eliminate short-term fluctuations and highlight trends. Each data point in our analysis represents the mean of predicted features in a window of 500 protein-IDs from a total of 20 857 protein-IDs. The protein-IDs were weighted to correspond to the number of AGI-IDs identified in the window. The moving average analysis in Figure 8(c) shows that the percentage of predicted secretory pathway proteins starts to be higher than chance at a $\log_2$(MBP/FP) value of approximately 0.322, which is equal to a ratio of 1.25, compared with the overall percentage. Similar results were obtained for the enrichment of predicted TMD-containing proteins (Figure 8d). This value was in good agreement with the lower level of enrichment for the selection of the MBP and FP datasets as used in Figures 5 and 7.

To determine the significance of the observed enrichment in the MBP and FP fractions, the chi-squared test for equality of proportions was performed (Table 1). The significance of enrichment for a particular cellular localization was calculated in a 2×2 contingency table containing the number of expected genes versus observed genes in a certain class and genes not belonging to this class, giving one degree of freedom. Table 1 shows that the MBP fraction is significantly enriched for predicted secretory pathway proteins, whereas the occurrence of predicted chloroplast- and mitochondrion-targeted proteins is reduced. For the FP fraction the opposite is the case, although with a lower significance compared with the MBP fraction. This lower significance is probably due to the smaller size of the dataset and lower level of enrichment.

To determine whether there was also a substantial enrichment for genes encoding membrane proteins in the MBP fraction, the selected datasets were also screened for proteins containing TMDs. The prediction algorithm TMHMM version 2.0 (Krogh *et al.*, 2001) was employed for this study as it has been shown to be one of the best TMD prediction algorithms available (Moller *et al.*, 2001). Proteins were classified into groups with 0, 1–3, and ≥ 4 TMDs. The cumulative histogram (Figure 8b) shows a clear enrichment of predicted TMD-containing proteins in the MBP fraction, whereas the occurrence of these proteins is notably lower in the FP fraction.

The significance of TMD enrichment in classes was calculated from chi-squared values (Table 1) as for the TARGETP analysis, and shows again a higher significance in the MBP fraction. This fraction is reduced for proteins predicted to contain no transmembrane domains, whereas proteins with predicted TMDs (classified 1–3 and ≥ 4) are enriched. In contrast, the appearance of proteins with predicted TMDs is reduced in the FP fraction and proteins without TMDs are enriched.

In the MBP fraction, a high occurrence of predicted SP- and TMD-containing proteins (43% and 54% respectively) was observed (Figure 8). As these two sets show considerable overlap, the percentage of protein-IDs containing at least one TMD) or a secretory pathway prediction without TMDs was determined to be 32% for the CATMA set, 69% for the MBP and 19% for the FP. This shows that in the MBP fraction the percentage of membrane-associated transcripts is more than doubled.

In our analysis, two prediction algorithms were used to determine putative membrane association. However, prediction algorithms are often in disagreement with each other. An alternative approach is to apply the 'majority vote' concept, giving a consensus prediction based on different algorithms (Nilsson *et al.*, 2000). This concept has been applied for Arabidopsis, *Oryza sativa* and other seed plants and is publicly available through the ARAMEMNON data-

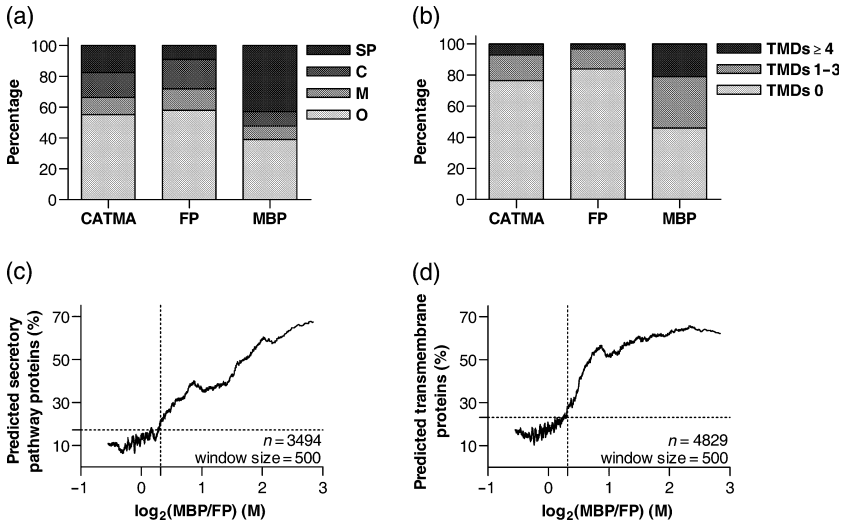

Figure 8. Enrichment of mRNAs encoding predicted transmembrane domain-containing and secreted proteins.
(a) Subcellular localization predicted by TARGETP. The predicted subcellular localizations (S, secretory pathway; C, chloroplast; M, mitochondrion; O, other) of the identified weighted protein-IDs in the FP and MBP fractions were compared with the predicted localization of the proteins on the CATMA array.
(b) As in (a) for the classes of predicted transmembrane domain-containing proteins by TMHMM.
(c) Moving average analysis of predicted secretory pathway proteins (*n*) in the selected CATMA set. Protein-IDs and occurrence of signal peptides (SPs) were given weighted scores to prevent double counting in the given window size. The dashed horizontal line represents the overall percentage of predicted secretory pathway proteins, whereas the vertical line represents a cut-off of $\log_2$(MBP/FP) of approximately 0.322, which is equal to a ratio of 1.25.
(d) Moving average analysis as in (c) for the occurrence of TMD-containing proteins.

Table 1 Significance of predicted subcellular localization classes by TARGETP and of TMD classes by TMHMM

	Expected	Observed	χ^2	*P*
Predicted subcellular localization				
MBP dataset of 2905 weighted protein-IDs (2609 AGI IDs)				
Secretory pathway	450	1121	410.05	<<0.001
Other	1437	1018	135.05	<<0.001
Chloroplast	429	242	59.81	<<0.001
Mitochondrion	293	228	8.98	0.003
Total	2609	2609		
FP dataset of 1208 weighted protein-IDs (1082 AGI IDs)				
Secretory pathway	187	98	32.01	<0.001
Other	596	626	1.69	0.193
Chloroplast	178	207	2.66	0.103
Mitochondrion	121	151	3.79	0.052
Total	1082	1082		
Predicted TM domains				
MBP dataset of 2905 weighted protein-IDs (2609 AGI IDs)				
0	1995	1194	517.41	<<0.001
1–3	426	862	195.96	<<0.001
≥4	188	553	209.55	<<0.001
Total	2609	2609		
FP dataset of 1208 weighted protein-IDs (1082 AGI IDs)				
0	827	908	23.40	<0.001
1–3	177	139	7.98	0.024
≥4	78	35	16.73	<0.001
Total	1082	1082		

Weighted protein IDs corresponding to MBP and FP transcripts were investigated for their subcellular localization, and corresponding chi-squared- and *P*-values were calculated between each expected and observed set for a given class.

base (http://aramemnon.botanik.uni.koeln.de; Schwacke *et al.*, 2003). To date, ARAMEMNON (release 3.0) accommodates consensus data from 16 TM domain, nine signal peptide and several other prediction algorithms with respect to beta sheets and glycosylphosphatidylinositol anchors. Currently, 6053 AGI-IDs are predicted to encode for 'putative' membrane and membrane-associated proteins. Of these, 4376 AGI-IDs are unambiguously present on the CATMA array and represent 23% of the 18 709 AGI-IDs. We recovered 1252 'ARAMEMNON' AGI-IDs in the MBP dataset of 2609 AGI-IDs (48%) and 187 'ARAMEMNON' AGI-IDs in the FP dataset of 1082 AGI-IDs (17%). The more than twofold enrichment of membrane-associated transcripts in the MBP fraction (approximately 2.1-fold) given by ARAMEMNON is in excellent agreement with that obtained with the individually tested algorithms, TARGETP and TMHMM (approximately 2.2-fold).

Gene ontologies. The prediction algorithm analyses showed that the MBP fraction is enriched for predicted secretory pathway and TMD-containing proteins. To obtain additional information about the localization of proteins, we investigated the available gene ontologies. The structured and controlled vocabulary of an ontology provides representations of knowledge that are excellently suited to investigate large datasets. Such an ontology has been developed over the past few years by the Gene Ontology (GO) consortium (http://www.geneontology.org/; Ashburner *et al.*, 2000). TAIR (The Arabiposis Information Resource, http://

www.arabidopsis.org/) is a member of this consortium and actively contributes to the development of the three organizing principles of GO (biological process, molecular function and cellular component) for Arabidopsis. The last organizing principle refers to the subcellular localization of a gene product and makes it well suited to investigate the properties of the MBP and FP fractions. The GOslim terms (http://www.arabidopsis.org/), which represent the higher-level GO entries, were examined for AGI-IDs of each dataset (CATMA, MBP and FP) and were given weighted scores to prevent double counting. Cellular components accommodate 16 classes that are shown in Table 2 for the AGI-IDs of CATMA, MBP and FP fractions with their corresponding chi-squared- and *P*-values. These values show that the MBP fraction is significantly enriched for proteins classified 'other membranes', whereas 'cellular component unknown', 'chloroplast', 'mitochondria' and 'nucleus' classes are reduced. In contrast to the MBP fraction, the FP fraction is significantly diminished for proteins classified 'other membranes' and enriched for 'ribosome' proteins. These data are consistent with the analysis described above for the prediction algorithms TARGETP and TMHMM. The fact that many cellular components show no significant enrichment is probably due to the low numbers of classified AGI-IDs in these higher-level GO entries.

Discussion

A method for isolating membrane-bound polysomal RNA from Arabidopsis was established and validated by DNA microarray analysis. Polysomes could be efficiently isolated by solubilization from microsomal fractions derived from Arabidopsis seedlings. Separation from free RNA was achieved by sucrose gradient centrifugation allowing the

Table 2 Cellular components by gene ontologies for the selected datasets

Cellular component	Expected	Observed	χ^2	*P*
MBP dataset of 2724 weighted protein-IDs (2469 AGI-IDs)				
Other membranes	567	1285	442.54	<<0.001
Cellular component unknown	769	445	114.26	<<0.001
Chloroplast	296	152	50.85	<<0.001
Nucleus	202	96	40.10	<<0.001
Mitochondria	258	178	16.08	<0.001
Ribosome	34	15	7.44	0.006
Other intracellular components	108	76	5.78	0.016
Other cytoplasmic components	63	39	5.77	0.016
ER	8	18	3.87	0.049
Plasma membrane	10	17	1.83	0.177
Cytosol	8	4	1.34	0.248
Cell wall	8	13	1.20	0.274
Plastid	5	2	1.29	NR
Extracellular	32	41	1.13	0.289
Other cellular components	125	111	0.87	0.350
Golgi apparatus	3	4	0.14	NR
Total	2469	2496		
FP dataset of 1105 weighted protein-IDs (1006 AGI-IDs)				
Other membranes	229	129	33.97	<0.001
Cellular component unknown	310	324	0.45	0.502
Chloroplast	120	161	6.58	0.010
Nucleus	81	67	1.44	0.229
Mitochondria	104	136	4.85	0.028
Ribosome	14	45	16.81	<0.001
Other intracellular components	44	57	1.76	0.184
Other cytoplasmic components	25	22	0.20	0.658
ER	3	2	0.20	NR
Plasma membrane	4	3	0.14	NR
Cytosol	3	4	0.14	NR
Cell wall	3	1	1.00	NR
Plastid	2	1	0.33	NR
Extracellular	13	6	2.60	0.107
Other cellular components	50	46	0.18	0.676
Golgi apparatus	1	3	1.00	NR
Total	1006	1006		

Presented are the higher-level categories of cellular components in each fraction compared with the expected values, with their corresponding chi-squared- and *P*-values. Note that frequencies below five were scored as not reliable (NR).

pelleting of polysomes from the Triton X-100-treated membrane fraction. The buffer conditions prevented RNA degradation as demonstrated by the intactness of ribosomal RNA and maintained integrity of the polysomes. The reproducibility of the method was shown by RNA profiling of three replicate isolations and calculated Pearson correlations between them, based on the hybridization signals of all genes on the CATMA arrays. Profiling of free and membrane-bound polysome-associated mRNA on CATMA microarrays allowed the genome-wide analysis of transcripts enriched for the respective fractions. The lower level of enrichment was set to 1.25 based on the statistical reliability of the signal intensities (q-values ≤ 0.001) and the moving average analyses. Transcripts encoding proteins predicted to contain TMDs and SPs were highly enriched in the MBP fraction, whereas their occurrence was clearly reduced in the FP fraction. In general, the observed significance of enrichment for proteins predicted to contain TMDs and SPs in the latter fraction was lower compared with the MBP fraction. This is probably due to the smaller dataset resulting from the overall lower level of enrichment in the FP fraction. Although the identification of transcripts from the FP fraction appears to be less reliable, the MBP-enriched dataset is robust, as indicated by the detection of the large number of proteins predicted to contain TMDs and/or SPs.

Of the 2609 MBP-enriched transcripts, approximately 69% were predicted to encode secretory pathway and/or TMD-containing proteins. Although the remaining 31% of transcripts (824) were not predicted to be translated on the ER, they were enriched in the MBP fraction. The ARAMEMNON dataset revealed that the 'majority vote' concept overruled the prediction of TARGETP and TMHMM for 69 proteins in this set, giving 755 unexpected predictions. We investigated the relative enrichment of these transcripts in the MBP fraction by calculating the average coordinates of the MBP data points and the unexpected transcripts in this set. The M–A coordinates were (1.17, 10.7) for the MBP set and (0.93, 10.7) for the unexpected transcripts. The unexpected transcripts are slightly over-represented in the lower regions of enrichment as expected, because the reliability of recovering 'true' membrane-associated transcripts becomes less in these areas. This indicates that a higher enrichment value gives a higher chance of recovering a known or predicted membrane-associated transcript. This is also evident from the moving average analyses (Figure 8), where the percentage of predicted features (SPs or TMDs) is generally higher as the enrichment factor increases.

The fact that these unknown mRNAs are enriched in the MBP fraction may indicate that the employed prediction algorithms are unable to detect localization features on these proteins. Another possibility is that some transcripts encoding chloroplast- and mitochondrion-targeted proteins traverse via the ER to their destination, suggesting distinct import pathways. Interestingly, recent advances in chloroplast proteome research suggest that 49 out of 604 nuclear-encoded proteins present in this organelle are predicted to contain SPs for ER translocation (reviewed by Jarvis, 2004; Kleffmann *et al.*, 2004). On the other hand, there is evidence that mRNAs encoding soluble proteins can at the same time be translated on, and in some cases even partition to the ER (Lerner *et al.*, 2003). This phenomenon may explain some of the overlap between the MBP and FP populations surrounding $M = 0$ (Figure 5), and gives a possible explanation for the recovery of unexpected transcripts in the MBP-enriched fraction.

Many studies have shed light on the widespread phenomenon of mRNA localization. For example, in animal and yeast cells, it has been shown that the recognition of several transcripts is mediated by *trans*-acting factors that identify *cis*-acting signals called zip codes (Oleynikov and Singer, 1998; Singer, 1993), which are localized mostly in the 3′-untranslated region of the transcript. Binding of the *trans*-acting factor(s) to the *cis*-element(s) leads to the formation of an mRNA ribonucleoprotein particle (mRNP). This transport particle can associate with microtubule or microfilament tracks that will direct it to its destination, mostly via active transport (Wilhelm and Vale, 1993). Other ways to create asymmetric mRNA distributions are by localized anchoring, diffusion and subsequent localized mRNA degradation/protection (reviewed by Kloc *et al.*, 2002). A prerequisite to creating these asymmetric mRNA distributions is translational arrest. Several repressors are known that bind to the 3′-untranslated region of the mRNA and suppress translation, thereby allowing the formation of asymmetric mRNA distributions (reviewed by Crofts *et al.*, 2004). It could be that several of the unexpected transcripts are arrested transcripts at the ER surface or other cell membranes and therefore enriched in our MBP fraction.

Interestingly, 301 transcripts from the 755 unexpected transcripts were classified as 'cellular component unknown' by gene ontologies. In addition, 334 AGI-IDs of the 755 embody a 'hypothetical', 'expressed' or 'putative' description, illustrating the limited knowledge concerning these Arabidopsis genes. The reason why these transcripts identified in our screen are associated with membranes is still unclear, but it could involve novel associations due to protein–protein interactions, translational regulation, protein modifications, mRNA interactions, and as yet unidentified targeting signals that are so far not detected by current prediction algorithms. We anticipate that future research including proteomics and tagging approaches will provide data on the localization of these transcripts and their encoded proteins.

Comparison of our findings with the research performed by Diehn *et al.* (2000) on yeast and mammalian cells showed that our presented method allowed a similar enrichment for MBP transcripts. In contrast, our observed enrichment for the FP fraction was less strong, which is probably caused by methodological differences. Their method was based on

isopycnic centrifugation, i.e. allowing membrane vesicles to float, giving a membrane-deprived fraction still containing free mRNA. In contrast, the method described in this paper removes free RNA from the FP fraction. It seems that both methods allow the enrichment of membrane-associated transcripts, but both have their limitations with respect to FP isolations. We also attempted to optimize their method for polysomal fractionation in Arabidopis, but we were unable to isolate sufficient amounts of RNA from floating membrane fractions. This may be caused by the limited amount of tissue that can be processed with their method or because the conditions are not appropriate to maintain the association of polysomes to membranes in Arabidopsis.

The use of mRNA from polysomes has an advantage compared with total mRNA. Total mRNA cannot account properly for the translated mRNA fraction (for examples, see Gibon *et al.*, 2004), but polysomal isolation is able to overcome this drawback, at least partially, as polysomes are in the final translation stage of mRNA. Polysomes integrate all events contributing to (localized) protein synthesis, such as transcription, maturation, export, stability, translation and partially localization. This has been demonstrated for individual protein levels compared with transcripts in polysomal isolations (Pradet-Balade *et al.*, 2001), but large correlation studies between proteome-wide and transcriptome-wide studies have not been published. This is mainly due to the lack of quantitative proteomics data. Profiling analyses of polysomes could provide researchers with genome-wide information and may partially contribute towards proteome characterization.

Overall, we have shown that the fractionation method described here is well suited to investigate transcripts encoding membrane and secreted proteins on a genome-wide scale. The analysis of membrane-bound polysomal transcripts can be extended to different tissues, growth conditions, and biotic/abiotic treatments to provide an additional level of information on the Arabidopsis transcriptome that will further aid the functional annotation of the Arabidopsis genome.

Experimental procedures

Growth of Arabidopsis

Plants were grown on potting soil (Primasta, Asten, The Netherlands) in a growth chamber (Snijders Scientific, Tilburg, The Netherlands) at 22°C with 16 h of light (approximately 100 μmol photons m^{-2} sec^{-1}) and a relative humidity of 70%.

Fractionation of membrane-bound polysomes (MBP) and free polysomes (FP)

Above-ground parts of 2-week-old seedlings (accession Col-0) were pulverized in liquid nitrogen. A volume of approximately 20 ml of tissue powder was slowly added to 10 ml extraction buffer (pH 8.5) containing 0.2 M Tris, 0.1 M KCl, 70 mM Mg-acetate, 50 mM EGTA and 0.25 M sucrose. The extraction buffer was autoclaved, and 10 mM DTT, 100 μg ml^{-1} cycloheximide (Sigma-Aldrich, St Louis, MO, USA) and 50 U ml^{-1} SUPERase·In (Ambion, Austin, TX, USA) were added before use. The slurry was allowed to thaw on ice and gently mixed with a spatula to enhance homogenization. The suspension was strained quickly through two layers of Miracloth (Calbiochem, San Diego, CA, USA) via successive centrifugation for 1 min at 100 ***g*** (4°C) and 5 min at 1000 ***g*** (4°C) in 50 ml sterile tubes in a benchtop centrifuge (Eppendorf AG, Hamburg, Germany; type 5804R). The cleared supernatant was centrifuged for 5 min at 10 000 ***g*** (4°C) to pellet additional debris. Membranes were pelleted by centrifugation at 30 000 ***g*** in Thickwall SW28 tubes (Beckman Coulter, Fullerton, CA, USA) for 30 min (4°C). The supernatant containing the FP was kept on ice and the membrane pellet was carefully resuspended in extraction buffer containing 1% v/v Triton X-100 to solubilize membranes. Both fractions went through an additional centrifugation step at 30 000 ***g*** under the conditions described above to pellet remaining membranes. Both crude MBP and FP supernatants were loaded on separate discontinuous sucrose gradients in SW41 Ultra-Clear tubes (Beckman Coulter). The gradients consisted of 0.75 ml 1.65 M sucrose and 1.25 ml 1.00 M sucrose both buffered with 10x concentrated sucrose salts buffer (pH 8.5, autoclaved) containing 0.4 M Tris, 1 M KCl, 0.3 M Mg-acetate and 50 mM EGTA. To maintain RNA integrity, 10 mM DTT, 100 μg ml^{-1} cycloheximide and 50 U ml^{-1} SUPERase·In (Ambion) were added to these sucrose solutions. The gradients were centrifuged for 4 h at 150 000 ***g*** (4°C) in an ultracentrifuge (Beckman Coulter; type L8–50 M/E) to pellet the polysomes, and carefully aspired after the run. The polysomal pellets were resuspended in 500 μl extraction buffer and stored at −80°C.

RNA isolation

RNA in the resuspended polysomal pellets was isolated with a RNeasy kit (Qiagen, Venlo, The Netherlands) and treated with the RNase-free DNase set (Qiagen) yielding between 50 and 70 μg of polysomal RNA per isolation. The quantity of RNA was investigated using a UVmini-1240 spectrophotometer (Shimadzu, Kyoto, Japan) and the quality with a bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA) using the RNA 6000 Nano Assay kit (Agilent Technologies).

Sucrose density gradient absorbance profile analysis

Polysomes were isolated as described above. Total polysomes were isolated from the post-mitochondrial supernatant made to 1% v/v Triton X-100. Residual membranes were pelleted, and polysomes in the supernatant were purified through discontinuous sucrose gradients as described above. Density sucrose gradients were made in SW50 Ultra-Clear tubes (Beckman Coulter) with the following volumes: 1.0 ml (60%), 1.5 ml (45%), 1.5 ml (30%) and 0.75 ml (20%) sucrose buffered in 1x sucrose salts buffer as described above. Each layer was frozen in liquid nitrogen before the next one was applied. Gradients were allowed to set for 1 h at 37°C and cooled for 1 h at 4°C (Kawaguchi *et al.*, 2004). Approximately 200 μg of polysomes for each isolation was added to 400 μl extraction buffer (as described above) and loaded onto separate gradients. The gradients were subjected to ultracentrifugation at 275 000 ***g*** for 90 min (4°C) for size separation of the polysomes. Gradients were drained at 0.5 ml min^{-1} with a Pharmacia LKB pump P-1 (Pfizer, New York, NY, USA), and absorbance units (mAU) at 260 and 320 nm were calculated from a moving average time frame of 1.28 sec using an

ÄKTA Design Monitor UV-900 (Amersham Biosciences, Piscataway, NJ, USA). The obtained data were analysed with the Unicorn software package version 3.21.02 (Amersham Biosciences). All 320 nm readings were subtracted from their corresponding 260 nm measurements. Background readings from an empty gradient were subtracted from the absorbance values and plotted as relative percentages.

CATMA arrays

Microarray analysis was performed with CATMA version 2 arrays (complete Arabidopsis transcriptome microarray; Allemeersch *et al.*, 2005; Hilson *et al.*, 2004). CATMA version 2 contained 24 411 gene-specific tags (GSTs). The GSTs, which are between 150 and 500 bp in length and show no more than 70% identity with any other sequence in the genome, were spotted on GAPSII glass slides (Corning Incorporated, Acton, MA, USA) using a BioRobotics Microgrid II TAS spotter (Genomic Solutions, Ann Arbor, MI, USA) and cross-linked for 4 h at 80°C.

Detailed information about CATMA and database access can be found at http://www.catma.org/ (Crowe *et al.*, 2003) and http://genomics.bio.uu.nl/.

Labelling, hybridization and scanning

mRNA from the fractionated polysomes was amplified with the MessageAmp aRNA kit (Ambion). Amplified mRNA (5 μg) was used as a template to synthesize modified cDNA with SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and random nonamers (Gene Link, Westchester County, NY, USA) for 2 h at 42°C with the incorporation of 5-(3-aminoallyl)-dUTP (Ambion; ratio dUTP/dTPP of 7/3), yielding approximately 2 μg cDNA. RNA template was removed by hydrolysis with 3 μl 2.5 M NaOH per 30 μl for 15 min at 37°C. The hydrolysis was stopped by neutralization with 15 μl 2 M MOPS and put on ice. cDNA was purified using the MineLute PCR purification kit (Qiagen) and labelled with Cy3 or Cy5 mono-reactive dye (Amersham, Buckinghamshire, UK). The reaction was quenched after 60 min using 4.5 μl 4 M hydroxylamine (Sigma-Aldrich) and incubated in the dark for 15 min. The labelled cDNA was purified as described above, and incorporation of Cy3 or Cy5 was determined using a UVmini-1240 spectrophotometer at 550 or 650 nm, respectively.

CATMA arrays were denatured in boiling demineralized water for 3 min and dipped in ethanol afterwards. Slides were spun dry for 3 min at 300 ***g*** in 50 ml tubes, covered with a LifterSlip (Erie Scientific Company, Portsmouth, NH, USA), and subsequently prehybridized at 55°C with 100 μl filtered prehybridization solution containing 25% formamide (Merck, Whitehouse Station, NJ, USA), 5x SSC, 0.1% SDS and 1% BSA, fraction V, minimum 96% (Sigma-Aldrich) for 15 min at 55°C. Arrays were transferred to demineralized water, dipped five times and immediately submerged in isopropanol. Slides were spun dry as before, covered with a LifterSlip, and hybridized within 1 h. For each hybridization, 50 μl filtered (0.2 μm) 2x hybridization mix was made, containing 50% deionized formamide (Sigma-Aldrich), 10x SSC and 0.2% SDS. Herring-sperm DNA (1 μl of 11 mg ml^{-1} stock; Sigma-Aldrich) was added, and the mix was heated to 42°C to prevent precipitation of SDS. For all the fractionated MBP and FP samples, a fixed amount of 800 ng cDNA was taken, corresponding to an incorporation of Cy3 or Cy5 of between 103 and 162 pmol. Each pair of labelled MBP and FP was concentrated in a Speed Vac (type SC100; Savant Instruments, Holbrook, NY, USA) to a volume of 50 μl, and added to 50 μl 2x hybridization mix. The probes were denatured for 5 min at 95°C and centrifuged for 2 min at 15 000 ***g*** in a standard tabletop centrifuge and immediately applied to the arrays. The arrays were put in hybridization chambers (Corning Incorporated) containing one drop of 20 μl water on each side, covered by foil, and placed for 16–20 h at 42°C in a water bath.

After the hybridization, the arrays were washed twice in a low-stringency wash solution containing 1x SSC, 0.2% SDS and 0.1 mM DTT for 4 min at 55°C. The arrays were subsequently washed in a high-stringency wash solution containing 0.1x SSC, 0.2% SDS and 0.1 mM DTT for 4 min at 55°C and in a final wash solution containing 0.1x SSC and 0.1 mM DTT each for 4 min at room temperature. The slides were dipped five times in demineralized water and immediately submerged in isopropanol. The slides were spun dry as described above and scans of the arrays were made using a ScanArray Express HT (PerkinElmer, Wellesley, MA, USA). Spot intensities of the scans were determined by ImaGene software version 6.5.1 (BioDiscovery, El Segundo, CA, USA).

Statistics

Spot intensities from CATMA arrays were analysed by LIMMA (Smyth, 2004) version 1.7.2 and limmaGUI version 1.2.5 (Wettenhall and Smyth, 2004) software packages from Bioconductor (Gentleman *et al.*, 2004) running in R version 1.9.1 (CR Foundation, Vienna, Austria; Ihaka and Gentleman, 1996). The intensities were normalized by the print tip LOESS to correct for possible within-array, dye and print-tip effects. Subsequently, all arrays were normalized between arrays by scaling to obtain the same median absolute deviation for each array, thereby enhancing the comparison between them (Smyth and Speed, 2003). A linear model was created to estimate the $\log_2$(MBP/FP) effect and a moderated *t*-test (empirical Bayes) was performed to determine differential transcripts in the MBP or FP fractions. In addition, a posterior residual standard deviation was employed instead of the common standard deviation, making this *t*-test more powerful than the conventional *t*-test, especially in cases where only a few measurements are available (Smyth, 2004). The obtained results were corrected for multiple testing errors using the false discovery rate approach (FDR), yielding *q*-values.

Transcripts were chosen so that the number of false discoveries was as low as 0.001, giving approximately five false discoveries in the selected dataset of 4714 transcripts. The Pearson correlations between the isolations were calculated by fitting *M*-values (least squared averages as calculated by LIMMA).

Quantitative PCR

cDNA was synthesized from the polysomal fractions using SuperScript III reverse transcriptase (Invitrogen) and oligo$(dT)_{15}$ (Promega, Madison, WI, USA). Cycle thresholds (C_T) were determined per transcript in triplicate using the ABI PRISM 7700 sequence detection system (Applied Biosystems, Foster City, CA, USA) using SYBR Green I (Applied Biosystems, Foster City, CA, USA) as reporter dye. Primer sets for the C_T determination of three MBP (MATs) and three FP (FTs) transcripts are listed in Table 3.

Prediction algorithms and gene ontologies

TARGETP version 1.01 and TMHMM version 2.0 predictions of proteins present in the complete ATH1.pep file (TIGR release 5.0) and the GOslim version of 13 August 2005 were obtained from TAIR.

Table 3 Primer sets for three enriched MBP and three enriched FP transcripts

Primers	AGI ID	Sequence (5′–3′)	Amplicon length
MAT-1-Fw	At1g09750	CCGTGTACGAAGCCATTAGA	99
MAT-1-Rv		CGTTATCCGCTGAGAAACAC	
MAT-2-Fw	At5g23660	TCGGATTTGGTTGAGAAAGA	99
MAT-2-Rv		CGCTACTGGTTCAGGAGATG	
MAT-3-Fw	At1g49750	CCGGTCCAATACCGTACTCT	99
MAT-3-Rv		CGCATACAATCTCCGGTATG	
FT-1-Fw	At2g22122	TTCGATGGGAGCAAACTATG	99
FT-1-Rv		ACACCGCCATCTCTTCTCTT	
FT-2-Fw	At3g07910	GCGCAGTAGGTGCTGTCTAT	99
FT-2-Rv		CGCTGCTGAGGGTAGTTTG	
FT-3-Fw	At5g12140	GGCTGATGGTGAGACCAATA	100
FT-3-Rv		AACATCGTGAAGGTGGTTGA	

Fw, forward; Rv, reverse.

Acknowledgements

We would like to thank Rainer Schwacke from the University of Cologne (Germany) for providing us with the consensus ARAMEMNON data (release 3.0) for Arabidopsis membrane proteins, and Annemiek Andel for in-house spotting of the CATMA arrays. This research was funded by a VIDI innovational research grant from The Netherlands Organization for Scientific Research (NWO) to Guido Van den Ackerveken. Bioinformatics analysis was supported by the Centre of BioSystems Genomics (CBSG), which is part of the Netherlands Genome Initiative.

Supplementary Material

The following supplementary material is available for this article online:

One Excel file with CATMA-IDs, corresponding AGI-IDs/protein-IDs, with their observed M, A and *q*-values. Predictions for each protein-ID by TARGETP version 1.01 and TMHMM version 2.0 are included. The ARAMEMNON consensus for each AGI-ID is represented as YES if it is predicted as a membrane protein, otherwise NO. Detailed information can be found at the ARAMEMNON website (http://aramemnon.botanik.uni-koeln.de/). In addition, we include the tag UNEXPECTED for the 755 identified transcripts in the MBP fraction for which the localization was not obvious from prediction data.

The data discussed in this publication have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/; Barrett *et al*., 2005; Edgar *et al*., 2002) and are accessible through GEO Series accession number GSE4023.

Supplementary data for 18 709 CATMA-IDs, AGI-IDs and their corresponding 20 857 Protein-IDs.

The combined microarray data is given for all CATMA-IDs (M, A and *q*-values).

Predictions for each Protein-ID are given for TARGETP (O = 'Other', M = 'Mitochondrion', C = 'Chloroplast' and S = 'Secretory pathway') and TMHMM (Number of transmembrane domains). The ARAMEMNON consensus for membrane proteins is given as YES or NO for each AGI-ID. Unexpected transcripts (AGI-IDs) in the MBP dataset are flagged as UNEXPECTED in the last column. These last two columns contain several times multiple instances due to splice variants (ATxGxxxxx.1 to max ATxGxxxxx.6) and should be considered as one consensus or UNEXPECTED flag. This material is available as part of the online article from http://www.blackwell-synergy.com

References

Allemeersch, J., Durinck, S., Vanderhaeghen, R. *et al.* (2005) Benchmarking the CATMA microarray. A novel tool for Arabidopsis transcriptome analysis. *Plant Physiol.* **137**, 588–601.

Ashburner, M., Ball, C.A., Blake, J.A. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29.

Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Ngau, W.C., Ledoux, P., Rudnev, D., Lash, A.E., Fujibuchi, W. and Edgar, R. (2005) NCBI GEO: mining millions of expression profiles – database and tools. *Nucleic Acids Res.* **33**, D562–D566.

Crofts, A.J., Washida, H., Okita, T.W., Ogawa, M., Kumamaru, T. and Satoh, H. (2004) Targeting of proteins to endoplasmic reticulum-derived compartments in plants. The importance of RNA localization. *Plant Physiol.* **136**, 3414–3419.

Crowe, M.L., Serizet, C., Thareau, V. *et al.* (2003) CATMA: a complete Arabidopsis GST database. *Nucleic Acids Res.* **31**, 156–158.

Davies, E. and Abe, S. (1995) Methods for isolation and analysis of polyribosomes. *Methods Cell Biol.* **50**, 209–222.

Diehn, M., Eisen, M.B., Botstein, D. and Brown, P.O. (2000) Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. *Nat. Genet.* **25**, 58–62.

Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210.

Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016.

Gentleman, R.C., Carey, V.J., Bates, D.M. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80, 1–16.

Gibon, Y., Blaesing, O.E., Hannemann, J., Carillo, P., Hohne, M., Hendriks, J.H., Palacios, N., Cross, J., Selbig, J. and Stitt, M. (2004) A Robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. *Plant Cell*, **16**, 3304–3325.

Hilson, P., Allemeersch, J., Altmann, T. *et al.* (2004) Versatile gene-specific tags for Arabidopsis functional genomics: transcripts profiling and reverse genetics applications. *Genome Res.* **14**, 2176–2189.

Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *J. Comp. Graph. Stat.* **5**, 299–314.

Jarvis, P. (2004) Organellar proteomics: chloroplasts in the spotlight. *Curr. Biol.* **14**, R317–R319.

Jarvis, P. and Robinson, C. (2004) Mechanisms of protein import and routing in chloroplasts. *Curr. Biol.* **14**, R1064–R1077.

Kawaguchi, R., Girke, T., Bray, E.A. and Bailey-Serres, J. (2004) Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *Plant J.* **38**, 823–839.

Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W. and Baginsky, S. (2004) The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* **14**, 354–362.

Kloc, M., Zearfoss, N.R. and Etkin, L.D. (2002) Mechanisms of subcellular mRNA localization. *Cell*, **108**, 533–544.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580.

Larkins, B.A. (1986) Polyribosomes. In *Modern Methods of Plant Analysis. Vol. 2: Cell Components* (Linskens, H.F. and Jackson, J.F., eds). New York: Springer-Verlag. pp. 331–352.

Lerner, R.S., Seiser, R.M., Zheng, T., Lager, P.J., Reedy, M.C., Keene, J.D. and Nicchitta, C.V. (2003) Partitioning and translation of mRNAs encoding soluble proteins on membrane-bound ribosomes. *RNA*, **9**, 1123–1137.

Liao, S., Lin, J., Do, H. and Johnson, A.E. (1997) Both lumenal and cytosolic gating of the aqueous ER translocon pore are regulated from inside the ribosome during membrane protein integration. *Cell*, **90**, 31–41.

Mechler, B.M. (1987) Isolation of messenger RNA from membrane-bound polysomes. *Methods Enzymol.* **152**, 241–248.

Moller, S., Croning, M.D. and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, **17**, 646–653. Erratum (2002) in *Bioinformatics*, **18**, 218.

Nilsson, J., Persson, B. and von Heijne, G. (2000) Consensus predictions of membrane protein topology. *FEBS Lett.* **486**, 267–269.

Oleynikov, Y. and Singer, R.H. (1998) RNA localization: different zipcodes, same postman? *Trends Cell Biol.* **8**, 381–383.

Pavy, N., Rombauts, S., Dehais, P., Mathe, C., Ramana, D.V., Leroy, P. and Rouze, P. (1999) Evaluation of gene prediction software using a genomic data set: application to *Arabidopsis thaliana* sequences. *Bioinformatics*, **15**, 887–899.

Pradet-Balade, B., Boulme, F., Beug, H., Mullner, E.W. and Garcia-Sanz, J.A. (2001) Translation control: bridging the gap between genomics and proteomics? *Trends Biochem. Sci.* **26**, 225–229.

Rapaport, D. (2003) Finding the right organelle. Targeting signals in mitochondrial outer-membrane proteins. *EMBO Rep.* **4**, 948–952.

Schiex, T., Moisan, A. and Rouzé, P. (2001) EuGène: an eukaryotic gene finder that combines several sources of evidence. In *Computational Biology, selected papers from JOBIM' 2000 number 2066 in LNCS*, Springer-Verlag, pp. 111–125.

Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., Frommer, W.B., Flügge, U. and Kunze, R. (2003) ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiol.* **131**, 16–26.

Siegel, V. (1997) Recognition of a transmembrane domain: another role for the ribosome? *Cell*, **90**, 5–8.

Singer, R.H. (1993) RNA zipcodes for cytoplasmic addresses. *Curr. Biol.* **3**, 719–721.

Smyth, G.K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 1–27.

Smyth, G.K. and Speed, T. (2003) Normalization of cDNA microarray data. *Methods*, **31**, 265–273.

Stoltenburg, R., Wartmann, T., Kunze, I. and Kunze, G. (1995) Reliable method to prepare RNA from free and membrane-bound polysomes from different yeast species. *Biotechniques*, **18**, 564, 568.

Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445.

Taliercio, E. and Ray, J.D. (2001) Identification of transcripts translated on free or membrane-bound polyribosomes by differential display. *Plant Mol. Biol. Rep.* **19**, 269a–269f.

Walter, P. and Johnson, A.E. (1994) Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10**, 87–119.

Wettenhall, J.M. and Smyth, G.K. (2004) limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics*, **20**, 3705–3706.

Wilhelm, J.E. and Vale, R.D. (1993) RNA on the move: the mRNA localization pathway. *J. Cell Biol.* **123**, 269–274.

Woolhead, C.A., McCormick, P.J. and Johnson, A.E. (2004) Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell*, **116**, 725–736.

