**Title: Elucidation of the Arabidopsis RNA-binding proteome *in planta***

Short Title: (≤40 characters) Arabidopsis mRNA-protein interactome

Marlene Reichel1\*, Yalin Liao2\*, Mandy Rettel3, Chikako Ragan2, Maurits Evers2, Anne-Marie Alleaume3, Rastislav Horos3, Jeroen Krijgsveld3, Matthias W. Hentze3, Thomas Preiss2,4, Anthony A. Millar1

\* these authors contributed equally to this work

1 Division of Plant Science, Research School of Biology, The Australian National University, Canberra ACT 2601, Australia

2 EMBL–Australia Collaborating Group, Department of Genome Sciences, The John Curtin School of Medical Research (JCSMR), The Australian National University, Canberra, ACT 2601, Australia

3 European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany

4 Victor Chang Cardiac Research Institute, Darlinghurst (Sydney), New South Wales 2010, Australia

Correspondence: tony.millar@anu.edu.au or thomas.preiss@anu.edu.au

Material distribution footnote: The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors ([www.plantcell.org](http://www.plantcell.org)) is: A.A.M

ORCID ID: …0000-0003-3906-2420 (Y.L.), …

Please change current text to 1.5 line spacing, 8.5x11 inch formatting, as required.

Add line number.

Use Arial font in all elements of supplemental data files.

Please check if the manuscript comply to this: in the Abstract, text, and Methods, organisms should be referred to by their common name at first use (if a standard common name applies), and the Latin name should be given in parentheses. Subsequent references to organisms can be either Latin or common names but should be consistent throughout the manuscript.

**ABSTRACT**

RNA-binding proteins (RBPs) are fundamental to controlling the fate and expression of the plant transcriptome. Despite this importance, our understanding of plant RBPs is rudimentary, being mainly derived via bioinformatic extrapolation from other kingdoms. Here, we have adapted the methodology of ‘mRNA-protein interactome capture’ to investigate the RNA-binding proteome *in planta*. From *Arabidopsis thaliana* (At) etiolated seedlings, we captured more than 700 proteins, 300 with high confidence (FDR below 1%), which we defined as the At-RBP set. Approximately 75% of the At RBP set are bioinformatically linked with RNA biology, containing a diversity of canonical RNA-binding domains (RBDs). As no prior experimental RNA-binding evidence exists for the majority of these proteins, their capture now authenticates them as RBPs. Moreover, we identified protein families harbouring emerging and potentially novel RBDs, including WHIRLY, LIM, ALBA, DUF1296 and YTH domain-containing proteins, the latter being homologous to animal RNA methylation readers. Other At-RBP set proteins included major signalling proteins, cytoskeleton-associated proteins, membrane transporters and enzymes, suggesting the potential scope and function of RNA-protein interactions within a plant cell is much broader than previously appreciated. Therefore, our foundation dataset has provided an unbiased insight into the RNA-binding proteome of plants, on which future investigations into plant RBPs can be based.**INTRODUCTION**

The diverse and dynamic interactions with RNA-binding proteins (RBPs) govern the life of cellular RNA, including its processing, modification, cellular localisation, translation and decay (Singh et al., 2015). Even though such post-transcriptional gene regulation events are ubiquitous across all kingdoms of life, relatively little is known about RNA-protein interactions in plants and how these events impact the fate and expression of the transcriptome. Instead, much of the research on post-transcriptional gene regulation in plants has focused on the role of small RNAs (sRNAs), which has been driven by the development of next-generation sequencing methodologies, enabling the relative ease of identification of sRNAs and their targets (Ma et al., 2015). By comparison, the cohort of RBPs of a plant cell has yet to be experimentally determined.

Knowledge on RBPs in plants comes mainly from targeted studies on individual proteins or from bioinformatic predictions based on sequence homology with canonical RNA-binding domains (RBDs) identified in other kingdoms (Silverman et al., 2013). For instance, there are hundreds of Arabidopsis genes that encode proteins exhibiting one or more canonical RBD, such as RNA recognition motif domains (RRM; 197 proteins), K homology domains (KH; 28 proteins), cold shock domains (CSD; 5 proteins), DEAD-box helicase domains (9 proteins), Pumilio RNA-binding repeats (PUF; 26 proteins), Like-Sm domains (LSM; 36 proteins), Zinc finger CCCH-type (C-x8-C-x5-C-x3-H; 5) proteins and pentatricopeptide repeat proteins (PPR; approximately 450 proteins) (Silverman et al., 2013). However, to date, only a few of these proteins have been functionally characterized. Examples include the RRM-containing GLYCINE-RICH RNA-BINDING PROTEINS (GR-RBPs) that have been implicated in mediating responses to various stresses such as cold, salinity and drought (Kim et al., 2007a; Kim et al., 2007b; Kwak et al., 2005; Lorković et al., 2009) and in regulating circadian rhythm (Nolte and Staiger, 2015). Similarly, CSD proteins and RNA helicases (RH) have been shown to be involved in abiotic stress responses (Jung et al., 2013). Proteins with known RBDs also play important roles in plant developmental processes such as flowering time (Macknight et al., 1997; Schomburg et al., 2001), floral morphogenesis (Lorkovic, 2009; Jung et al., 2013), embryogenesis (Tripurani et al., 2011), as well as ovule development and cell size homeostasis (Bush et al., 2013). Nevertheless, for the vast majority of bioinformatically predicted plant RBPs there is no experimental evidence for their RNA-binding activity or their molecular function. Additionally, what non-canonical RBDs exist in plants remains to be determined, as there has been no global, unbiased experimental approach taken to determine the cohort of plant RBPs (Silverman et al., 2013).

Recently, a methodology termed “mRNA interactome capture” was developed that can identify the portion of cellular proteomes that is bound to polyadenylated RNA (Castello et al., 2012; Baltz et al., 2012). The method uses irradiation of live cells with short-wave UV light (254 nm), which, unlike formaldehyde, is known to selectively cross-link proteins in direct contact to RNA (Greenberg, 1979; Dreyfuss et al., 1983; Wagenmakers et al., 1980). Following cell lysis, cross-linked mRNA-protein complexes are isolated using oligo(dT) beads under stringent conditions, prior to RNase treatment and protein identification by mass spectrometry (MS). mRNA interactome capture has been performed on a range of mammalian cell lines (Castello et al., 2012; Baltz et al., 2012; Beckmann et al., 2015; Kwon et al., 2013; Liao et al., 2016), *Caenorhabditis elegans* (Matia-Gonzalez et al., 2015), *Drosophila melanogaster* (Wessels et al., 2016) and *Saccharomyces cerevisiae* (Beckmann et al., 2015; Mitchell et al., 2013; Matia-Gonzalez et al., 2015). These studies have revealed unexpectedly high numbers of diverse RBPs in eukaryotic cells, indicating that many unforeseen RNA-based regulatory mechanisms have yet to be elucidated. Currently such an approach has not been reported for plants.

Here, we detail the successful adaption of mRNA interactome capture to a living intact plant. Using Arabidopsis etiolated seedlings as source material, we have identified 300 Arabidopsis proteins as RNA-binding, and present another set of over 400 proteins as candidate RBPs, underscoring the prevalence of RNA-binding and RBD diversity within the plant proteome. Corroborating our approach, many known RBPs were isolated, along with a multitude of bioinformatically predicted RBPs, providing the first direct experimental evidence of their *in vivo* RNA-binding activity. Moreover, we identified potential novel plant RBDs and a diverse set of proteins not previously associated with RNA biology including proteins involved in signalling pathways, cytoskeleton organization and membrane transport. Our study thus reports a novel method of broad utility in plant research, as well as providing the first experimental census of the Arabidopsis RNA-binding proteome as a unique resource for future research into RBP function in plants.

**RESULTS AND DISCUSSION**

**Development of an mRNA interactome capture protocol for Arabidopsis seedlings**

Four-day old etiolated seedlings were chosen as source material for mRNA interactome capture (Figure 1A). To adapt the original protocol (Castello et al., 2012) for use with plant material, we increased the dosage of 254 nm UV light to three cycles of irradiation at 150 mJ/cm2 (Au et al., 2014), to establish crosslinks (CL) between proteins and RNA. This did not appreciably increase RNA degradation compared to lower UV dosages or a non-crosslinked (noCL) control sample (Supplemental Figure 1A). Snap-frozen seedlings were ground in liquid nitrogen and thawed into lysis buffer that was adapted for use in plants by supplementing with -mercaptoethanol and polyvinylpyrrolidone 40 (PVP40). Lysates were further passed through a shredding column to clear the lysate and allow for efficient isolation of mRNA-protein complexes after two rounds of capture on oligo(dT) beads (Supplemental Figure 1B, C). Finally, proteins were released from the beads and treated with RNase. Analysis of aliquots taken before and after the oligo(dT) capture (referred to as input and eluate, respectively) by SDS-PAGE and silver staining showed purification of a distinct set of proteins in the CL eluate but not the noCL control sample (Figure 1B). Western blot analysis confirmed that this set of proteins contained known RBPs such as ARGONAUTE 1 (AGO1) and the chloroplast RBP CP29A, whereas non-RBPs such as the inositol phosphatase FIERY1 (FRY1)/SAL1 protein were absent (Figure 1C). Together, this demonstrated that the modified mRNA interactome capture protocol could efficiently and selectively purify plant RBPs.

**Identification of an Arabidopsis RNA-binding proteome**

Three independent biological replicates for both CL and noCL eluate samples were prepared using the SP3 (Single-Pot Solid-Phase-enhanced Sample Preparation) method (Hughes et al., 2014) for analysis by quantitative MS. Scatter plots comparing CL/noCL fold changes between biological replicates showed reproducibility (Supplemental Figure 1D). Together, this identified 746 proteins in the eluates and for 333 of these, a CL/noCL ratio was quantified in at least two out of three replicates. Nine proteins were enriched in the noCL sample and were not considered for further analysis, while 324 proteins were enriched in the CL sample. Of the latter, 300 proteins were enriched at a false discovery rate (FDR) < 1% and were defined as the Arabidopsis RBP set (At-RBPs; Figure 1A and listed in Supplemental Data Set 1). The 24 proteins enriched at a higher FDR and those 413 proteins without a CL/noCL ratio (i.e. ’non-quantified’), were defined as candidate At-RBPs (Figure 1A; listed in Supplemental Data Set 1). Although candidate At-RBPs did not pass stringent criteria, their features are nevertheless examined below as they likely contain additional RBPs that warrant further investigation. MS was also performed on two independent input samples, which identified 8264 proteins (termed input proteome).

**Interactome capture enriches for proteins related to RNA biology**

Compared to the input proteome, both At-RBPs and candidate At-RPBs were enriched for RNA-related Gene Ontology (GO) terms and canonical RBDs (Figure 1D, E). Based on GO annotations, 75% of At-RBPs had prior experimentally determined or predicted links to RNA biology. The remaining 25% had no known or predicted function in RNA biology and therefore represent novel RBPs in Arabidopsis. Similarly, about 80% of the At-RBPs contained known RBDs (based on pfam and Interpro annotations and previous interactome datasets (Castello et al., 2012; Beckmann et al., 2015; Liao et al., 2016)), while the other 20% did not (Figure 1D). The candidate At-RBPs grouping also showed these enrichments albeit at a lower level; 46% of them were classified as ‘linked to RNA biology’ and 48% of proteins had recognized RBDs (Figure 1E). Multiple molecular function GO terms were enriched among At-RBPs and candidate At-RBPs, the majority of which were associated with RNA, nucleic acid binding or translation (Figure 1F, G). Similarly, Biological Process (Supplemental Figure 2 A, C) and Cellular Component (Supplemental Figure 2 B, D) GO terms referred to a range of processes and components typically associated with RNA, such as translation and the ribosome, as well as splicing and several types of RNA granules.

Approximately 25% of At-RBPs lack known functions in RNA biology, a proportion similar to other interactomes (Beckmann et al., 2016). However, the number of 300 proteins that qualify as At-RBPs is less than what had been observed in the two other interactomes on multicellular organisms, i.e. *C. elegans* (594 proteins) and *D. melanogaster* (476 proteins) (Matia-Gonzalez et al., 2015; Wessels et al., 2016). This may suggest the presence of fewer RBP networks in plants compared to animals, but a lower UV cross-linking efficiency for plants due to their abundant UV-absorbing pigments is a more likely explanation, although we tried to reduce this through the use of etiolated seedlings.

Next, we investigated the conservation of RBPs across the major eukaryotic kingdoms. 200 proteins of the At-RBP set were predicted to have orthologs in human, mouse and/or yeast (Supplemental Figure 3A) as determined by the InParanoid database (Sonnhammer and Östlund, 2015). Of these, 64 proteins were only found in the At-RBP set (discussed below), while 136 proteins have been detected as RBPs in other mRNA interactomes including several without prior association to RNA-binding (Supplemental Figure 3A, Supplemental Data Set 1). Such strong overlap strengthens the confidence that most of the At-RBPs are *bona fide* RBPs. Finally, 52 At-RBPs were present in interactomes of all three kingdoms (Supplemental Figure 3B). This group mainly comprises proteins involved in mRNA translation, splicing and helicase activity (Supplemental Data set 1), all of which are core eukaryotic RNA functions.

**Biophysical properties of captured proteins are characteristic of RBPs**

Next, we examined the biophysical and amino acid sequence features of captured proteins (Figure 2). We used the properties of the input proteome as a reference for diverse proteins, whereas input proteins with the GO annotation ‘RNA-binding’ was a reference for expected properties of known/predicted RBPs. These two groups were compared to the properties of the At-RBP set, of those At-RBPs without known RBDs and of candidate At-RBPs. All five groupings spanned the full range of protein sizes, with some bias towards larger proteins among the At-RBPs with unknown RBDs (Figure 2A). Compared to the input proteome, all four groups showed significantly increased proportions of residues in intrinsically disordered regions (Figure 2B), which have been linked to protein-protein, protein-DNA and protein-RNA interactions (Wright and Dyson, 2015; Calabretta and Richard, 2015). Furthermore, all four RBP sets exhibited significant shifts towards a more alkaline isoelectric point and a lower hydrophobicity compared to the input proteome (Figure 2C, 2D). This is also reflected in their amino acid composition. Hydrophobic and aromatic amino acids such as leucine (L), isoleucine (I), and valine (V), as well as amino acids with aliphatic side chains such as tryptophan (W) and phenylalanine (F), which have all been shown to have low propensity to bind RNA (Jeong et al., 2003; Lejeune et al., 2005), were depleted relative to the input proteome (Figure 2F and Supplemental Figure 4B). Cysteine (C) also showed strong depletion, which is consistent with its low propensity to bind RNA (Lejeune et al., 2005) and its depletion in intrinsically disordered regions (Theillet et al., 2013; Williams et al., 2001). By contrast, proline (P) is enriched in all four RBP sets, in agreement with its strong enrichment in disordered regions (Theillet et al., 2013; Williams et al., 2001). Furthermore, positive and polar amino acids such as arginine (R), glutamine (Q), asparagine (N) and histidine (H), which have a high propensity to bind RNA, were enriched among all RBP sets (Figure 2F and Supplemental Figure 4A). The smallest amino acid glycine (G), which can form strong interaction with the nucleotide guanine (Lejeune et al., 2005), also showed strong positive enrichment (Supplemental Figure 4A). Overall, the At-RBPs showed the strongest biases in these features. At-RBPs without known RBDs and candidate At-RBPs also followed these trends albeit to a lesser extent. Taken together, these findings mirror those reported for other mRNA interactomes (Castello et al., 2012; Liao et al., 2016), and indicate that mRNA interactome capture strongly enriched for *bona fide* plant RBPs.

**Interactome capture identifies a diverse set of proteins with a range of recognized RBDs**

Next, we grouped At-RBPs based on their annotated protein domains (Figure 3A, B). mRNA interactome capture identified a broad array of proteins with known RBDs, covering more than 30 different known types. Of these, proteins containing RRM domains constituted the largest class (80 in the At-RBP set and 50 within candidate At-RBP set) and thus we captured the majority of the 197 bioinformatically predicted RRM domain proteins in Arabidopsis (Silverman et al., 2013). Similarly, of the 28 predicted KH domain proteins (Silverman et al., 2013), 19 have been detected in the interactome (7 proteins) and candidates (12 proteins) (Figure 3A), indicating that the majority of RRM and KH domain proteins are expressed and bound to poly(A) RNA in the seedling. Proteins harbouring diverse canonical RBDs such as the Nuclear Transport Factor 2 (NTF2), Like-Sm (LSM), Pumilio (PUF) and La domains, were all readily captured (Figure 3A). Additionally, multiple zinc-finger proteins were identified as RNA-binding (Figure 3A), including those sub-types known to interact with RNA, such as zf-CCCH, zf-CCHC and zf-C2H2 (Li et al., 2013; Ciftci-Yilmaz and Mittler, 2008), as well as others such as zf-RanBP, which has been shown to interact with RNA in humans (Nguyen et al., 2011; Vandevenne et al., 2014), but not in plants. Within these zinc-finger protein classes, ten proteins not previously associated with RNA-binding have been identified (Supplemental Table 1), expanding our knowledge on zinc-finger containing RBPs.

Other canonical RBPs detected include ARGONAUTE (AGO) proteins that contain PAZ and PIWI domains (Figure 3A). Both AGO1 and AGO2 were detected in the At-RBP set and AGO4 in the candidate At-RBP set. This is the first report of successful UV cross-linking of AGO proteins to mRNA in plants, and provides the basis for future determination of sRNA targets via methods such as immunoprecipitation followed by high-throughput sequencing (HITS-CLIP), a long established method in animal cell yet to be applied in plants (Chi et al., 2009). Lastly, we have captured many canonical RBP families that are involved in responses to various abiotic stresses including GR-RBPs, cold shock domain (CSD) proteins, tudor-SN proteins and DEAD box RNA helicases (Kim et al., 2007b; Kwak et al., 2005; Lorković et al., 2009; Jung et al., 2013; dit Frey et al., 2010) (Figure 3A).

**Limited capture of mitochondrial and chloroplastic RBPs in etiolated seedlings**

Consistent with previous studies (Castello et al., 2012; Liao et al., 2016), we found multiple cytoplasmic ribosomal proteins (RPs) within the plant mRNA interactome (27 proteins of the large and 32 of the small ribosomal subunit; Supplemental Data Set 1). Several RPs are known to be in direct contact with mRNA (Pisarev et al., 2008) and a number of RPs also have extra-ribosomal functions in mRNA regulation (Warner and McIntosh, 2009). The capture of polyadenylated rRNA processing intermediates (Sikorski et al., 2015) is also likely to have contributed to the isolation of cytoplasmic RPs. By contrast, only one chloroplastic and one mitochondrial RP were captured despite a total of 46 mitochondrial and chloroplastic RPs being present in the input proteome. This stark contrast in the capture of cytoplasmic and mitochondrial/chloroplastic RPs is consistent with most mature, translatable transcripts in these organelles not harbouring a poly(A) tail (Chang and Tong, 2012). Likewise, only 18 PPR proteins were detected in the interactome (six in At-RBPs, 12 in candidate At-RBPs), out of the 60 PPR proteins identified in the input proteome (Figure 3A). Again, this poor ratio of interactome/input proteome of PPR proteins is possibly explained by their mitochondrial and chloroplastic location (Colcombet et al., 2013). Although, known chloroplast RBPs such as CP29A, CP31A and CP31B were strongly enriched by interactome capture, CP29A has also been shown to interact with nuclear mRNAs (Gosai et al., 2015), so it is not certain that this RBP is being captured from the organelle. With regards to the PPR proteins, there were only 60 in the input proteome, a fraction of the 450 predicted PPR proteins in Arabidopsis (Silverman et al., 2013). Again this strongly contrasts to other RBD classes (RRM; 160 in the input proteome of the predicted 197; Silverman et al., 2013), and therefore the lack of PPRs in general may reflect that the etioplasts have yet to differentiate, hence the results may be considerably different in light-grown seedlings.

**Interactome capture provides experimental evidence of RNA-binding activity for many predicted plant RBPs**

Although approximately 80% of the At-RBP set are annotated to contain canonical RBDs and RNA-related GO terms, this is primarily based on *in silico* predictions and not yet on experimental evidence. For example, of the 25 members of the RNA-binding (RRM/RBD/RNP motifs) protein family that are present in the At-RBP set, only one has a demonstrated biological function and RNA-binding activity (Zhang et al., 2015). Another example is the family of CTC-interacting-domain (CID) proteins, which bind to the C-terminal domain of poly(A)-binding proteins (PABPs) via their poly(A)-binding protein interacting motif 2 (PAM2) (Bravo et al., 2005). CID proteins are categorized into four groups based on their other domains (Bravo et al., 2005). Until now, only CID12, which contains PAM2 and RRM domains, has been shown to interact with RNA *in vitro* (Hecht et al., 1997). We have identified CID12 as well as CID1, 3, 4, 7, 8, 10 and 11 in the interactome and candidate proteins (Figure 4A), and thus provide evidence of their RNA-binding activity *in vivo* for the first time.

Similarly, plant PUF proteins are largely only predicted to bind RNA. Of the 25 family members in Arabidopsis, only PUM2 and PUM5 have been experimentally demonstrated to bind RNA (Francischini and Quaggio, 2009; Huh et al., 2013; Huh et al., 2014). Here, we provide evidence that PUM1-6, all of which belong to group I with the strongest homology to the Drosophila PUF domain (Francischini and Quaggio, 2009), bind to RNA *in vivo* (Figure 4A). Most of the other PUF proteins were not present in the input proteome, suggesting that PUF1-6 are the major players of this family in early Arabidopsis growth and development.

Other predicted RBPs include the Nuclear Transport Factor 2 (NTF2) protein family that contains an NTF2 domain, which is required for protein-protein interactions (Ribbeck et al., 1998; Fribourg et al., 2001), but has been shown to interact with RNA in humans (Katahira et al., 2015). In Arabidopsis, 18 proteins are predicted to contain NTF2 domains; these proteins can be classified into two groups – group I proteins, which contain both NTF2 and RRM domains and group II proteins, which only have NTF2 domains. We have identified six out of eight group I proteins in the At-RBP set and one among the candidate At-RBPs (Figure 4A), thereby verifying their RNA-binding activity. As all the group I proteins contain both NTF2 and RRM domains, and none of the group II proteins (NTF2-like only) were present in At-RBPs or candidate At-RBPs, this suggests that the NTF2 domain per se may not be an RBD in plants.

We also captured both Arabidopsis homologs of Barentsz [BTZ; also known as METASTATIC LYMPH NODE 51 (MLN51)] harbouring the Btz domain (Figure 4A), which is a known RBD in animals (Bono et al., 2006). Barentsz, eIF4A3, Y14 and Mago Nashi (MAGO) form the core of the exon-junction complex (EJC). The EJC is deposited on nascent mRNA at splice junctions and functions in subsequent mRNA utilization (Nyikó et al., 2013). In addition to the Barentsz homologs, we identified eIF4A3 and Y14 in the At-RBP set and MAGO among the candidate At-RBPs (Figure 4A). In humans, only Bartentsz and eIF4A3 are thought to be in direct contact with RNA when bound to the EJC (Bono et al., 2006; Le Hir and Andersen, 2008). The fact that we detected RNA-binding by all four core proteins of the plant EJC suggests that Mago and the RRM-containing Y14 may have RNA-binding functions which may be distinct to their role in the EJC. The EJC also serves as a mark for nonsense-mediated decay (NMD) (Kim et al., 2001). Other components of NMD include UP-FRAMESHIFT proteins (UPF1-3), all three of which have been identified in the At-RBP set (Figure 4A). While the role of plant UPF1 as activator of NMD is relatively well studied, UPF2 and UPF3 are less well characterized and the molecular mechanism of NMD in plants is still poorly understood (Dai et al., 2016).

Further examples of protein families that are annotated as RBPs with little experimental evidence in plants include MEI2-like proteins, La-related proteins (LARP), BRUNO-LIKE proteins and others listed in Figure 4A. Many members of these protein families were shown here as plant RBPs for the first time.

**Identification of proteins harbouring domains linked to RNA in other kingdoms, but thus far not in plants**

**The majority of YTH and ALBA domain containing proteins are present in the At-RBP set.**

**Identification of proteins with well characterized RNA-binding functions in other kingdoms, but not in plants**

The At-RBP set contains a number of proteins that have no annotated links to plant RNA biology, despite harbouring domains associated with RNA-binding. One standout example is the Arabidopsis YT521-B homology (YTH) domain containing family of proteins. There are thirteen YTH proteins in Arabidopsis (Li et al., 2014), which show distinct developmental expression patterns and responses to stress and include the eleven Evolutionary Conserved C-terminal domain (ECT) domain family proteins (Ok et al., 2005). Ten of the eleven ECT proteins were captured here, nine in the At-RBP set and two in the candidate At-RBPs (Figure 4B). The mammalian YTH proteins YTHDF1-3 and YTHDC1 were recently shown to bind the epitranscripomic mark N6-methyladenosine (m6A), impacting mRNA splicing, export, translation or turnover (Schumann et al., 2016; Schwartz, 2016). YTH domains bind m6A in a cage of aromatic amino acids that is conserved among all Arabidopsis YTH domain proteins (Fray and Simpson, 2015). The fact that all these proteins only have the YTH domain in common (Figure 4B), makes this domain a strong candidate for an RBD also in plants.

A further YTH domain protein in the At-RBP set is the Cleavage and Polyadenylation Specificity Factor 30 (AtCPSF30), which functions as part of a larger complex in mRNA 3’ end formation (Hunt et al., 2012). Analyses with mutant Arabidopsis plants deficient in CPSF30 indicated roles in stress responses (Zhang et al., 2008), immunity and programmed cell death (Bruggeman et al., 2014), as well as demonstrating altered mRNA 3’ end cleavage site choice in a large number of genes (Thomas et al., 2012). Interestingly, Arabidopsis expresses two CPSF30 protein variants from a single gene due to alternative mRNA 3’ end formation: a shorter form of approximately 28kD that harbours three zinc finger domains and is homologous to yeast and mammalian CPSF30, and a longer form of approximately 70kD that adds a YTH domain and is unique to plants (Hunt et al., 2012; Delaney et al., 2006). These observations, together with evidence that m6A is enriched near 3’ends of Arabidopsis mRNAs (Luo et al., 2014; Wan et al., 2015; Bodi et al., 2012), led to recent speculation about a role of m6A in governing mRNA 3’ end formation in plants (Fray and Simpson, 2015; Chakrabarti and Hunt, 2015; Burgess et al., 2016). To formally investigate such a link, we obtained Arabidopsis data on m6A site distribution (Wan et al., 2015) and 3’ end cleavage and polyadenylation maps (polyadenylation cluster (PAC) sites from Wu et al., 2011 and cleavage site from Sherstnev et al., 2012)and performed spatial co-location analyses. Indeed, we found an enrichment of m6A sites within a 100 nt window upstream of mRNA 3’ ends (Figure 4C). No such enrichment was seen in mammals where m6A peaks were examined within a 50 nt window upstream of known poly(A) cleavage sites, but no significant association was found (Meyer et al., 2012). Cleavage and polyadenylation regions in plant mRNAs consist of three signals: far upstream (FUE) and near upstream elements (NUE; equivalent to the mammalian AAUAA motif), as well as sequences immediately surrounding the cleavage site (Loke et al., 2005; Thomas et al., 2012). We detected maximally enriched m6A abundances at around -45 nt relative to PAC and cleavage sites (Figure 4C), which lies upstream of the NUE (-13 to -30 nt) at the 3’ boundary of the FUE region (-50 to -130 nt). As AtCPSF30 has a role in the function of the NUE (Thomas et al., 2012), this finding suggests complementary roles of RNA-binding through its zinc finger and YTH domains, perhaps conveying an m6A dependence for a subset of mRNA 3’end cleavage sites that should be tested in future work.

Another example is the Alba domain-containing protein family, where we identified four Alba proteins amongst the At-RBPs and the fifth among the candidate At-RBPs (Figure 4D), which accounts for all Alba proteins of Arabidopsis and demonstrates for the first time that this family of proteins can bind to RNA *in planta*. Currently, nothing is known about the molecular or functional role of Alba proteins in plants. Alba proteins are widely distributed in archaea where they are a major component of chromatin and involved in transcriptional repression through binding to DNA, but are also know to interact with RNA (Bell et al., 2002; Jelinska et al, 2005; Forterre et al., 1999; Guo et al., 2003). They are structurally similar to prokaryotic translation initiation factor 3 (Aravind et al., 2003) and have been reported to play a role in translational control in multiple eukaryotes (Mani et al., 2011; Gissot et al., 2013; Mair et al., 2010), so there is precedence for proteins containing this domain to bind RNA. Interestingly, the Alba gene family of the protozoan parasite *Trypanosoma brucei* share a similar structure to that in Arabidopsis, where there are members encoding small proteins that only contain the ALBA domain, and members encoding longer proteins with a C-terminal region rich in RGG (arginine-glycine-glycine) boxes (Figure 4D; Subota et al., 2011). These motifs are known to promote RNA-binding (Thandapani et al., 2013), and are often found in combination with other RBDs (Castello et al., 2012). Thus, they might function synergistically with the Alba domain to facilitate RNA-protein interactions.

**Proteins without known RBDs and no prior association with RNA biology**

About one sixth of the At-RBPs have neither recognized RBDs nor annotated roles in RNA biology. Many of these are plant specific, where they are either only found in the At-RBPs but not in other interactomes, or have no identified orthologs in other kingdoms (Supplemental Figure 3A). The latter group includes WHIRLY (WHY) domain-containing proteins, which form a small family of single stranded DNA-binding proteins localized to organelles where they maintain genome stability (Krause et al., 2005; Marechal et al., 2008; Cappadocia et al., 2010; Marechal et al., 2009). In maize, the chloroplast-localized WHY1 has been shown to bind to both DNA and a subset of plastid RNAs *in vitro* (Prikryl et al., 2008). Here, we have identified all three Arabidopsis WHY proteins as At-RBPs (Figure 5A) and therefore provide the first *in vivo* evidence that these proteins are a family of RBPs in Arabidopsis.

Another example is a group of largely uncharacterized proteins with the Domain of Unknown Function 1296 (DUF1296). The Arabidopsis genome encodes for eight proteins containing DUF1296 domains; seven kinase-related proteins of unknown function and a G-BOX TRANSCRIPTION FACTOR-INTERACTING PROTEIN1 (GIP1), which is involved in regulating the activity of transcription factors involved in plant development (Lee et al., 2014; Shaikhali, 2015). We have identified three kinase-related proteins among the At-RBPs, and GIP1 as well as another kinase-related protein in the candidate At-RBPs (Figure 5B). The only domain that is common to these proteins is the DUF1296 domain (Figure 5B), suggesting that this domain might be a novel RBD in plants.

Furthermore, we have identified several proteins containing the LIM-domain, which consists of tandem zinc-finger structures. LIM domains are present in a wide range of eukaryotic organisms and have been shown to mediate protein-protein interactions (Kadrmas and Beckerle, 2004). We have captured four of the six Arabidopsis *LIM* genes in our RBP sets (Figure 5C). LIMs function in cytoskeleton organization by binding to actin filaments (Papuga et al., 2010; Ye and Xu, 2012), but have no known role in RNA-binding. RNA transport along the cytoskeleton is a major mechanism of mRNA localization and requires motor proteins that move the RNA cargo in form of ribonucleoprotein (RNP) particles along the cytoskeletal tracks (Bullock 2011; Gagnon and Mowry, 2011; Jansen, 1999). Despite some well-studied examples in *Drosophila* and yeast (Bullock, 2011), relatively little is known about how RNPs are connected to the motor proteins in plants. LIM proteins might carry out this connecting role in plants thereby mediating mRNA transport along actin filaments.

**The scope of non-canonical RBPs in Arabidopsis**

**No obvious domain.**

LIM proteins are not the only cytoskeletal proteins among the At-RBPs, as we have also captured actin and tubulin (Supplemental Table 2, Figure 5D), which form microfilaments and microtubules, respectively, the major components of the cytoskeleton. Actin is also present in the nucleus where its function is less well studied (Falahzadeh et al., 2015). Interestingly, in animals, nuclear actin was found to be part of hnRNPs, proteins that are involved in mRNA processing, export, localization and stability (Hofmann, 2009), but no such functional role is known of nuclear actin in plants.

Curiously, we have also identified a subset of aquaporin proteins, which belong to the well-studied family of major intrinsic proteins (MIP) (Figure 5D, Supplemental Table 2). These proteins form transmembrane channels that transport water, other small solutes and gases (Quigley et al., 2002; Biela et al., 1999; Gaspar, 2003; Holm et al., 2005; Uehlein et al., 2003), but there are no reports of aquaporins transporting or interacting with RNA. We mainly found plasma membrane intrinsic proteins (PIPs); PIP2;1, PIP2;2, PIP2;7 among the At-RBPs and PIP1;1, PIP1;2, PIP1;3 as well as the tonoplast intrinsic protein TIP1;2 in the candidate At-RBPs. Considering the ever expanding types of substrates assigned to aquaporins (Maurel et al., 2015), it is intriguing to speculate that RNAs, either as a protein template or as active signalling molecule, may also travel through aquaporins, similar to protein assisted cell-to-cell transport of RNA during virus infection (Peña and Heinlein, 2013).

Other examples of well-characterized proteins identified as At-RBPs include several plant-specific proteins involved in major signal transduction pathways of etiolated seedlings (Supplemental Table 2, Figure 5D). Firstly, we captured ETHYLENE-INSENSITIVE 2 (EIN2), which is a classic ethylene signalling protein expressed strongly in etiolated seedlings, in which *ein2* mutants display the triple response of reduced apical hook formation and elongated hypocotyl and roots. Independently validating our mRNA interactome data, EIN2 has recently been shown to bind to EIN3-BINDING F-BOX 1 (EBF1) and EBF2 mRNA in the presence of ethylene, thereby promoting their translational repression and activating ethylene responsive genes (Li et al., 2015; Merchante et al., 2015). Another signalling protein among the At-RBPs is PHYTOCHROME A (PHYA), the primary red-light photoreceptor in seedlings that mediates many aspects of seedling de-etiolation in response to light (Casal et al., 2014). PHYA has no previous association with RNA-binding, but wide-spread changes in alternative splicing were observed in *phyA.phyB* double mutants (Shikata et al., 2014) and PHYB has recently been found to regulate translation of mRNAs in the cytosol (Paik et al., 2012). Finally, a co-acting receptor of PHYA, the blue-light receptor PHOTOTROPIN 1 (PHOT1) that is required for the inhibition of hypocotyl growth during de-etiolation, was also detected by interactome capture. Similar to PHYA, PHOT1 has not been ascribed an RNA-binding function, however it is required for blue-light-mediated mRNA destabilization (Folta and Kaufman, 2003). Thus, photoreceptors are emerging as regulators of post-transcriptional events through mechanisms that still remain to be determined, but possibly involve RNA-binding functions.

A number of proteins in the At-RBP set have been implicated in transcriptional control. Firstly, PURA1 (PURIN-RICH ALPHA 1) and AtNFXL1 are Arabidopsis homologs of known human transcription factors (TFs). PURA1 has been shown to interact with the 5’ region of many Arabidopsis genes (Tremousaygue et al., 1999) and AtNFXL1 is a NF-X1 type zinc finger protein. Additionally, MULTIPROTEIN BRIDGING FACTOR 1A, a highly conserved transcriptional co-activator (Tsuda et al., 2004;), is found in the At-RBP set as well as in human, mouse and yeast interactomes (Supplemental Data Set 1). A number of transcription factors (TFs) were previously identified in mammalian mRNA interactomes, for example, twenty TFs were found in mouse cardiomyocytes (Liao et al., 2016), setting up intriguing possibilities of RNA-binding being involved in the regulation of transcription.

We have also captured several stress- responsive proteins that have no RNA related GO annotations, such as HEAT SHOCK PROTEIN 90.2 (HSP90.2; AT5G56030), which was identified in the At-RBPs, and HSP70.1 and HSP40, which was found in the candidate At-RBPs (Figure 5D). These proteins are highly conserved molecular chaperones involved in protein folding, stability and activation (Wang et al., 2004). RNA-binding roles are not unprecedented for HSPs as mammalian HSP70 is able to bind to the 3’UTR of labile mRNAs (Henics et al., 1999) and NbHSP90 of *Nicotiana bethamiana* can interact with *Bamboo mosaic virus* genomic RNA enhancing its replication (Huang et al., 2012). Moreover, HSP90 is involved in RNA interference in human and yeast (Wang et al., 2013) and has been captured in human, animal and yeast interactomes. Therefore, this protein is possibly a conserved RBP, although a direct RNA-binding function remains to be determined.

Another example is Annexin 4 (ANN4), which is a member of a multigene family of Ca2+-dependent membrane-binding proteins (Loahavisit and Davies, 2011; Figure 5D; Supplemental Table 2). ANN4 is involved in osmotic stress and ABA signalling in a Ca2+ dependent manner (Lee et al., 2004), but has no prior association with RNA-binding. ECT1 and ECT2 have also been implicated in Ca2+ signalling, where they were shown to interact with Calcineurin B-Like-Interacting Protein Kinase 1 (CIPK1), the primary target of the Ca2+ sensor Calcineurin B-Like 1 (CBL1) (OK et al., 2005). Furthermore, the RNA-binding activity of CPSF30 is affected by the Ca2+-binding protein calmodulin (Delaney et al., 2006). These findings suggest that there might be a link between Ca2+ signalling pathways and RNA metabolism in plants.

Finally, we have captured a number of metabolic enzymes such as RAFFINOSE SYNTHASE 6 (RS6), which is involved in carbohydrate metabolism and whose expression is dark inducible and repressed by sugar (Fujiki et al., 2001) (Figure 5D). Additionally, S-ADENOSYLMETHIONINE SYNTHASE 4 (METK4) was captured, along with its closely associated homolog METK3 that was present in the candidate At-RBP set. DNA topoisomerase type IA, CATALASE-3, peroxidase 69 and the SNF1-RELATED PROTEIN KINASE 2.5 (SRK2G) were also present in the At-RBP set (Figure 5D, Supplemental Table 2). It is tempting to speculate that these enzymes may be regulated by RNAs, altering their activities or specificities. Recent mRNA interactome data in higher eukaryotes revealed the striking aspect that many enzymes bear RNA-binding functions (Castello et al., 2012; Beckmann et al., 2015; Lia et al., 2016; Matia-Gonzalez et al., 2015). Elucidating such interactions may lead to the discovery of novel RNA-based regulatory mechanisms.

**mRNA interactome capture - a method for studying RNA-protein interactions *in planta***

Our study presents the first system-wide, *in vivo* analysis of proteins bound to mRNA in plants. We have identified more than 700 proteins, 300 of them with high confidence (FDR below 1%), which constitute the Arabidopsis mRNA-binding proteome. The study gives a snapshot of the extent of RNA-protein interactions occurring in etiolated seedlings. Given that hundreds of RBPs have been identified, many of which are likely regulating/targeting multiple mRNAs, it is not unreasonable to propose that the majority of seedling mRNAs are possibly under post-transcriptional control via RBP regulation. As the biological role is unknown for the vast majority of these RBP, it is clear that our knowledge of post-transcriptional gene regulation of a plant is superficial at best. Our successful development of interactome capture *in planta* will not only allow us to identify the portion of the proteome that is interacting with RNA, but will enable the analysis of the dynamic nature of these interactions via the comparison of multiple interactomes from different tissues, developmental stages and environmental conditions of intact living plants. Moreover, we envision that our study will facilitate the development of other methods based on UV cross-linking such as CLIP and variations thereof, long standing and widely utilized methodologies applied in animal cells (Ule et al., 2003), but as yet to be applied to plants. Development of such methodologies will be important to elucidate the RNA targets of RBPs, uncovering RBP networks and the true scope of post-transcriptional gene regulation in plants.

**METHODS**

**Plant growth**

Arabidopsis thaliana ecotype Columbia (Col-0) seeds were sown on plates containing half-strength MS medium and stratified in the dark at 4°C for 48 hours. Plates, wrapped in aluminium foil, were then placed into 22°C growth cabinets for 4 days.

**mRNA interactome capture**

*UV-crosslinking*

For *in vivo* cross-linking (CL), plates of four-day old etiolated seedlings were placed on ice and irradiated in a Stratalinker (Stratagene) with 254 nm UV light at 150 mJ/cm2. The irradiation was performed three times with a one-minute pause in-between treatments. Seedlings were harvested immediately after irradiation and frozen in liquid N2. Seventy plates (h x w: 90 x 15 mm) were used per replicate.

*Lysis and oligo-dT capture*

Frozen tissue was then ground into fine powder in liquid N2 and resuspended in lysis buffer (20 mM Tris HCl pH 7.5, 500 mM LiCl, 0.5% LiDS, 1 mM EDTA, 0.02% NP-40, 5 mM DTT, 2.5% (w/v) PVP40, 1% (v/v) beta-mercaptoethanol, 1× EDTA-free Roche protease inhibitor). The lysate was cleared by passing it through a QIAshredder column (QIAGEN), centrifuging for 2 min at 14 000 r.p.m. Non-crosslinked seedlings (noCL) were processed side-by-side as a control. Aliquots from the lysate (input) were taken for quality controls (silver stain, western blot) and for protein identification by MS (referred to as total proteome).

RNA-protein complexes were isolated using oligo(dT)25 magnetic beads (beads from 500 µL original bead suspension, New England Biolabs), by incubating for 1 hour at 4°C on a rotator (lysate should not be frozen before oligo(dT) capture). Beads were collected on a magnet and washed twice with 1 mL of lysis buffer, followed by two washes with 1 ml of buffer I (20 mM pH 7.5 Tris HCl, 500 mM LiCl, 0.1% LiDS, 1 mM EDTA, 0.02% NP-40, 5 mM DTT), buffer II (20 mM pH 7.5 Tris HCl, 500 mM LiCl, 1 mM EDTA, 0.02% NP-40, 5 mM DTT) and buffer III (20 mM pH 7.5 Tris HCl, 200 mM LiCl, 1 mM EDTA, 5 mM DTT) for 5 min at 4°C on a rotator. RNA-protein complexes were eluted by incubating the beads with 200 µL of elution buffer (20 mM pH 7.5 Tris HCl, 1 mM EDTA) at 50°C for 3 min. After elution, oligo(dT) beads were reactivated in lysis buffer according to the manufacturer’s recommendations, a second round of oligo(dT) capture was performed for each sample and the two eluates were combined.

**RNase treatment and protein extraction**

The eluate was supplemented with ¼ volume of 5x RNase buffer (50 mM pH 7.5 Tris HCl, 750 mM NaCl, 0.25% NP-40, 2.5 mM DTT) and treated with 0.11 µg RNase A (Sigma) and 0.035 U of RNase T1 (Sigma) at 37°C for 1 hour. Proteins were extracted using TCA/acetone precipitation: 3 volumes of TCA/acetone solution (13.3% w/v TCA, 0.07% w/v DTT in acetone; chilled at -20°C) were added and samples were incubated at -20°C overnight. Samples were then centrifuged at 4°C for 15 min at 14,000 r.p.m., supernatant was removed and the protein pellet was washed by adding 1 mL washing solution (0.07% DTT in acetone; chilled at -20°C), vortexing and incubation at -80°C for at least 30 min. After repeating the centrifugation and wash steps, the protein pellet was dried in a SpeedVac for 2 min and resuspended in solubilisation buffer (8 M Urea, 0.5% SDS, 1% DTT, 35 mM HEPES, pH 7.5). Samples were sonicated twice for 30 sec, centrifuged for 10 min at 14,000 rpm and the supernatant containing the proteins was transferred to a new tube. An aliquot (20%) was taken for protein analysis (silver staining and western blot) and the remainder of the eluate was used for mass spectrometry.

**SDS-PAGE, Silver staining and Western blot**

For SDS-PAGE, protein samples in 1x NuPAGE LDS sample buffer and 1x NuPAGE reducing reagent (Thermo Fisher) were loaded on NuPAGE Novex 4-12% Bis-Tris protein gels (Thermo Fisher) and electrophoresed at 130 V for 2 hours in 1x MES buffer (Thermo Fisher). For silver staining, the gel was fixed in 50% methanol and 5% acetic acid for 30 min, followed by a wash with 50% ethanol and a second wash with 30% ethanol, each for 5 min. After a wash with water for 10 min, the gel was sensitized with 0.02% sodium thiosulfate for 60 sec and washed three times with water for 30 sec each. The gel was then placed in silver solution (6 mM silver nitrate, 0.0185% formaldehyde) for 20 min followed by three washes with water for 30 sec each. The gel was developed with 2% sodium carbonate, 0.0185% formaldehyde and 0.0004% sodium thiosulfate. The developing reaction was stopped with 5% acetic acid. All solutions were prepared freshly and all procedures were performed on a rocking platform at room temperature.

When restain was necessary, the gel was incubated in destain solution (prepared by mixing two solutions, 30 mM potassium ferricyanide and 100 mM sodium thiosulfate, at equal volume immediately prior to use) until gels were clear again, typically this occurred within 10 min. The gel was then washed with water to remove any yellow color. The gel was then restained as described above.

For western blot, gels were electroblotted onto nitrocellulose membrane (GE Healthcare Life Sciences). The membrane was blocked in 5% non-fat milk in 1×PBST (1×PBS with 0.2% Tween-20) for 30 min at room temperature, followed by incubation with primary antibodies in 5% non-fat milk in 1×PBST on a rotating wheel at 4 °C overnight. The membrane was then washed three times with 1×PBST, for 5 min each at room temperature. The membrane was incubated with secondary antibody in 5% non-fat milk in 1×PBST for 1 hr at room temperature. Protein signals were detected using Super Signal Femto chemiluminescent reagent (Pierce), visualized on ImageQuant LAS 4000 system (GE Healthcare Life Sciences). Primary antibodies used were anti-CP29A (kind gift from Christian Schmitz-Linneweber), anti-AGO1 (Agrisera) and anti-SAL1 (kind gift from Barry Pogson). An anti-rabbit HRP-conjugated antibody (AP132P, EMD Millipore) was used as secondary antibody.

**Mass spectrometry for mRNA interactome samples**

*Sample preparation using SP3*

For reduction and alkylation of the samples, 1 µL 200 mM DTT in 200 mM HEPES was added to the eluates followed by incubation at 56°C for 30 min. After placing the samples on ice for 2 min, 2 µL of 400 mM iodoacetamide (IAA) in 200 mM HEPES (pH 8.5) was added and samples were incubated at 24°C for 30 min in the dark. Then, 2 µL of 200 mM DTT was added followed by incubation at RT for 5 min.

For protein clean up and digestion, 2 µL of Sera-Mag bead mix (Thermo Scientific) were added to the eluates followed by the addition of 5 µL 5% formic acid. After ensuring that samples were acidic, acetonitrile was added to a final concentration of 50% and the samples were incubated for 8 min at RT. Samples were then placed on a magnetic rack and incubated for 2 min at RT. The supernatant was removed and samples were washed by adding 200 µL 70% ethanol and incubation for 15 sec on the magnetic rack. The wash was repeated once followed by the addition of 180 µL acetonitrile and incubation on the magnetic rack for 15 sec. The supernatant was removed and samples were air-dried for 30 sec. Samples were then taken off the rack and digested by adding 800 ng of trypsin in 50mM HEPES (pH8) and incubation at 37°C for 14 hours.

After the digest, beads were resuspended by pipetting and samples were placed on a magnetic rack. Dimethyl-labeling was performed by adding 1 µL of formaldehyde (CH2O for light, 13CD2O for heavy) and 1 µL of sodium cyanoborohydride (NaBH3CN for light, NaBD3CN for heavy) followed by incubation on a magnetic rack for 30 min at RT. 1 µL of the respective formaldehyde and sodium cyanoborohydride solutions were added again for more efficient labeling and samples were incubated for another 30 min at RT on the magnetic rack. After that, 1 µL of quench mix was added and samples were incubated for 5 min at RT. Samples were then taken off the magnetic rack and beads were resuspended by pipetting. Acetonitrile was added to the samples to a final percentage of 95% or higher, samples were mixed by pipetting followed by incubation off the magnetic rack for 8 min at RT and further 2 min on the magnetic rack at RT. After removal of the supernatant, samples were washed by adding 180 µL of acetonitrile and incubation for 15 sec on magnetic rack. The wash was repeated once, supernatant was removed and beads were air-dried for 30 sec. The beads were then resuspended in 9 µL of 2% DMSO and sonicated in a water bath for 5 min. Finally, samples were placed on a magnet, the supernatant was recovered to a new tube, acidified with 1 µL of 10% formic acid, and used for MS.

Cysteine’s were reduced with 1 ul of 200 mM DTT (Biomol) in 50 mM HEPES for 30 min. at 57 °C. The sample was cooled to 24 °C and 2 μL of 400 mM IAA (Merck) dissolved in 50 mM HEPES was added and incubated in dark for 30 min. at 24 °C. A novel protocol using paramagnetic beads, termed Single-Pot Solid-Phase-enhanced Sample Preparation (SP3) described by Hughes et al., 2014 was used to prepare the samples for LC-MS/MS. The proteins were digested using 800ng trypsin (Promega) in 50mM HEPES (pH8) and incubated at 37°C for 14 hours. Peptides were captured on magnetic beads, released, dimethyl labelled (Boersema et al., 2009) and recaptured again as described in Hughes et al., 2014.

(Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJR (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc 4: 484 – 494)

*LC-MS/MS analysis*

Samples were analyzed on an LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Scientific) coupled to a nanoAcquity UPLC system (Waters). Peptides were loaded onto a trapping column (nanoAcquity Symmetry C18, 5 μm, 180 μm × 20 mm) at a flow rate of 15 μL/min with solvent A (0.1% formic acid). Peptides were separated over an analytical column (nanoAcquity BEH C18, 1.7 μm, 75 μm × 200 mm) at a constant flow of 0.3 μL/min using the following gradient: 3% solvent B (Acetonitrile, 0.1% formic acid) for 10 min, 7-25% solvent B within 210 min, 25-40% solvent B within 10 min, 85% solvent B for 10 min. Peptides were introduced into the mass spectrometer using a Pico-Tip Emitter (360 μm outer diameter × 20 μm inner diameter, 10 μm tip, New Objective). MS survey scans were acquired from 300-1700 m/z at a nominal resolution of 30,000. The 15 most abundant peptides were isolated within a 2 Da window and subjected to MS/MS sequencing using collision-induced dissociation in the ion trap (activation time 10 msec, normalized collision energy 40%). Only 2+/3+ charged ions were included for analysis. Precursors were dynamically excluded for 30 sec (exclusion list size was set to 500).

**Mass spectrometry for the input proteome of etiolated seedlings**

*Sample preparation*

Proteins were extracted from aliquots (500 µL) saved from the input by TCA/acetone extraction as described above, and protein pellets were resuspended in 40 µL of 50 mM Tris-HCl (pH 8), 1% SDS, 1x Protease inhibitor. Samples were digested with 2 µL (25 Units/µL) benzonase for 45 min at 37°C. Samples were prepared for MS using SP3 as described above, except that the protein digestion was performed with trypsin/LysC, at 2 µg per sample at 37°C overnight.

*High pH reverse phase offline fractionation*

Offline high pH reverse phase fractionation was performed using an Agilent 1200 Infinity high-performance liquid chromatography (HPLC) system equipped with a quaternary pump, degasser, variable wavelength UV detector (set to 254 nm), peltier-cooled autosampler, and fraction collector (both set at 10 °C for all samples). The column was a Gemini C18 column (3 μm, 110 Å, 100 x 1.0 mm, Phenomenex) with a Gemini C18, 4 x 2.0 mm SecurityGuard (Phenomenex) cartridge as a guard column. The solvent system consisted of 20 mM ammo-nium formate (pH 10.0) as mobile phase (A) and 100% acetonitrile as mobile phase (B). The separation was accomplished at a mobile phase flow rate of 0.1 mL/min using the following linear gradient: 100% A for 2 min, from 100% A to 35% B in 59 min, to 85% B in a further 1 min, and held at 85% B for an additional 15 min, before returning to 100% A and re-equilibration for 13 min. Thirty-two fractions were collected along with the LC separation that were subsequently pooled into 10 fractions. Pooled fractions were dried under vacuum centrifugation, reconstituted in 10 μL 0.1% formic acid and then stored at -80 °C until LC-MS analysis.

*LC-MS Analysis for total proteome*

Peptides in the pooled fractions were separated using the nanoAcquity ultra performance liquid chromatography (UPLC) system (Waters) fitted with a trapping (nanoAcquity Sym-metry C18, 5 μm, 180 μm x 20 mm) and an analytical column (nanoAcquity BEH C18, 1.7 μm, 75 μm x 200 mm). The outlet of the analytical column was coupled directly to a LTQ (linear trap quadrupole) Orbitrap Velos Pro (Thermo Fisher Scientific) using the Proxeon nanospray source. Solvent A was water, 0.1% formic acid and solvent B was acetonitrile, 0.1% formic acid. The samples (7.5 μL out of 10 μL for total proteome analysis) were loaded with a constant flow of solvent A at 5 μL/min onto the trapping column. Trapping time was 6 min. Peptides were eluted via the analytical column with a constant flow of 0.3 μL/min. During the elution step, the percentage of solvent B increased in a linear fashion from 3% to 7% in 10 min, then increased to 25% in 100 min and finally to 40% in a further 10 min. The peptides were introduced into the mass spectrometer (Orbitrap Velos, Thermo) via a Pico-Tip Emitter 360 μm OD x 20 μm ID; 10 μm tip (New Objective) and a spray voltage of 2.2 kV was applied. The capillary temperature was set at 300 °C. Full scan MS spectra with mass range 300-1700 m/z were acquired in profile mode in the FT (Fourier transform) with resolution of 30000. The filling time was set at maximum of 500 ms with limitation of 1.0 x 106 ions. The most intense ions (up to 15) from the full scan MS were selected for sequencing in the LTQ. Normalized collision energy of 40% was used, and the fragmentation was performed after accumulation of 3.0 x104 ions or after filling time of 100 ms for each precursor ion (whichever occurred first). MS/MS data was acquired in centroid mode. Only multiply charged (2+, 3+, 4+) precursor ions were selected for MS/MS. The dynamic exclusion list was restricted to 500 entries with maximum retention period of 30 s and relative mass window of 10 ppm. In order to improve the mass accuracy, a lock mass correction using the ion (m/z 445.12003) was applied.

**Peptide identification and quantification**

Raw data were processed using MaxQuant (version 1.4.1.2) (Cox and Mann, 2008). MS/MS spectra were searched against the UniProt Arabidopsis database (input proteome and interactome capture: version 05/06/2015 including 54193 entries) concatenated to a database containing protein sequences of common contaminants. Enzyme specificity was set to trypsin/P, allowing a maximum of two missed cleavages. Cysteine carbamidomethylation was set as fixed modification, and methionine oxidation and protein N-terminal acetylation were used as variable modifications. For the mRNA interactome study, the required modifications for the dimethyl labelling were added as variable modification (DimethyLys0, DimethyNter0, DimethylLys8, DimethylNter8). The minimal peptide length was set to six amino acids and a minimum of one unique peptide was required for the identification. The mass tolerances were set to 20 ppm for the first search, 6 ppm for the main search and 0.5 Da for product ion masses. False discovery rates (FDR) for peptide and protein identification were set to 1%. Match between runs (time window 2 min) and re-quantify options were enabled. Protein quantification was based on razor and unique peptides with a CL/noCL ratio count of at least 2.

**Definition of mRNA interactome proteins**

Statistical analysis for CL/noCL enrichment of protein groups quantified in at least two out of three biological replicates was performed using an empirical Bayes moderated t-test within the R/Bioconductor package limma (Smyth, 2004). P-values were adjusted for multiple testing using the method of Benjamini-Hochberg. The UniProt accession numbers of each protein group were converted into Arabidopsis gene IDs. Where multiple gene IDs applied, the gene ID corresponding to the majority protein ID were used. Proteins with a CL/noCL enrichment >0 at an FDR below 1% were defined as At-RBPs, whereas proteins with an FDR > 1% as well as proteins where no CL/noCL ratio could be determined were defined as candidate At-RBPs.

**Bioinformatic analyses**

*Gene Ontology analyses*

TAIR (version 10) ATH\_GO\_GOSLIM.txt .gz (version 2015-08-02) downloaded from <https://www.arabidopsis.org/> was used for Gene Ontology annotation. Enrichment of GOMF, GOBP and GOCC categories was analyzed for interactome proteins and candidate binders compared to the background of proteins identified from the total proteome using 'fisher.test' (Fisher’s exact test), and multiple testing was performed using 'p.adjust' with Benjamini & Hochberg (BH) method in R package (R Core Team, 2015). GO terms with BH p-value <0.05 was defined as enriched/depleted.

*Classification of RNA biology status and protein domains*

The RNA biology status and RBDs of At-RBPs and candidate At-RBPs was classified as described in Beckmann et al., (2015).

*Analysis of biophysical properties and sequence features*

Analyses of disordered regions, length of proteins, hydrophobicity and amino acid composition were performed as described in Castello et al., 2012. Hydrophobicity for each amino acid residue was obtained from https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/midas/hydrophob.html. Isoelectric Point was calculated using the 'IsoelectricPoint' module in Biopython package (<http://biopython.org/DIST/docs/api/Bio.SeqUtils.IsoelectricPoint-pysrc.html>). Distribution biases for the sequence features are evaluated using R packages, adk.test (Anderson-Darling k-sample test) and ks.test (two-sample Kolmogorov-Smirnov test).

*Conservation of At-RBPs*

The list of predicted orthologues between plant and human, mouse and yeast were obtained from InParanoid (version 8.0) (Sonnhammer and Östlund, 2015). The list of two-way predictions (*A*.*thaliana*-*H*.*sapiens*, *A*.*thaliana*-*M*.*musculus*, and *A*.*thaliana*-*S*.*cerevisiae)* were downloaded from <http://inparanoid.sbc.su.se/download/>. InParanoid uses a clustering method based on genome-wide pairwise sequence similarity matches to identify putative orthologous proteins between two species, and predicts ortholog groups, where each group contains one (the highest sequence similarity matches) or more (with high pairwise similarity matches relative to the best pair) pairs those including in-paralogs within defined cut-off value (Sonnhammer and Östlund, 2015). We included all pairs in each group, which may include more than one combination (in-paralogs), i.e. all plant proteins which are predicted orthologs to either human, mouse or yeast. The list of interactome proteins are obtained from Baltz et al., (2012) (HEK293, *H*.*sapiens*), Kwon et al., (2013) (mESC, *M*.*musculus*), Liao et al., (2016) (HL-1, *M*.*musculus*) and Beckmann et al., (2015) (HuH-7, *H.sapiens*; *S*.*cerevisiae*).

*Meta-transcript analysis of Arabidopsis m6A and poly(A) sites*

We use RNAModR [Evers et al., submitted to Bioinformatics] to perform a meta-transcript analysis of published m6A (Wan et al., 2015) and polyadenylation sites (PAC sites from Wu et al., 2011; cleavage sites from Sherstnev et al., 2012). In a first step, all reported m6A and polyadenylation sites are mapped to a custom reference transcriptome; the latter is constructed by collapsing all TAIR10-based transcript isoforms per gene, and keeping the transcript isoform with the longest coding sequence (CDS) and longest CDS-adjoining 5’/3’ untranslated regions. RNAModR then evaluates transcript-level spatial enrichment of m6A and suitable null sites relative to polyadenylation sites by using multiple Fisher’s exact tests to calculate odds ratios



and test against a null hypothesis corresponding to OR = 1. Here *d* corresponds to the minimum distance between an m6A or null site and all polyadenylation sites in the same transcript



The total number of m6A and null sites, respectively, at distance *d* is denoted by *N*.

Resulting *p*-values are corrected for multiple hypotheses testing using the method of Benjamini and Hochberg (Benjamini and Hochberg, 1995). Suitable null sites are constructed within RNAModR from the genomic loci of all non-methylated adenosines in transcripts that contain at least one m6A site.

**Supplemental Data files**

Supplemental Figure 1: Optimization of the Arabidopsis mRNA interatome capture protocol.

Supplemental Figure 2. GO enrichment analysis of At-RBPs and candidate At-RBPs.

Supplemental Figure 3. Conservation of RBPs across kingdoms.

Supplemental Figure 4. Analysis of amino acid enrichment/depletion in the Arabidopsis mRNA interactome.

Supplemental Table 1: Zinc-finger proteins not associated with RNA-binding identified by mRNA interactome capture.

Supplemental Table 2: Examples of non-canonical RBPs identified by mRNA interactome capture.

Supplemental Data Set: Arabidopsis mRNA interactome.

**AUTHOR CONTRIBUTIONS**

M.R, Y.L, M.W.H., T.P and A.A.M designed the project; M.R. and Y.L performed the majority of experiments with assistance from A.-M.A; M.Rettel carried out the MS experiments and initial data analysis, C.R. performed the bioinformatic analysis of the MS data; M.E. carried out the meta-transcript analysis of m6A and poly(A) sites R.H gave conceptual advice and assisted with experiments; M.R, Y.L, T.P and A.A.M wrote the manuscript with input from all authors. All authors approved the manuscript.

**ACKNOWLEDGEMENTS**

We thank Christian Schmitz-Linneweber and Barry Pogson for kindly providing the CP29A and SAL1 antibody, respectively. We also thank the EMBL Proteomics core facility for the MS analysis. This work was supported by an EMBO short-term fellowship and an International ANU PhD scholarship to M.R., a John Stocker Postdoctoral Fellowship (PF14-079) from the Science and Industry Endowment Fund to Y.L., an ERC Advanced Grant (ERC-2011-ADG\_20110310) to M.W.H., and a grant from the National Health and Medical Research Council of Australia (#1045417) to T.P. and M.W.H.

**REFERENCES**

Ambrosone A, Costa A, Leone A, Grillo S. (2012) Beyond transcription: RNA-binding proteins as emerging regulators of plant response to environmental constraints. Plant Sci. 182:12-8.

Aravind L, Iyer LM, Anantharaman V. (2003) The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. Genome Biol. 4(10):R64.

Au PC, Helliwell C, Wang MB. (2014) Characterizing RNA-protein interaction using cross-linking and metabolite supplemented nuclear RNA-immunoprecipitation. Mol Biol Rep. 41(5):2971-7.

Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cel l 46(5):674-90.

Baumberger N, Baulcombe DC. (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci U S A. 102(33): 11928–11933. (page number format is different from others)

Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume AM, Schwarzl T, Curk T, Foehr S, Huber W, Krijgsveld J, Hentze MW (2015) The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat Commun. 6:10127.

Bell SD, Botting CH, Wardleworth BN, Jackson SP, White MF. (2002) The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. Science. 296(5565):148-51.

Benjamini Y, Hochberg Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Statist. Soc. B 57, 289-300.

Biela A, Grote K, Otto B, Hoth S, Hedrich R, Kaldenhoff R. (1999) The Nicotiana tabacum plasma membrane aquaporin NtAQP1 is mercury-insensitive and permeable for glycerol. Plant J. 18(5):565-70.

Bodi Z, Zhong S, Mehra S, Song J, Graham N, Li H, May S, Fray RG. (2012). Adenosine Methylation in Arabidopsis mRNA is Associated with the 3' End and Reduced Levels Cause Developmental Defects. Front Plant Sci 3, 48.

Bono F, Ebert J, Lorentzen E, Conti E. (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell. 126(4):713-25.

Branco-Price C, Kaiser KA, Jang CJ, Larive CK, Bailey-Serres J. (2008) Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in Arabidopsis thaliana. Plant J. 56(5):743-55.

Bravo J, Aguilar-Henonin L, Olmedo G, Guzmán P. (2005) Four distinct classes of proteins as interaction partners of the PABC domain of Arabidopsis thaliana Poly(A)-binding proteins. Mol Genet Genomics. 272(6):651-65.

Bruggeman Q, Garmier M, de Bont L, Soubigou-Taconnat L, Mazubert C, Benhamed M, et al. (2014). The Polyadenylation Factor Subunit CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR30: A Key Factor of Programmed Cell Death and a Regulator of Immunity in Arabidopsis. Plant Physiology, 165(2), 732–746.

Bullock SL. (2011) Messengers, motors and mysteries: sorting of eukaryotic mRNAs by cytoskeletal transport. Biochem Soc Trans. 39(5):1161-5.

Burgess A, David R, Searle IR. (2016). Deciphering the epitranscriptome: A green perspective. Journal of Integrative Plant Biology. http://doi.org/10.1111/jipb.12483

Bush MS, Crowe N, Zheng T, Doonan JH (2015) The RNA helicase, eIF4A-1, is required for ovule development and cell size homeostasis in Arabidopsis. Plant J. 84(5):989-1004.

Calabretta S, Richard S. (2015) Emerging Roles of Disordered Sequences in RNA-Binding Proteins. Trends Biochem Sci. 40(11):662-72.

Cappadocia L, Maréchal A, Parent JS, Lepage E, Sygusch J, Brisson N. (2010) Crystal structures of DNA-Whirly complexes and their role in Arabidopsis organelle genome repair. Plant Cell. 22(6):1849-67.

Casal JJ, Candia AN, Sellaro R. (2014) Light perception and signalling by phytochrome A. J Exp Bot. 65(11):2835-45.

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149(6):1393-406.

Cha JY, Ahn G, Kim JY, Kang SB, Kim MR, Su'udi M, Kim WY, Son D. (2013) Structural and functional differences of cytosolic 90-kDa heat-shock proteins (Hsp90s) in Arabidopsis thaliana. Plant Physiol Biochem. 70:368-73.

Chakrabarti, M., & Hunt, A. G. (2015). CPSF30 at the Interface of Alternative Polyadenylation and Cellular Signaling in Plants. Biomolecules 5(2), 1151–1168.

Chang JH, Tong L. (2012) Mitochondrial poly(A) polymerase and polyadenylation. Biochim Biophys Acta. 1819(9-10):992-7.

Chen F, Shi X, Chen L, Dai M, Zhou Z, Shen Y, Li J, Li G, Wei N, Deng XW. (2012) Phosphorylation of FAR-RED ELONGATED HYPOCOTYL1 is a key mechanism defining signaling dynamics of phytochrome A under red and far-red light in Arabidopsis. Plant Cell. 24(5):1907-20.

Chen F, Li B, Li G, Charron JB, Dai M, Shi X, Deng XW. (2014) Arabidopsis Phytochrome A Directly Targets Numerous Promoters for Individualized Modulation of Genes in a Wide Range of Pathways. Plant Cell. 26(5):1949-66.

Chi SW, Zang JB, Mele A, Darnell RB. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature. 460 (7254):479-86.

Ciftci-Yilmaz S, Mittler R. (2008) The zinc finger network of plants. Cell Mol Life Sci. 65(7-8):1150-60.

Colcombet J, Lopez-Obando M, Heurtevin L, Bernard C, Martin K, Berthomé R, Lurin C. (2013) Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. RNA Biol. 10(9):1557-75.

Dai Y, Li W, An L. (2016) NMD mechanism and the functions of Upf proteins in plant. Plant Cell Rep. 35(1):5-15.

Delaney, K.J., Xu, R., Zhang, J., Li, Q.Q., Yun, K.Y., Falcone, D.L., and Hunt, A.G. (2006). Calmodulin interacts with and regulates the RNA-binding activity of an Arabidopsis polyadenylation factor subunit. Plant Physiol. 140: 1507–21.

Dreyfuss G, Adam SA, Choi YD (1983) Physical Change in Cytoplasmic Messenger Ribonucleoproteins in Cells Treated with Inhibitors of mRNA Transcription. Molecular and cellular Biology.415-423.

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, Sorek R,Rechavi G. (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 485(7397):201-6.

Falahzadeh K, Banaei-Esfahani A, Shahhoseini M. (2015) The potential roles of actin in the nucleus. Cell J. 17(1):7-14.

Folta KM, Kaufman LS. (2003) Phototropin 1 is required for high-fluence blue-light-mediated mRNA destabilization. Plant Mol Biol. 51(4):609-18.

Forterre P, Confalonieri F, Knapp S (1999) **Identification of the gene encoding archeal-specific DNA-binding proteins of the Sac10b family.** Mol Microbiol.**32:**669–670.

Francischini CW, Quaggio RB. (2009) Molecular characterization of Arabidopsis thaliana PUF proteins--binding specificity and target candidates. FEBS J. 276(19):5456-70.

Fray RG, Simpson GG. (2015) The Arabidopsis epitranscriptome. Curr Opin Plant Biol. 27:17-21.

Fribourg S, Braun IC, Izaurralde E, Conti E. (2001) Structural basis for the recognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor. Mol Cell. 8(3):645-56.

Fujiki Y, Yoshikawa Y, Sato T, Inada N, Ito M, Nishida I, Watanabe A. (2001) Dark-inducible genes from Arabidopsis thaliana are associated with leaf senescence and repressed by sugars. Physiol Plant. 2001 Mar;111(3):345-352.

Gagnon JA, Mowry KL. (2011) Molecular motors: directing traffic during RNA localization. Crit Rev Biochem Mol Biol. 46(3):229-39.

Gaspar M., Bousser A., Sissoëff I., Roche O., Hoarau J. & Mahe A. (2003) Cloning and characterization of ZmPIP1–5b, an aquaporin transporting water and urea. PlantScience 165, 21–31.

Gissot M, Walker R, Delhaye S, Alayi TD, Huot L, Hot D, Callebaut I, Schaeffer-Reiss C, Dorsselaer AV, Tomavo S. (2013) Toxoplasma gondii Alba proteins are involved in translational control of gene expression. J Mol Biol. 425(8):1287-301.

Glisovic T, Bachorik JL, Yong J, Dreyfuss G. (2008) RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. 582(14):1977-86.

Gosai SJ, Foley SW, Wang D, Silverman IM, Selamoglu N, Nelson AD, Beilstein MA, Daldal F, Deal RB, Gregory BD. (2015) Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the Arabidopsis nucleus. Mol Cell. 57(2):376-88.

Greenberg JR (1979) Ultraviolet light-induced crosslinking of mRNA to proteins. Nucleic Acids Res. 6, pp. 715–32.

Guo R, Xue H, Huang L. (2003) Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo. Mol Microbiol. 50(5):1605-15.

Hall TM. (2005) Multiple modes of RNA recognition by zinc finger proteins. Curr Opin Struct Biol. 15(3):367-73.

Hecht V, Stiefel V, Delseny M, Gallois P. (1997) A new Arabidopsis nucleic-acid-binding protein gene is highly expressed in dividing cells during development. Plant Mol Biol. 34(1):119-24.

Henics T, Nagy E, Oh HJ, Csermely P, von Gabain A, Subjeck JR. (1999) Mammalian Hsp70 and Hsp110 proteins bind to RNA motifs involved in mRNA stability. J Biol Chem. 274(24):17318-24.

Hofmann WA. (2009) Cell and molecular biology of nuclear actin. Int Rev Cell Mol Biol. 273:219–63.

Holm LM, Jahn TP, Møller AL, Schjoerring JK, Ferri D, Klaerke DA, Zeuthen T. (2005) NH3 and NH4+ permeability in aquaporin-expressing Xenopus oocytes. Pflugers Arch. 450(6):415-28.

Huang YW, Hu CC, Liou MR, Chang BY, Tsai CH, Meng M, Lin NS, Hsu YH. (2012) Hsp90 interacts specifically with viral RNA and differentially regulates replication initiation of Bamboo mosaicvirus and associated satellite RNA. PLoS Pathog. 8(5):e1002726.

Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. Mol Syst Biol. 10:757.

Huh SU, Kim MJ, Paek KH. (2013) Arabidopsis Pumilio protein APUM5 suppresses Cucumber mosaic virus infection via direct binding of viral RNAs. Proc Natl Acad Sci U S A. 110(2):779-84.

Huh SU, Paek KH. (2014) APUM5, encoding a Pumilio RNA binding protein, negatively regulates abiotic stress responsive gene expression. BMC Plant Biol. 14:75.

Hunt AG. (2014) The Arabidopsis polyadenylation factor subunit CPSF30 as conceptual link between mRNA polyadenylation and cellular signalling. Curr Opin Plant Biol. 21:128-32.

Hunt AG., Xing D, Li Q Q. (2012). Plant polyadenylation factors: conservation and variety in the polyadenylation complex in plants. BMC Genomics, 13(1), 641.

Jansen RP. (1999) RNA-cytoskeletal associations. FASEB J. 13(3):455-66.

Jelinska C, Conroy MJ, Craven CJ, Hounslow AM, Bullough PA, Waltho JP, Taylor GL, White MF (2005) Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. Structure. 13(7):963-71.

Jeong E, Kim H, Lee SW, Han K. (2003) Discovering the interaction propensities of amino acids and nucleotides from protein-RNA complexes. Mol Cells. 16(2):161-7.

Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang Y-G et al.(2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol 7:885-887.

Jung HJ, Park SJ and Kang H (2013) Regulation of RNA Metabolism in Plant Development and Stress Responses. J. Plant Biol 56:123-29.

Kadrmas JL, Beckerle MC. (2004) The LIM domain: from the cytoskeleton to the nucleus. Nat Rev Mol Cell Biol. 5(11):920-31.

Katahira J, Dimitrova L, Imai Y, Hurt E. (2015) NTF2-like domain of Tap plays a critical role in cargo mRNA recognition and export. Nucleic Acids Res. 43(3):1894-904.

Kawaguchi R, Girke T, Bray EA, Bailey-Serres J. (2004) Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in Arabidopsis thaliana. Plant J. 38(5):823-39.

Kim YO, Pan S, Jung CH, Kang H. (2007) A zinc finger-containing glycine-rich RNA-binding protein, atRZ-1a, has a negative impact on seed germination and seedling growth of Arabidopsis thaliana under salt or drought stress conditions. Plant Cell Physiol. 48(8):1170-81.

Kim JY, Park SJ, Jang B, Jung CH, Ahn SJ, Goh CH, Cho K, Han O, Kang H. (2007) Functional characterization of a glycine-rich RNA-binding protein 2 in Arabidopsis thaliana under abiotic stress conditions. Plant J. 50(3):439-51.

Kim VN, Kataoka N, Dreyfuss G (2001) Role of the nonsensemediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. Science 293(5536):1832–1836.

Krause K, Kilbienski I, Mulisch M, Rödiger A, Schäfer A, Krupinska K. DNA-binding proteins of the Whirly family in Arabidopsis thaliana are targeted to the organelles. FEBS Lett. 579(17):3707-12.

Kwak KJ, Kim YO, Kang H. (2005) Characterization of transgenic Arabidopsis plants overexpressing GR-RBP4 under high salinity, dehydration, or cold stress. J Exp Bot. 56(421):3007-16.

Kwon SC, Yi H, Eichelbaum K, Föhr S, Fischer B, You KT, Castello A, Krijgsveld J, Hentze MW, Kim VN. (2013) The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol. 20(9):1122-30.

Lambermon MH, Simpson GG, Wieczorek Kirk DA, Hemmings-Mieszczak M, Klahre U, Filipowicz W. (2000) UBP1, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. EMBO J. 19(7):1638–49.

Lange H, Sement FM, Canaday J, Gagliardi D. (2009) Polyadenylation-assisted RNA degradation processes in plants. Trends Plant Sci. 497-504.

Laohavisit A1, Davies JM. (2011) Annexins. New Phytol. 189(1):40-53.

Lau CK, Bachorik JL, Dreyfuss G. (2009) Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains. Nat Struct Mol Biol. 16(5):486-91.

Le Hir H, Andersen GR. (2008) Structural insights into the exon junction complex. Curr Opin Struct Biol. 18(1):112-9.

Lee HW, Park JH, Park MY, Kim J. (2014) GIP1 may act as a coactivator that enhances transcriptional activity of LBD18 in Arabidopsis. J Plant Physiol. 171(5):14-8.

Lejeune D, Delsaux N, Charloteaux B, Thomas A, Brasseur R. (2005) Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure. Proteins. 61(2):258-71.

Li D, Zhang H, Hong Y, Huang L, Li X, Zhang Y, Ouyang Z, Song F (2014) Genome-Wide Identification, Biochemical Characterization, and Expression Analyses of the YTH Domain-Containing RNA-Binding Protein Family in Arabidopsis and Rice. Plant Mol Biol Rep 32: 1169-86.

Li F, Zhao D, Wu J, Shi Y. (2014) Structure of the YTH domain of human YTHDF2 in complex with an m(6)A mononucleotide reveals an aromatic cage for m(6)A recognition. Cell Res. 24(12):1490-2.

Li W, Ma M, Feng Y, Li H, Wang Y, Ma Y, Li M, An F, Guo H. (2015) EIN2-directed translational regulation of ethylene signaling in Arabidopsis. Cell. 163(3):670-83.

Liu H, Qin C, Chen Z, Zuo T, Yang X, Zhou H, Xu M, Cao S, Shen Y, Lin H, He X, Zhang Y, Li L, Ding H, Lübberstedt T, Zhang Z, Pan G. (2014) Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. BMC Genomics. 15:25.

Loke JC, Stahlberg EA, Strenski DG, Haas BJ, Wood PC, Li QQ. (2005). Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures. Plant physiology 138, 1457-1468.

Lorković ZJ. (2009) Role of plant RNA-binding proteins in development, stress response and genome organization. Trends Plant Sci. 14(4):229-36.

Lorković ZJ, Wieczorek Kirk DA, Klahre U, Hemmings-Mieszczak M, Filipowicz W. (2000) RBP45 and RBP47, two oligouridylate-specific hnRNP-like proteins interacting with poly(A)+ RNA in nuclei of plant cells. RNA 6(11):1610-24.

Luo GZ, MacQueen A, Zheng G, Duan H, Dore LC, Lu Z, Liu J, Chen K, Jia G, Bergelson J, He C. (2014) Unique features of the m6A methylome in Arabidopsis thaliana. Nat Commun. 5:5630.

Luo S., Tong L. (2014) Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain.Proc. Natl. Acad. Sci. *USA*. 111: 13834–39.

**Ma X, Tang Z, Qin J, Meng Y. (2015) The use of high-throughput sequencing methods for plant microRNA research. RNA Biol. 12(7):709-19.**

**Macknight R, Bancroft I, Page T, Lister C, Schmidt R, Love L, Westphal L, Murphy G, Sherson S, Cobbett C, Dean C.** (1997). FCA, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. Cell 89**,** 737–45.

Mair GR, Lasonder E, Garver LS, Franke-Fayard BM, Carret CK, Wiegant JC, Dirks RW, Dimopoulos G, Janse CJ, Waters AP. (2010) Universal features of post-transcriptional gene regulation are critical for Plasmodium zygote development. PLoS Pathog. 6(2):e1000767.

Mani J, Güttinger A, Schimanski B, Heller M, Acosta-Serrano A, Pescher P, Späth G, Roditi I. (2011) Alba-domain proteins of Trypanosoma brucei are cytoplasmic RNA-binding proteins that interact with the translation machinery. PLoS One. 6(7):e22463.

Maréchal A, Parent JS, Véronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. (2009) Whirly proteins maintain plastid genome stability in Arabidopsis. Proc Natl Acad Sci U S A. 106(34):14693-8.

Matia-González AM, Laing EE, Gerber AP. (2015) Conserved mRNA-binding proteomes in eukaryotic organisms. Nat Struct Mol Biol. 22(12):1027-33.

Maurel C, Boursiac Y, Luu DT, Santoni V, Shahzad Z, Verdoucq L. (2015) Aquaporins in Plants. Physiol Rev. 2015 Oct;95(4):1321-58.

Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enríquez P, Binder BM, Heber S, Stepanova AN, Alonso JM. (2015) Gene-specific translation regulation mediated by the hormone-signaling molecule EIN2. Cell. 163(3):684-97.

Mielecki D, Zugaj DŁ, Muszewska A, Piwowarski J, Chojnacka A, Mielecki M, Nieminuszczy J, Grynberg M, Grzesiuk E (2012). Novel AlkB dioxygenases — alternative models for in silico and in vivo studies. PLoS ONE 7:e30588.

Mitchell SF, Jain S, She M, Parker R. (2013) Global analysis of yeast mRNPs. Nat Struct Mol Biol. 20(1):127-33.

Nguyen CD, Mansfield RE, Leung W, Vaz PM, Loughlin FE, Grant RP, Mackay JP. (2011) Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA. J Mol Biol. 407(2):273-83.

Nicolaï M, Roncato MA, Canoy AS, Rouquié D, Sarda X, Freyssinet G, Robaglia C. (2006) Large-scale analysis of mRNA translation states during sucrose starvation in arabidopsis cells identifies cell proliferation and chromatin structure as targets of translational control. Plant Physiol. 141(2):663-73.

Nobuta K, McCormick K, Nakano M, Meyers BC. (2010) Bioinformatics analysis of small RNAs in plants using next generation sequencing technologies. Methods Mol Biol. 592:89-106.

Nolte C, Staiger D. (2015) RNA around the clock - regulation at the RNA level in biological timing. Front Plant Sci. 6:311.

Nyikó T, Kerényi F, Szabadkai L, Benkovics AH, Major P, Sonkoly B, Mérai Z, Barta E, Niemiec E, Kufel J, Silhavy D. (2013) Plant nonsense-mediated mRNA decay is controlled by different autoregulatory circuits and can be induced by an EJC-like complex. Nucleic Acids Res. 41(13):6715-28.

Ok SH, Jeong HJ, Bae JM, Shin JS, Luan S, Kim KN. (2005) Novel CIPK1-associated proteins in Arabidopsis contain an evolutionarily conserved C-terminal region that mediates nuclear localization. Plant Physiol. 139(1):138-50.

Paik I, Yang S, Choi G. (2012) Phytochrome regulates translation of mRNA in the cytosol. Proc Natl Acad Sci U S A. 109(4):1335-40.

Papuga J, Hoffmann C, Dieterle M, Moes D, Moreau F, Tholl S, Steinmetz A, Thomas C. (2010) Arabidopsis LIM proteins: a family of actin bundlers with distinct expression patterns and modes of regulation. Plant Cell. 22(9):3034-52.

Peña EJ1, Heinlein M. (2013) RNA transport during TMV cell-to-cell movement. Front Plant Sci. 3:193.

Pisarev AV, Kolupaeva VG, Yusupov MM, Hellen CU, Pestova TV (2008). Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. EMBO J. 27(11):1609-21.

Prikryl J, Watkins KP, Friso G, van Wijk KJ, Barkan A. (2008) A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. Nucleic Acids Res. 36(16):5152-65.

Quigley F, Rosenberg JM, Shachar-Hill Y, Bohnert HJ. (2002) From genome to function: the Arabidopsis aquaporins. Genome Biol. 3(1): research0001.1–research0001.17.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Raab S, Toth Z, de Groot C, Stamminger T, Hoth S. (2006) ABA-responsive RNA-binding proteins are involved in chloroplast and stromule function in Arabidopsis seedlings. Planta. 224(4):900-14.

Ribbeck K, Lipowsky G, Kent HM, Stewart M, Görlich D. (1998) NTF2 mediates nuclear import of Ran. EMBO J. 17(22):6587-98.

Schomburg FM, Patton DA, Meinke DW, Amasino RM (2001) FPA, a Gene Involved in Floral Induction in Arabidopsis, Encodes a Protein Containing RNA-Recognition Motifs. Plant Cell 13(6): 1427–36.

Schumann U, Shafik A, Preiss T. (2016). METTL3 Gains R/W Access to the Epitranscriptome. Molecular Cell, 62(3), 323–324. http://doi.org/10.1016/j.molcel.2016.04.024

Schuster G, Stern D. (2009) RNA polyadenylation and decay in mitochondria and chloroplasts. Prog Mol Biol Transl Sci. 85:393-422.

Schwartz S. (2016) Cracking the epitranscriptome. (2016). RNA 22(2), 169–174. http://doi.org/10.1261/rna.054502.115

Shaikhali J. (2015) GIP1 protein is a novel cofactor that regulates DNA-binding affinity of redox-regulated members of bZIP transcription factors involved in the early stages of Arabidopsis development. Protoplasma. 252(3):867-83.

Sherstnev A, Duc C, Cole C, Zacharaki V, Hornyik C, Ozsolak F, Milos PM, Barton GJ, Simpson GG. (2012). Direct sequencing of Arabidopsis thaliana RNA reveals patterns of cleavage and polyadenylation. Nature structural & molecular biology 19, 845-852.

Shikata H, Hanada K, Ushijima T, Nakashima M, Suzuki Y, Matsushita T. (2014) Phytochrome controls alternative splicing to mediate light responses in Arabidopsis. Proc Natl Acad Sci U S A. 111(52): 18781–86.

Singh G, Pratt G, Yeo GW, Moore MJ (2015) The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu Rev Biochem. 84:325-54.

Sikorski, P. J., Zuber, H., Philippe, L., Sement, F. M., Canaday, J., Kufel, J., et al. (2015). Distinct 18S rRNA precursors are targets of the exosome complex, the exoribonuclease RRP6L2 and the terminal nucleotidyltransferase TRL in Arabidopsis thaliana. Plant J, 83(6), 991–1004.

Silverman IM, Li F, Gregory BD. (2013) Genomic era analyses of RNA secondary structure and RNA-binding proteins reveal their significance to post-transcriptional regulation in plants. Plant Sci. 205-206:55-62.

Sonnhammer EL, Östlund G. (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. 43(Database issue):D234-9.

Subota I, Rotureau B, Blisnick T, Ngwabyt S, Durand-Dubief M, Engstler M, Bastin P. (2011) ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. Mol Biol Cell. 22(22):4205-19.

Theillet F-X, Kalmar L, Tompa P, Han K-H, Selenko P, Dunker AK, Daughdrill GW, Uversky VN (2013) The alphabet of intrinsic disorder, Intrinsically Disordered Proteins, 1:1, e24360, DOI: 10.4161/idp.24360.

Thandapani P, O'Connor TR, Bailey TL, Richard S. (2013) Defining the RGG/RG motif. Mol Cell. 50(5):613-23.

Thomas PE, Wu X, Liu M, Gaffney B, Ji G, Li QQ, Hunt AG (2012). Genome-wide control of polyadenylation site choice by CPSF30 in Arabidopsis. Plant Cell. 24(11):4376-88.

Tremousaygue D, Manevski A, Bardet C, Lescure N, Lescure B. (1999) Plant interstitial telomere motifs participate in the control of gene expression in root meristems. Plant J. 1999 Dec;20(5):553-61.

Tripurani SK, Nakaminami K, Thompson KB, Crowell SV, Guy CL, Karlson DT (2011) Spatial and temporal expression of cold-responsive DEAD-box RNA helicases reveals their functional roles during embryogenesis in Arabidopsis thaliana. Plant Mol. Biol. Rep. 29: 761-68.

Tsuda K1, Tsuji T, Hirose S, Yamazaki K. (2004) Three Arabidopsis MBF1 homologs with distinct expression profiles play roles as transcriptional co-activators. Plant Cell Physiol. 2004 Feb;45(2):225-31.

Uehlein N, Lovisolo C, Siefritz F, Kaldenhoff R. The tobacco aquaporin NtAQP1 is a membrane CO2 pore with physiological functions. Nature. 425(6959):734-7.

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302 1212–1215.

Vandevenne M, O'Connell MR, Helder S, Shepherd NE, Matthews JM, Kwan AH, Segal DJ, Mackay JP. (2014) Engineering specificity changes on a RanBP2 zinc finger that binds single-stranded RNA. Angew Chem Int Ed Engl. 53(30):7848-52.

Vaucheret H. (2008) Plant ARGONAUTES. Trends Plant Sci. 13(7):350-8.

Wagenmakers AJM, Reinders RJ, Van venrooij WJ (1980) Cross-linking of mRNA to Proteins by Irradiation of Intact Cells with Ultraviolet Light. Eur. J. Biochem. 112, 323-30.

Wan Y, Tang K, Zhang D, Xie S, Zhu X, Wang Z, Lang Z. (2015). Transcriptome-wide high-throughput deep m(6)A-seq reveals unique differential m(6)A methylation patterns between three organs in Arabidopsis thaliana. Genome biology 16, 272.

Wang W, Vinocur B, Shoseyov O, Altman A. (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci. 9(5):244-52.

Wang Y, Mercier R, Hobman TC, LaPointe P. (2013) Regulation of RNA interference by Hsp90 is an evolutionarily conserved process. Biochim Biophys Acta. 1833(12):2673-81.

Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, Ren B, Pan T, He C. (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. Nature. 505(7481):117-20.

Warner, J. R., & McIntosh, K. B. (2009). How common are extraribosomal functions of ribosomal proteins? Molecular Cell, 34(1), 3–11.

Williams RM, Obradovi Z, Mathura V, Braun W, Garner EC, Young J, Takayama S, Brown CJ, Dunker AK. (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder. Pac Symp Biocomput. 2001:89-100.

Wright PE, Dyson HJ. (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 16(1):18-29.

Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, Hunt AG (2011). Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. Proceedings of the National Academy of Sciences of the United States of America 108, 12533-12538.

Xu C, Liu K, Ahmed H, Loppnau P, Schapira M, Min J. (2015) Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. J Biol Chem. 290(41):24902-13.

Yang YC, Di C, Hu B, Zhou M, Liu Y, Song N, Li Y, Umetsu J, Lu ZJ. (2015) CLIPdb: a CLIP-seq database for protein-RNA interactions. BMC Genomics. 16:51.

Ye J, Xu M. (2012) Actin bundler PLIM2s are involved in the regulation of pollen development and tube growth in Arabidopsis. J Plant Physiol. 169(5):516-22.

Zhang J, Addepalli B, Yun K., Hunt AG, Xu R, Rao S, Li QQ, Falcone DL (2008). A polyadenylation factor subunit implicated in regulating oxidative signaling in Arabidopsis thaliana. PLoS One 3, e2410.

Zhang C, Zhang F. (2015) The Multifunctions of WD40 Proteins in Genome Integrity and Cell Cycle Progression. J Genomics. 3:40-50.

Zhang Y, Gu L, Hou Y, Wang L, Deng X, Hang R, Chen D, Zhang X, Zhang Y, Liu C, Cao X. (2015) Integrative genome-wide analysis reveals HLP1, a novel RNA-binding protein, regulates plant flowering by targeting alternative polyadenylation. Cell Res. 25(7):864-76.

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang C-M, Li CJ, Va˚ gbø CB, Shi Y, Wang W-L, Song S-H et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell 49:18-29.

Zhao Q, Leung S, Corbett AH, Meier I. (2006) Identification and characterization of the Arabidopsis orthologs of nuclear transport factor 2, the nuclear import factor of ran. Plant Physiol. 140(3):869-78.

Zhong S, Li H, Bodi Z, Button J, Vespa L, Herzog M, Fray RG. (2008) MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. Plant Cell. 20(5):1278-88.

**FIGURE LEGENDS**

**Figure 1**. mRNA interactome capture in Arabidopsis.

(A). Overview of the mRNA interactome capture procedure in Arabidopsis and categorization of interactome proteins.

(B-C). RNA-protein complexes from CL and noCL samples isolated by oligo(dT) capture were treated with RNases, separated by SDS-PAGE and analysed by silver staining (B) and western blot (C) alongside input samples. Results are representative of three independent interactome capture experiments.

(D-E). Annotations of At-RBPs and candidate At-RBPs compared to the input proteome according to functional characteristics (RNA biology) (D) and RNA-binding domains (RBDs) (E).

(F-G). The most significantly over- and under- represented Gene Ontology (GO) terms for molecular function in At-RBPs (F) and candidate At-RBPs (G), compared to the input proteome.

**Figure 2**. Biophysical and sequence features of captured protein sets.

(A-D). Density of protein length (A), proportion of amino acid residues in disordered regions (B), isoelectric point (pI) (C) and hydrophobicity (HI) (D) were analysed for At-RBPs (green), proteins from the input proteome annotated as ‘RNA-binding’ (light purple), At-RBPs with unknown RBD (light orange), candidate At-RBP (blue), and input proteome (red). The significance of differences between RBP subsets in panels A-D was tested by the Kolmogorov-Smirnov test. This showed that protein size distribution (A) does not differ between all five groups. Compared to input proteome, all four subsets are significantly different in disordered region (B), isoelectric point (C), and hydrophobicity (D) (*p* < 0.001).

(E).

(F). Log2 enrichment of amino acids groups in At-RBPs (green), proteins from the input proteome annotated as ‘RNA-binding’ (light purple), At-RBPs with unknown RBD (light orange) and candidate At-RBP (blue) compared to the input proteome. The significance of enrichment/depletion was tested by two-sample test for population proportion. Comparing all RBP subsets to the input proteome, results were significant (*p* < 0.001) for all amino acid groups, except for aromatic amino acid in At-RBPs, negative amino acid in candidate At-RBPs, positive/aromatic amino acid in At-RBPs with unknown RBD.

**Figure 3**. Recognized and unknown RBDs isolated by mRNA interactome capture.

(A). Number of proteins harbouring domains associated with RNA-binding in At-RBPs (green), candidate At-RBPs (blue), or only identified in the input proteome (white).

(B). Number of proteins harbouring domains not associated with RNA-binding in At-RBPs (green), candidate At-RBPs (blue), or only identified in the input proteome (white).

**Figure 4**. mRNA interactome capture provides experimental evidence of RNA-binding for many predicted Arabidopsis RBPs.

1. Members of predicted Arabidopsis RBPs families identified as At-RBPs (green), candidate At-RBPs (blue), or only in input proteome (black). Proteins with prior experimental evidence of RNA-binding are underlined, proteins not detected in the input proteome are in italic. CID: CTC-interacting domain, PUM: Pumilio, NTF2: Nuclear Transport Factor 2, LARP: La-related protein, UPF: Up-frameshift protein, GR-RBP: Glycine-rich RNA-binding protein, PTB: polypyrimidine tract-binding protein, EJC: exon junction complex.
2. Arabidopsis YTH domain proteins identified in input proteome, At-RBPs, candidate At-RBPs and schematic representation of annotated domains. CPSF30 gives rise to two transcripts, a longer one containing a YTH domain and a shorter one without this domain.
3. Meta-transcript analysis showing enrichment of m6A sites within a 100 nucleotide window upstream*,* and depletion of m6A sites downstream of polyadenylation sites, based on polyadenylation cluster (PAC) sites from Wu et al., 2011 (left panel) and cleavage sites from Sherstnev et al., 2012 (right panel). Effect sizes are given as log10-transformed odds ratios (OR) and are shown by the red lines; the red shaded areas show 95% confidence intervals of the log10-transformed ORs. Statistical significance of the enrichment/depletion at the FDR = 0.001 level is denoted by the top blue bars.
4. Arabidopsis Alba domain family members identified in input proteome, At-RBPs, candidate At-RBPs, and schematic representation of annotated domains.

**Figure 5**. Novel Arabidopsis RBPs.

(A-C).Novel families of Arabidopsis RBPs (A) Whirly proteins; (B) DUF1296 proteins; (C), LIM proteins.

(D) Illustration of diverse functions of non-canonical RBPs in Arabidopsis.

**Supplemental Figures**

**Supplemental Figure 1**. Optimization of the Arabidopsis mRNA interatome capture protocol.

(A). RNA extracted from four-day-old etiolated seedlings was cross-linked (CL) either once, twice or three times at UV 150 mJ/cm2. Non-cross-linked (noCL) seedlings were used as control. RNA integrity was analysed on a Bioanalyzer 2100 using an RNA pico chip, and is depicted as the gel representation of in-chip electrophoresis.

(B). Lysates were prepared from noCL four-day-old etiolated seedlings using a lysis buffer with or without polyvinylpyrrolidone 40 (PVP40) and -mercaptoethanol (-ME). RNA captured on oligo(dT) beads was extracted and used for qRT-PCR. Measurements are the average of three technical replicates and are shown as fold change in RNA recovery upon addition of PVP40 and -ME to lysis buffer.

(C). Two rounds of oligo(dT) capture were performed using a lysate from noCL four-day-old etiolated seedlings. Aliquots of the lysate were taken after the first and second round of oligo(dT) capture, respectively, followed by RNA extraction and qRT-PCR analysis. Measurements are the average of three technical replicates with error bars representing the SD. Measurements are shown relative to the input.

(D). Scatter plots comparing the protein enrichment in CL over noCL based on LC-MS/MS intensities of two biological replicates. Proteins significantly enriched (FDR < 5%) in CL or noCL are depicted in red. Proteins that lack enrichment are depicted in black.

**Supplemental Figure 2**. GO enrichment analysis of At-RBPs and candidate At-RBPs.

(A-B) The most significantly over- and under- represented Gene Ontology (GO) terms for biological process in At-RBPs (A) and candidate At-RBPs (B).

(C-D) The most significantly over- and underrepresented GO terms for cellular compartment in candidate At-RBPs (C) and candidate At-RBPs (D).

**Supplemental Figure 3**. Conservation of RBPs across kingdoms.

(A) Categorization of At-RBPs based on conservation between Arabidopsis, yeast, mouse, and human.

(B) Overlap of At-RBPs with proteins identified in mRNA interactomes of yeast (Beckmann et al. 2015), mouse (mESC (Kwon et al., 2013, HL-1 (Liao et al. 2016)) and human (HEK293 (Baltz et al., 2012), HuH7 (Beckmann et al. 2015)).

**Supplemental Figure 4**. Analysis of amino acid enrichment/depletion in the Arabidopsis mRNA interactome.

Log2 enrichment (A) and depletion (B) of amino acids of At-RBPs (green), proteins from the input proteome annotated as ‘RNA-binding’ (light purple), At-RBPs with unknown RBD (light orange) and candidate At-RBP (blue) compared to the input proteome.

**Supplemental Data Sets**

**Supplemental Data Sets 1**. Arabidopsis mRNA interactome.

List of Arabidopsis interactome and candidate proteins including protein and gene IDs, log2 fold enrichement, p-values, protein domains, classifications and associated GO terms (obtained from TAIR (version 10) ATH\_GO\_GOSLIM.txt .gz downloaded from <https://www.arabidopsis.org/> on 20th August 2015). HEK293, mESC, HL-1, HuH-7 and yeast mRNA interactomes (Baltz et al., 2012; Kwon et al., 2013, Liao et al., 2016; Beckmann et al., 2015) were surveyed for the presence of Arabidopsis RBPs.

**Supplemental Tables**

**Supplemental Table 1**. Zinc-finger proteins not associated with RNA-binding identified by mRNA interactome capture.

List of zinc-finger proteins within At-RBPs not previously known to bind RNA including their annotated protein domains and FDRs indicating enrichment in CL compared to noCL in mRNA interactome capture.

**Supplemental Table 2**. Examples of non-canonical RBPs identified by mRNA interactome capture.

List of non-canonical At-RBPs and their annotated protein domains, known functions and FDRs indicating enrichment in CL compared to noCL in mRNA interactome capture.