RESEARCH ARTICLE   **Open Access**

Changes in the Arabidopsis RNA-binding proteome reveal novel stress response mechanisms

Claudius Marondedze[1,2,4*] (ID), Ludivine Thomas[2,5], Chris Gehring[2,3] and Kathryn S. Lilley[1]

Abstract

Background: RNA-binding proteins (RBPs) are increasingly recognized as regulatory component of post-transcriptional gene expression. RBPs interact with mRNAs via RNA-binding domains and these interactions affect RNA availability for translation, RNA stability and turn-over thus affecting both RNA and protein expression essential for developmental and stimulus specific responses. Here we investigate the effect of severe drought stress on the RNA-binding proteome to gain insights into the mechanisms that govern drought stress responses at the systems level.

Results: Label-free mass spectrometry enabled the identification 567 proteins of which 150 significantly responded to the drought-induced treatment. A gene ontology analysis revealed enrichment in the "RNA binding" and "RNA processing" categories as well as biological processes such as "response to abscisic acid" and "response to water deprivation". Importantly, a large number of the stress responsive proteins have not previously been identified as RBPs and include proteins in carbohydrate metabolism and in the glycolytic and citric acid pathways in particular. This suggests that RBPs have hitherto unknown roles in processes that govern metabolic changes during stress responses. Furthermore, a comparative analysis of RBP domain architectures shows both, plant specific and common domain architectures between plants and animals. The latter could be an indication that RBPs are part of an ancient stress response.

Conclusion: This study establishes mRNA interactome capture technique as an approach to study stress signal responses implicated in environmental changes. Our findings denote RBP changes in the proteome as critical components in plant adaptation to changing environments and in particular drought stress protein-dependent changes in RNA metabolism.

Keywords: *Arabidopsis thaliana*, Drought stress, Proteomics, Mass spectrometry, mRNA interactomics, RNA-binding proteins, Systems analysis

Background

RNA-binding proteins (RBPs) determine RNA fate from synthesis to decay and are increasingly recognized as critical post-transcriptional gene regulators. RBPs bind mRNAs through RNA-binding domains (RBDs) and consequently affect RNA availability for processing and translation [1] essential for stimulus-specific responses [2]. Remarkably, in the animal systems it was noted that modifying the expression pattern or mutating RBPs and/or their target binding sites influence alternative splicing events and can trigger diseases such as neurological disorders and cancers [3–5]. Additionally, transcriptional arrest inducing stress granule formation in response to stresses such as low oxygen, oxidative and heat stresses has been observed [6–10]. Stress granules are cytoplasmic foci formed from cytoplasmic aggregates of non-translated messenger ribonucleoproteins and are described as sites of mRNA storage, sorting and triage. However, they are yet to be fully characterized in plants and other systems.

Systems level detection of the RNA-binding proteome (RBPome) has been made possible by the use of the interactome capture technology and has yielded genome-wide mRNA interactomes in several species [11–17] and they

* Correspondence: cmarondedze@gmail.com
[1]Cambridge Centre for Proteomics, Cambridge Systems Biology Centre, and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK
[2]Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, 23955-6900 Thuwal, Kingdom of Saudi Arabia
Full list of author information is available at the end of the article

Marondedze *et al. BMC Plant Biology*   (2019) 19:139

Page 2 of 11

have revealed high degree of similarity between mammalian cells and yeasts [13], as well as plants [12, 18, 19] suggesting an ancient origin. Interactome capture involves in vivo fixing of proteins to their target mRNAs by UV crosslinking followed by purification of mRNA-protein complexes through affinity capture of polyadenylated RNA and then analyzing interacting proteins by tandem mass spectrometry (MS/MS). This technique has a great advantage over other crosslinking techniques based on chemical fixation in that it generates covalent linkages between physically interacting proteins and mRNAs in vivo [20, 21]. It also permits time resolved isolation of RBPs allowing characterization of targeted developmental and physiological states of cellular systems. Furthermore, it has been established that cellular reactions to stress signals compel tight regulation of gene expression including the timely up-regulation of genes encoding for specific stress-responsive factors. However, the respective stress responsive RBPome remain yet to be established. Therefore, we set out to determine whether a defined abiotic stress induces an RBP response signature and did so using a response to drought stress in Arabidopsis as an experimental test system. We firstly established the interactome capture technique as an approach to study stress response, in particular drought responsive RBPome. We argue that changes in the latter will afford insights into the mechanisms that govern metabolic changes during stress and our results would afford a unique systems view RBP changes. Finally, we suggest that stress-induced RBPs may be an evolutionarily conserved mechanism governing post-transcriptional responses to stress.

Results and discussion
Here we used *Arabidopsis thaliana* cell suspension cultures (ecotype Columbia-0) to obtain and characterize the plant stress RBPome and gain insight into mechanisms that govern stress responses at the systems level. Three biological replicates were treated with 40% (*v/v*) PEG, a dehydration-inducing agent to mimic drought stress, collected samples at 1 h and 4 h and measured ABA levels.

Abscisic acid assay
To confirm whether the treatment of cell suspension cultures with 40% (*v/v*) PEG was sufficient to induce or mimic drought stress, we performed an ABA assay using the Phytodetek® ABA Immunoassay kit. A rapid increase in ABA levels at 1 h and 4 h after treatment as compared to the control samples was observed (Fig. 1a). We noted a three-fold increase in ABA at 1 h and a 1.5-fold at 4 h, which is consistent with the induction of drought stress (Fig. 1a). This further validated that exposure to drought stress signals a rapid cellular signal that leads to an increase in the hormonal levels of ABA, a canonical stress marker for drought or water dehydration stress.

Identification of RBPs responsive to drought stress
Mass spectrometry identified 1408 proteins of which 567 proteins showed specific time-dependent responses to drought stress and these represent the drRBPome. Within the 567 responsive proteins, 178 proteins were detected either at 1 h or 4 h after treatment, 191 proteins were detected at both 1 h and 4 h after treatment while 48 proteins were consistently detected only in the control samples or rather could not be detected upon exposure to stress (Fig. 1b, Additional file 1: Table S1). It is important to note that these time-dependent transient changes that occurred as result of drought stress were considered only when the proteins were detected consistently in all the three biological replicates used in this study. This group of RBPs represent proteins that are present or absent only in the stress-treated samples. The remaining 150 significantly (*p*-value ≤0.05) altered proteins (Fig. 1b, Additional file 1: Table S1) represent the drought stress responsive RBPs (drRPBs). The drRBPs contain the ribosomal protein S15A (AT1G07770), ILITYHIA (AT1G64790) and ABA hypersensitive 1 (CBP80, AT2G13540) that significantly increase in abundance (log$_2$-fold change ≥1.5, *p*-value ≤0.05) (Additional file 1: Table S1) at both 1 h and 4 h after treatment. In contrast, the abundance of RNA-binding protein (AT3G15010), flowering time control protein (AT4G16280), hyaluronan (AtRGGA, AT4G16830), co-chaperone GrpE family protein (AT4G26780), glycine-rich RNA-binding protein 8 or cold, circadian rhythm and RNA-binding protein 1 (GR-RBP8, AT4G39260) and nuclear transport factor 2 (NTF2, AT5G43960) decrease. Most of these proteins have previously been linked to drought and/or ABA responses, for example CBP80 and AtRGGA whose role as RBPs have been proposed to be important for a proper response to osmotic stress [22–25]. AtRGGA gene expression was observed to increase in seedlings following a prolonged exposure to either ABA or PEG [22]. This is consistent with the view that these RBPs operate as post-transcriptional modulators of the drought stress response signaling.

Next we assessed the changes at the proteome level of stimulus-specific RBPs bound to RNA. At the protein level, most drRBPs did not change except for five proteins that show significant (*p*-value ≤0.05) changes in their abundance both at the protein and RBP levels post treatment (Additional file 2: Figure S1). Three of these five proteins, the calcium-binding EF hand protein, rotamase CYP1 and the RNA-binding protein (AT1G60650), show contrasting responses at protein abundance and RNA-binding levels i.e. increasing RNA-binding level and decreasing at protein abundance. Responses of two proteins, SWIB/MDM2 domain protein and stress inducible protein, show similar trends both at protein and
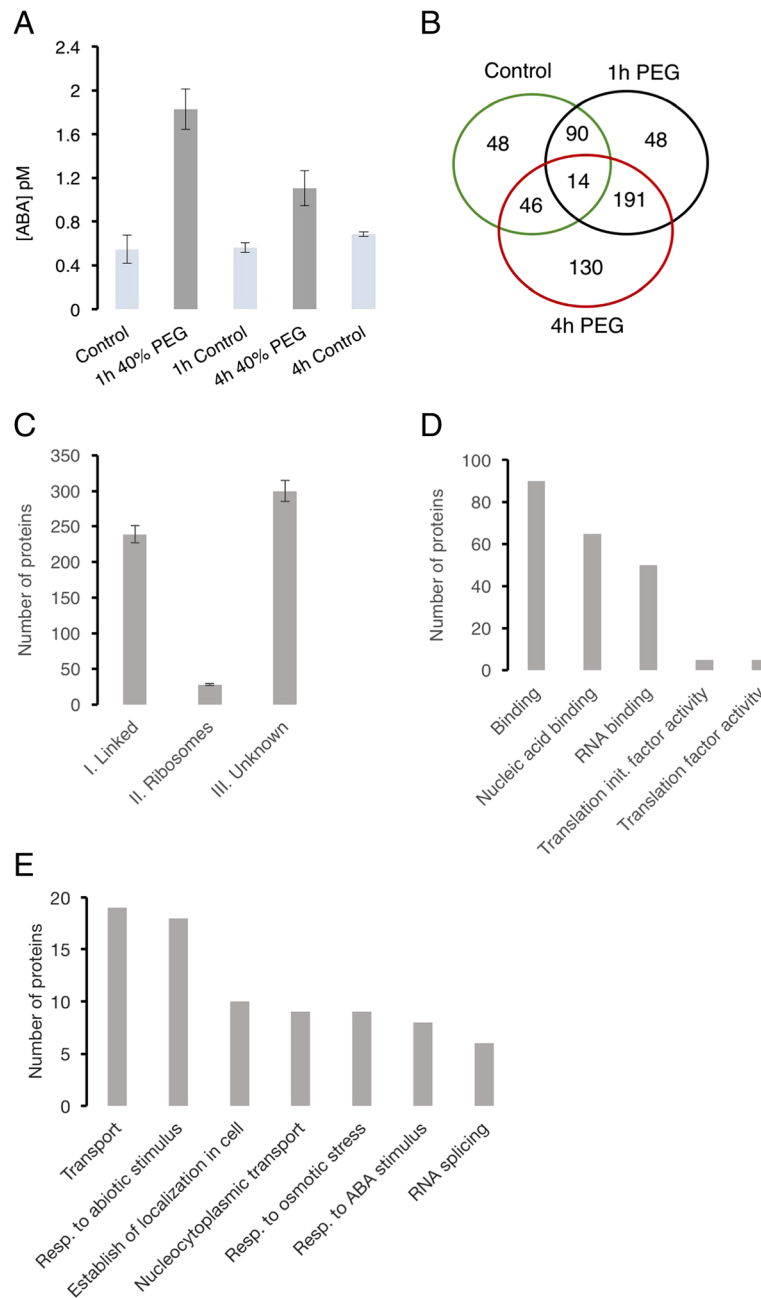
Fig. 1 Abscisic acid assay, drought stress responsive RNA-binding proteome (RBPome) and systems classification of drought stress response RNA-binding proteins in *Arabidopsis thaliana*. **a** Abscisic acid assay performed using the Phytodetek® ABA Immunoassay kit showing significant (p-value ≤0.05) increases in ABA after drought stress treatment on Arabidopsis cell suspension cultures. **b** Distribution of the RBPome before and after treatments at 1 h and 4 h. **c** Classification of drought stress responsive proteins based on three categories: category 1 proteins linked to RNA biology, category II ribosomal proteins and category III proteins with unknown RNA biology or RNA-binding domains. The bars are standard deviation error bars signifying whether a protein is present in at least two biological replicates. **d-e** Gene ontology (GO) molecular function (**d**) and biological processes (**e**) enrichment analyses of the significantly enriched (150 proteins) RBPs linked to drought stress responses

RNA-binding levels. Overall, the data show that changes observed in RNA-binding to the protein are due to RBP-RNA interactions and therefore part of the cellular response to the stress stimulus.

Domain organization of drought stress responsive RBPome

To further characterize the composition of the drRBPome, we looked at the RBD enrichment and noted that 50 proteins contain the RNA recognition

motifs (RRM), representing the most prominent RBD in this dataset (Additional file 3: Figure S2). Six of the RRM containing proteins also contain the NTF2 domain. Other predominant "classical" RBDs include the zinc finger (ZF)-CCCH, K-homology, DEAD helicases, like-Smith (commonly know as LSm), poly(A) binding protein and cold shock domain (Additional file 3: Figure S2). Additionally, 300 proteins contain unknown or unconfirmed RBDs (Additional file 4: Table S2). The RDBs gave rise to a three-way classification. Category I comprises proteins linked to RNA biology based on their RBDs and/or role in RNA processing (42% of the RBPome), category II contains ribosomal proteins (5%) and category III contains proteins with currently unknown RNA-interactions (53%) (Fig. 1c). The later is indicative of a large set of potential RNA-interacting proteins that are yet to be fully characterized and in particular, their mode of action and target RNAs.

The drRBPs shows enrichments in gene ontology (GO) categories

Perhaps not surprisingly, drRPBs are enriched in the functional categories, "nucleic acid binding" and "RNA binding" (Fig. 1d) and drought-specific processes, which are among the most enriched biological processes. The latter include "response to stress" and "response to stimulus" including hormonal and temperature stimulus, "response to ABA stimulus", "response to osmotic stress" and "response to water deprivation" (Fig. 1e, Additional file 5: Table S3). In addition, signal transduction relay associated processes are also enriched including "transport" and "establishment of localization in cell". In the latter, proteins enriched in this category include six NTF2 proteins that are involved in nucleocytoplasmic transport of mRNA, a process that enables translation of the respective mRNAs at their destination site [26]. All the six NTF2 proteins decrease in abundance within the first hour of treatment potentially signaling a reduction in nucleocytoplasmic transport of their target mRNAs. In contrast, nuclear pore anchor, a protein that mediates the transport of RNA and other cargo between the nucleus and the cytoplasm, increase in abundance. The nuclear pore anchor has been shown to be necessary for RNA homeostasis between the nucleus and cytoplasm and is required for e.g. flowering time and auxin signaling [27].

Co-expression analysis was performed on a selected set of the most up-regulated and/or down-regulated proteins and the top 300 co-expressed proteins for further characterization. Overall, co-expression analysis shows that among the most ranked proteins were a general bias towards proteins that are time specific upon stress treatment, although some proteins that are differentially regulated at 1 h and 4 h also existed (Additional file 6: Table S6).

Examples of the latter include ILITHYA, which had 56 co-expressed proteins from the drRBP responsive proteins, of which 17 are up-regulated upon drought stress. Up-regulated proteins included classical drought stress responsive proteins such as CBP80, proteins involved in intermediary metabolism such as phosphofructokinase (AT1G20950) and carbohydrate binding like fold (AT3G62360) and various RNA binding proteins including NTF2 (AT1G13730), eIF2 gamma (AT1G04170) and ribosomal protein L4/L1 (AT3G09630). The second highly represented co-expressed protein is the nucleolar GTP-binding protein (AT1G50920) with 55 proteins that are drought stress responsive and 13 of these are differentially regulated. The third is adenine nucleotide alpha hydrolases-like super protein (AT5G54430) with 41 (of which 15 are differentially regulated) proteins, followed by guanylate-binding protein (AT5G46070) that showed 38 (of which 13 are differentially regulated) proteins. Among the least represented is the flowering time control (AT4G16280) with 15 proteins (five are differentially regulated). Besides their classical biological process of developmental role, proteins co-expressed with flowering time control protein are also involved in RNA binding. Proteins co-expressed with adenine nucleotide alpha hydrolases-like super protein show a bias towards enrichment of biological processes such as "response to stress" and "primary metabolic process". Of interest to note is that all the enriched biological process of the selected proteins and their respective co-expressed ones are involved in RNA metabolic processes, translational activities or intermediary metabolism and functionally are biased towards RNA-binding. Co-expression analysis suggests that the drRBPs have a strong connectivity network and biotechnologically may be important targets towards improving tolerance to drought stress in crop plants.

EnigmaRBPs detected responsive to drought stress

A pathway analysis of the unique UV-enriched and drRBPs was undertaken. The KEGG annotated pathways reveal a bias towards metabolic enzymes especially for proteins increasing in abundance after treatment. Seven stress-responsive proteins belong to the carbohydrate metabolism pathway (Additional file 7: Table S4) and six (glyceraldehyde 3-phosphate dehydrogenase C-2 (GAPC2), pyruvate dehydrogenase E1 component α-subunit (PDHA), phosphofructokinase, aldehyde dehydrogenase 7B4 (ALDH7B4), cytosolic NAD-dependent malate dehydrogenase 1 and aconitase 3 (ACO)) have a role in glycolysis and the citric acid cycles. Two proteins (ALDH7B4 and monodehydroascorbate reductase 1) are part of the ascorbate and aldarate metabolism (Fig. 2).

Identification of RBPs with a role in metabolism link post-transcriptional gene regulation to stress-induced
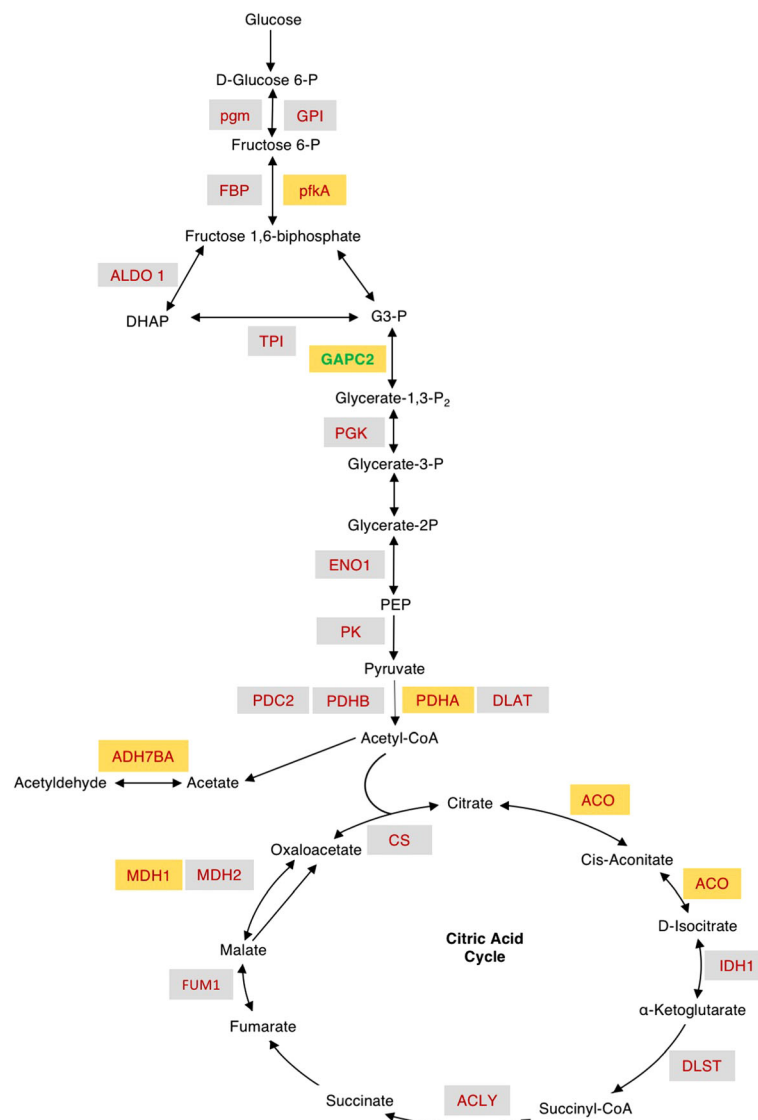
Fig. 2 Schematic glycolytic and citric acid pathways illustrating the proteins identified responsive to drought stress at RBP level (orange boxes) with the bold green GAPC2 only identified in the current study and others including the enzymes in grey boxes have been previously identified [12]. Enzymes in orange boxes namely pfkA, GAPC2, ALDH7B4, ACO and MDH1 significantly increase and PDHA decrease in abundance after drought stress treatment (Additional file 1: Table S1). The figure is adapted from [12]. Abbreviations in order of appearance: pgm: phosphoglucomutase 2; GPI: glucose-6-phosphate isomerase; FBP: fructose 1.6-bisphosphate phosphatase or high cyclic electron flow 1; pfkA: phosphofructokinase 3; ALDO1: fructose bisphosphate aldolase 1; P: phosphate; TPI: triosephosphate isomerase; GAPC2: glyceraldehyde-3-phosphate dehydrogenase C2; P_2: bisP; PGK: phosphoglycerate kinase; G3-P: glyceraldehyde 3-phosphate; ENO1: enolase; PK: pyruvate kinase; PDC2: pyruvate decarboxylase 2; PDHB: transketolase or pyruvate dehydrogenase E1 component beta subunit; PDHA: pyruvate dehydrogenase E1 component subunit alpha-3, chloroplastic; DLAT: mitochondrial pyruvate dehydrogenase subunit 2–2 or dihydrolipoyllysine-residue acetyltransferase component 2 of pyruvate dehydrogenase complex; ALDH7B4: aldehyde dehydrogenase 7B4; CS: citrate synthase; ACO: aconitase; IDH1: NADP-specific isocitrate dehydrogenase; DLST: dihydrolipoamide succinyltransferase; ACLY: ATP-citrate lyase; FUM1: fumarase 1; MDH1: cytosolic NAD-dependent malate dehydrogenase; MDH2: mitochondrial malate dehydrogenase 2

metabolic changes and may suggest that RBPs exert their effect by (auto-)regulating their own or other mRNA species. Moreover, four of the carbohydrate metabolism proteins (GAPC2, ALDH7B4, PDHA and ACO) are also enriched in gene ontology categories "response to ABA stimulus", "response to water derivation" or "response to

oxidative stress". In animals, besides their glycolytic activity, GAPC2 has non-glycolytic functions that depend on its subcellular localization, e.g. in the nucleus it acts as a signal for programmed cell death [28] and is involved in posttranscriptional regulation and maintenance of DNA integrity [29]. At protein level, GAPC2

expression has been observed to increase in response to cold stress [30]. In the present study, we note an increase of GAPC2 in response to drought stress at post-transcriptional level denoting a potential transcriptional rise of its target RNA. Aldehyde dehydrogenase 7B4, a member of the "turgor-responsive" ALDH genes [31] also increases in abundance after stress treatment. The ALDH protein family detoxifies aldehydes generated in plants when exposed to environmental stresses such as salinity and dehydration [32, 33]. Knockout mutants, *ALDH3I1* and *ALDH7B4* T-DNA, displayed higher sensitivity to dehydration and salinity stress compared to the wild-type plants consistent with a role of ALDH genes in stress responses [34]. At transcriptional level, abundance of *ALDH7B4* increases in plantlets and roots after dehydration and ABA treatments and declines in a time-dependent manner after stress relief [35]. In a previous study, we identified ALDH7B4 as a candidate RBP [12], and consistently, we find it enriched in the stress-responsive RBPome. It appears that ADLH7B4 has a dual function as a glycolytic enzyme and interacting with RNA thereby acting as a post-transcriptional gene regulator during drought stress. Another well-characterized glycolytic enzyme that we also noted to be drought stress responsive is ACO. Aconitase is an iron regulatory protein 1 (IRP1) that catalyzes the conversion citrate to isocitrate (Fig. 2). In animals, ACO1 is a bifunctional protein that becomes catalytically active in the presence of an iron-sulfur cluster in its catalytic center, while in the absence of the cluster, it operates as RBP, modulating the translation or stability of transcripts [36]. In plants, nitric oxide and oxidative stress have been shown to modulate the expression of ferritins [37] and to inactivate ACO catalytic activity [38] converting it to IRP1 through structural changes to its 4Fe-4S cluster. ACO3 is responsive to oxidative stress [39, 40] and interacts with mRNA in vivo [12]. The increase in abundance of ACO3 during drought stress is consistent with a post-transcriptional regulatory role that is likely to affect the transcriptome and eventually the proteome and metabolome during responses to stress. Taken together, it appears that drought stress-induced differential accumulation of RNA-interacting proteins is over-represented in specific functional groups.
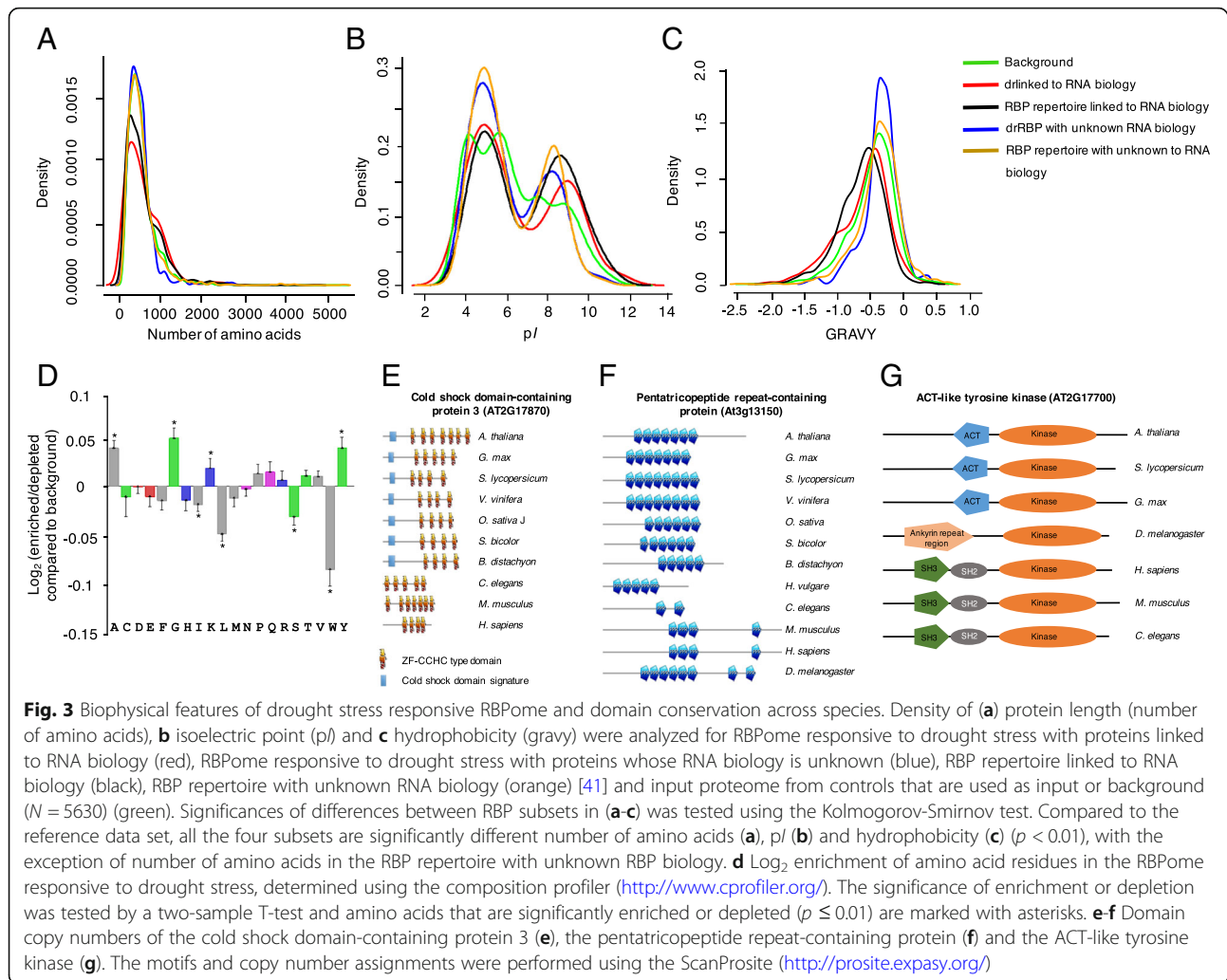
Biophysical characteristics and sequence topology of drRBPome

Biophysical and amino acid (aa) sequence characteristics were also analyzed to determine the physical properties that enable RBPs to interact with RNA. The drRBPome, much like the input reference and the RBP repertoire data [41], span the full spectrum of protein sizes, with the majority of proteins being < 1000 aa long (Fig. 3a). However, compared to the reference data, we notice that

drRBPs linked to RNA biology behave the same as the RBP repertoire linked to RNA biology compared to the drRBP with unknown RNA biology and RBP repertoire with unknown RNA biology. Proteins linked to RNA biology show a high density for proteins with amino acid sequence length of between 1000 and 2000 compared to proteins whose RNA biology is unknown. A similar trend is noted on the isoelectric point (pI) distribution (Fig. 3b). The pIs of proteins enriched in RNA interaction show similar patterns distinct from the reference data. In addition, proteins with unknown RNA biology from both drRBP and RBP repertoire sets have the same configuration and the same for drRBP and RBP repertoire proteins linked to RNA biology. The pI distribution of the latter significantly shifts towards higher pI (≥ 8) as compared to the reference proteome. A slight hydrophobicity bias is noted on the proteins with unknown RNA biology compared to the proteins linked to RNA biology, however, the enhanced density peak for drRBP with unknown RNA biology could be attributed to the much smaller number of proteins in this data set (Fig. 3c). Overall, the consistent trend observed on the pI, number of amino acids and hydrophobicity distribution on proteins with unknown RNA biology compared to the proteins whose RNA biology are known, may suggest additional properties with implications in RNA interactions of the novel proteins in this set. If we consider overall aa frequencies in the drRBPome and input reference as the basis for the analysis, we note that aa residues with polar side chains are favored since they have high affinity for RNA such as lysine, which is significantly ($p < 0.05$) enriched. Additionally, glycine that interacts strongly with guanine [42], is also significantly enriched while aa with aliphatic side chains such as phenylalanine (F) and tryptophan (W) are generally underrepresented (Fig. 3d).

Conservation of drRBPs across different species

Many of the drRBPs identified (85%, 127 proteins) have orthologs in other plants (notably in *Brachypodium distachyon*) (Additional file 8: Table S5) and 70% (101 proteins) have orthologs in human, mouse, drosophila, *Caenorhabditis elegans* and yeast hinting at ancient origin RBP-dependent responses. A comparative analysis of domain architectures reveals similarities and loss or gain of domain copies across different species. The Arabidopsis cold shock domain-containing protein 3 (AtCSD3), for example, has orthologs in nearly all organisms examined (Additional file 8: Table S5). AtCSD3 is the longest ortholog and similarly to mouse has seven ZF-CCHC-type domains. CSD3 contains glycine-rich regions and at least four ZF-CCHC-type domains (Fig. 3e). Importantly, in addition to the ZF-CCHC, the CSD domain is present and seems unique to plants and may have evolved to

Fig. 3 Biophysical features of drought stress responsive RBPome and domain conservation across species. Density of (**a**) protein length (number of amino acids), **b** isoelectric point (p*I*) and **c** hydrophobicity (gravy) were analyzed for RBPome responsive to drought stress with proteins linked to RNA biology (red), RBPome responsive to drought stress with proteins whose RNA biology is unknown (blue), RBP repertoire linked to RNA biology (black), RBP repertoire with unknown RNA biology (orange) [41] and input proteome from controls that are used as input or background (N = 5630) (green). Significances of differences between RBP subsets in (**a**-**c**) was tested using the Kolmogorov-Smirnov test. Compared to the reference data set, all the four subsets are significantly different number of amino acids (**a**), p*I* (**b**) and hydrophobicity (**c**) ($p < 0.01$), with the exception of number of amino acids in the RBP repertoire with unknown RBP biology. **d** Log$_2$ enrichment of amino acid residues in the RBPome responsive to drought stress, determined using the composition profiler (http://www.cprofiler.org/). The significance of enrichment or depletion was tested by a two-sample T-test and amino acids that are significantly enriched or depleted ($p \leq 0.01$) are marked with asterisks. **e**-**f** Domain copy numbers of the cold shock domain-containing protein 3 (**e**), the pentatricopeptide repeat-containing protein (**f**) and the ACT-like tyrosine kinase (**g**). The motifs and copy number assignments were performed using the ScanProsite (http://prosite.expasy.org/)

optimize survival under drought conditions that incidentally are also induced by freezing [43].

A pentatricopeptide repeat (PPR)-containing protein (At3g13150) showed higher PPR-repeat copies in plants and drosophila than in animals (Fig. 3f). PPR proteins are an emerging class of RBPs with a 35-aa motif, repeated in tandem up to 30 times and have been proposed to function as molecular adaptors for RNA processing [44]. RNA-binding selectivity is conferred by dimers where an AsnAsp (ND) interacts with uracil, AsnSer (NS) with cytosine, SN with adenine and ThrAsp (TD) with guanine [45].

The number of PPR-containing proteins in land plants is higher (> 450) as compared to algae, as well as protozoa, yeast or animals (< 50) [44]. Furthermore, PPRs have been reported to be involved in RNA metabolism in plant mitochondria and chloroplasts and are likely to have a regulatory role in the responses to abiotic stress [46]. It has also been demonstrated that mutations of PPR proteins can result in severe phenotypes due to

disrupted expression of target genes, many of which are essential for plant survival (e.g. the Arabidopsis PPR mutant *high chlorophyll fluorescence (hcf)152* struggle to survive the seedling stage under autotrophic conditions due to defective carbon fixation [47]). These findings are therefore consistent with important functions of the RNA-binding PPR proteins in the adaptation to terrestrial environments.

Besides, CSD and PPR, which are protein already known to interact with RNA, we examined domain conservation among the most regulated proteins with no known RNA binding role. We noted a high degree of conservation in domains across species among the most highly up-regulated proteins including pyridoxal-dependent decarboxylase protein (AT5G11880), guanylate-binding protein (AT5G46070), rotamase CYP 1 (AT4G38740), serine-rich protein (AT5G25280). Similar observation has been made from the most down-regulated proteins including leucine-rich repeat protein (AT5G22320), calcium-binding EF hand protein (AT2G41100) and

structural maintenance of chromosomes protein (AT3G54670), with the exceptions of ACT-like tyrosine kinase, also called serine/threonine/tyrosine kinase 8 (AT2G17700, Fig. 3g). The latter protein is implicated in chloroplast organization in addition to its protein phosphorylation role [48]. It contains a highly conserved kinase domain, which is common in all species. However, aspartate kinase, chorismate mutase and tyrosine A (ACT) domain is detected only in plant species, ankyrin domain only in drosophila and SH2 and 3 domains present in animal systems. The ACT domain is proposed to be a conserved regulatory binding fold that is linked to a wide range of metabolic enzymes that are regulated by amino acid concentration.

Conclusions

In summary, this study characterizes systems level changes occurring in the RBPome during drought stress responses. It highlights that qualitative and quantitative changes in RBPome are likely to affect metabolic processes and carbohydrate metabolism in particular. Control and stability of metabolic processes during exposure to stress are known to increase survival thus implicating the significant changes in the RBPome in post-transcriptional mechanisms that enable regulatory plasticity essential for a timely stress response that in turn enhances short- and long-term adaptations. In addition, it turns out that RBPs have an important biological function during drought stress as changes in RBPs are indicative of a stress response signaling. Finally, our findings are also consistent with evolutionarily conserved roles of RBPs in post-transcriptional drought stress response mechanisms.

Methods

Cell culture and treatment

Cells derived from roots of *Arabidopsis thaliana* (ecotype Columbia-0) were grown in liquid medium, as previously described [39, 49, 50]. The cell cultures used in this study were obtained from Mrs. Xiaolan Yu in the Department of Biochemistry at the University of Cambridge. Cells were treated with 40% (*v*/v) polyethylene glycol (PEG) 6000, a dehydration-inducing agent to mimic drought stress or with equal volumes of media as a negative control. Three biological replicates of cells treated with PEG or mock-treated cells were collected at 1 h and 4 h post-treatment. Each time-point treatment has a corresponding mock treatment per replicate. The medium was drained using Stericup® filter unit (Millipore, Billerica, MA), and cells were rinsed with $1 \times$ phosphate buffered saline immediately before UV-crosslinking [12].

Abscisic acid (ABA) assay

Three biological replicates of cell suspension cultures for each time-point (controls at 0 h, 1 h and 4 h, and 40%

PEG treated samples at 1 h and 4 h) were subjected to Phytodetek® ABA Immunoassay (Agdia Inc., Elkhart, Indiana, US) following the manufacturer's instructions. ABA levels were measured and statistically evaluated between each control and treatment time-point.

UV-crosslinking and interactome capture

In vivo UV-crosslinking and isolation of Arabidopsis RBPs was performed, as previously described [12], using a protocol that utilizes a modified method originally optimized for HeLa cells [11]. Sample from each time-point were split into two, one set for UV-crosslinking and the second set for non UV-crosslinking. Samples for UV-crosslinking were irradiated in vivo with UV (254 nm) and the mRNA-protein complexes were pulled down using oligo(dT) beads. Purified proteins were analyzed by label free tandem mass spectrometry. Similarly to [12], the quality of the mRNA-protein crosslinked complex pull-down was assessed by performing an additional control whereby the sample was treated with RNase T1/A mix (Thermo-Fisher Scientific) according to the manufacturer's recommendations. To isolate RBPs, mRNA-protein samples were treated with RNase A/T1 mix to release them from the captured RNA molecules. Crosslinking and isolation of RBPs were evaluated by western blotting using antibodies against polypyrimidine tract-binding protein 1, β-actin (Sigma Aldrich, St Louis, MO, USA) and histone 3 (Abcam, Cambridge, UK) following manufacturer's recommendations (see [12]).

Protein digestion and mass spectrometry

Protein samples were reduced, alkylated, buffer exchanged and digested, as described elsewhere [12]. Dried peptides were resuspended in 20 μL of 5% (*v*/v) acetonitrile and 0.1% (v/v) formic acid and analyzed with Q-Exactive™ Hybrid Quadrupole-Orbitrap™ using nano-electrospray ionization (Thermo-Fisher Scientific, San Jose, CA) coupled with a nano-Liquid Chromatography (LC) Dionex Ultimate 3000 Ultra High Performance Liquid Chromatography (UHPLC) (Thermo-Fisher Scientific). Mass spectrometry parameters and run analysis were performed following the protocol described in [51].

Mass spectrometry data analysis

Raw files were processed using the Proteome Discoverer v2.1 (Thermo-Fisher Scientific) interlinked with the local MASCOT server (Matrix Science, London, UK). MASCOT searches were carried out against *Arabidopsis thaliana* database (built using the Arabidopsis information resource (TAIR; release 10)) using a precursor mass tolerance of 20 ppm, a fragment ion mass tolerance of ±0.5 Da and trypsin specificity allowing up to two missed cleavages, peptide charges of + 2, + 3 and + 4. Carbamidomethyl modification on cysteine residues was used as a fixed modification,

oxidation on methionine residues as variable modifications and the decoy database was selected. Further stringency was applied on the peptide spectrum matches (PSMs) by allowing "forward" and "decoy" searches by MASCOT to be re-scored using the Percolator algorithm in Proteome Discoverer v2.1 thus yielding a robust false discovery rate (FDR) of < 1%.

UV-crosslink enrichment

Protein enrichment upon UV-crosslinking was performed as previously described [12] using Microsoft Excel. Proteins that were detected in both the UV-crosslinked samples and the control (non-UV crosslinked samples) were quantitatively analyzed to assess UV-crosslinking enrichment. Normalized intensities of UV-crosslinked samples were compared quantitatively against normalized intensities of the control (non-UV crosslinked samples), and a $\log_2$-fold change of ≥2 and *p*-value of ≤0.05 calculated using Student's T-test corrected for multiple testing using a method described previously [52] were applied for proteins to be categorized as enriched RBPs and to be considered for further analysis.

Drought stress responsive RB-proteome (drRBPome) analysis

After normalization of the data and UV-crosslink enrichment analysis, proteins from the UV-crosslink enrichment and those that were only identified in the UV-crosslinked samples were used for quantitative analyses. Proteins only detected in at least two biological replicates were included. In this analysis, samples collected at 1 h time-point, that is 1 h PEG treated samples and mock treated controls were compared against each other and similarly for samples collected at 4 h time-point. Proteins with a $\log_2$-fold change ≥1.5 and *p*-value ≤0.05 corrected for multiple testing a method detailed elsewhere (Benjamini and Hochberg [52]) represented the significantly responsive proteins and were categorized as the significantly regulated drought stress responsive RBPs (drRBPs).

Bioinformatics analyses
Classification of RBPs and gene ontology analyses

Classical and non-classical RNA-binding domains (RBDs) were detected from the drRBPome identified both in this study using pfam (http://pfam.xfam.org; February 2017). RBPs and candidate RBPs were classified, as described previously [13]. Furthermore, three categories were extrapolated to give clarity to the data, as reported previously [19]. Category I contains all proteins that have been reported or shown to have a role in RNA associated processes (linked to RNA biology), category II comprises of all detected ribosomal proteins, and category III contains the remaining proteins that have either no known RBDs

or known association with RNA. Gene ontology (GO) enrichments were performed using AGRIGO (http://bioinfo.cau.edu.cn/agriGO/) and pathway analysis was done with the KEGG mapper (http://www.kegg.jp/kegg/tool/annotate_sequence.html; February 2017), which annotates sequences by BlastKOALA. BlastKOALA is an internal annotation tool in KEGG that assigns KEGG Orthology numbers by BLAST searches against a non-redundant set of KEGG GENES using SSEARCH computation [53]. Co-expression for functional and data correlation analysis of selected up- and down- regulated proteins was performed using ATTED database (http://atted.jp).

Biophysical characteristics and sequence topographies analyses

Analyses of biophysical properties including length of proteins (number of amino acids), isoelectric points (p*I*) and hydrophobicity were performed using R (version 3.3.1). Amino acid composition enrichment between the drought stress responsive RBPome and input total proteome as reference as the background set was determined using the web-based composition profiler program (http://www.cprofiler.org/) using default setting and ordering amino acids by hydrophobicity (Kyte-Doolittle) and the significance level was further assessed using Bonferroni correction [54]. Length and sequences of amino acid were retrieved from TAIR (https://www.arabidopsis.org/tools/bulk/sequences/index.jsp), p*I* were obtained from TAIR (https://www.arabidopsis.org/tools/bulk/protein/index.jsp) and hydrophobicity values were calculated using the GRAVY calculator (http://www.gravy-calculator.de).

Evolutionary conservation of drRBPs

To understand the conservation and potentially, the role of drRBPs, InParanoid version 8 (http://inparanoid.sbc.su.se/cgi-bin/index.cgi, [55]) was used to identify their predicted orthologs among Arabidopsis, selected dicots (*Glycine max*, *Solanum lycopersicum*, *Vitis vinifera*), monocots (*Brachypodium distachyon*, *Hordeum vulgare*, *Oryza sativa*, *Sorghum bicolor*), *Saccharomyces cerevisae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus* and *Homo sapiens*. Here, a two-way prediction was possible and data was compiled in Excel. The InParanoid program generates ortholog groups including all inparalogs with scoring below 0.05, which is achieved by using clustering rules based on genome-wide pairwise sequence similarity matches between two species [55].

Additional files

Additional file 1: mRNA-interacting proteins responsive to drought stress treatment at 1 h and 4 h treatment. (XLSX 103 kb)

Additional file 2: Proteins identified as responsive to polyethylene glycol treatment. (A) RNA binding protein (AT1G60650), (B) Rotamase

CYP1 (AT4G38740), (C) SWIB/MDM2 domain protein (AT2G35605), (D) Calcium binding EF hand (At calmodulin like 4, AT2G41100), (E) Stress-inducible protein (AT1G62740). Total soluble protein changes are represented by the grey bars and RNA-binding protein or mRNA-interacting protein changes by the black bars. The asterisk represents significantly ($p < 0.05$) changing protein at a given time. (PDF 23 kb)

Additional file 3: Classical and non-classical RNA-binding domains in *Arabidopsis thaliana* drought stress responsive RBPs mined using pfam database. (A) Most represented classical RNA-binding domains. (B) Most represented non-classical RNA-binding domains. Bars in blue represent number of protein domains mined from differentially expressed drought responsive proteins compared to domains present in drought responsive time specific proteins in red. (PDF 33 kb)

Additional file 4: Protein domains of the drought stress responsive proteins extracted from the pfam database (http://pfam.xfam.org; February 2017). (XLSX 154 kb)

Additional file 5: Gene Ontology analysis of the significantly enriched drought stress responsive proteins performed using AgriGO software (http://bioinfo.cau.edu.cn/agriGO/). (XLSX 49 kb)

Additional file 6: Co-expression analysis of the most up - and/or down- regulated proteins mined using the ATTED database (http://atted.jp/top_search.shtml#GeneTable). (XLS 190 kb)

Additional file 7: KEGG BlastKOALA (https://www.kegg.jp/blastkoala/) pathways represented by the differentially abundant proteins responsive to drought stress (XLSX 44 kb)

Additional file 8: Inparalog and Orthologs clusters for drought stress responsive protein of *Arabidopsis thaliana* and selected organism mined using the Inparanoid database (http://inparanoid.sbc.su.se/). (XLSX 308 kb)

Abbreviations
aa: Amino acid; ABA: Abscisic acid; ACO: Aconitase; ALDH: Aldehyde dehydrogenase; AtCSD3: *Arabidopsis thaliana* cold shock domain-containing protein 3; BLAST: Basic Local Alignment Search Tool; CBP80: Cap-binding protein 80 or ABA hypersensitive 1; D or Asp: Aspartate; drRB: Drought stress responsive RNA-binding protein; drRPPome: Drought stress responsive RNA-binding proteome; FDR: False discovery rate; Fe: Iron; GAPC2: Glyceraldehyde 3-phosphate dehydrogenase C-2; GO: Gene ontology; GR-RBP8: Glycine-rich RNA-binding protein 8; hr: Hour; IRP1: Iron regulatory protein 1; KEGG: Kyoto encyclopedia of genes and genomes; LC: Liquid chromatography; LSm: Like-Smith; MS/M: Tandem mass spectrometry; N or Asn: Asparagine; NTF2: Nuclear transport factor 2; PDHA: Pyruvate dehydrogenase E1 component α-subunit; PEG: Polyethylene glycol; p*I*: Isoelectric point; PPR: Pentatricopeptide repeat; PSMs: Peptide spectrum matches; RBD: RNA-binding domain; RBP: RNA-binding protein; RRM: RNA recognition motif; S or Ser: Serine; S: Sulphur; T or Thr: Tryptophan; TAIR: The Arabidopsis Information Resource; UHPLC: Ultra high performance liquid chromatography; UV: Ultra violet; *v*/*v*: Volume/volume; ZF: Zinc finger

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
CM, CG, KL and LT designed the research; CM performed all the experiments and data analysis. KL and LT provided technical assistance to CM. CM and CG drafted manuscript. All authors participated in revising the manuscript and approved the final version.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not Applicable

Competing interests
The authors declare that the research was conducted in the absence of any commercial or financial affiliations that could be construed as a potential conflict of interest.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Cambridge Centre for Proteomics, Cambridge Systems Biology Centre, and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK. [2]Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, 23955-6900 Thuwal, Kingdom of Saudi Arabia. [3]Department of Chemistry, Biology and Biotechnology, University of Perugia, 74 Borgo XX Giugno, 06121 Perugia, Italy. [4]Present Address: Laboratoire de Physiologie Cellulaire et Végétale, Université Grenoble Alpes, CEA/DRF/BIG, INRA UMR1417, CNRS UMR5168, 38054 Grenoble Cedex 9, France. [5]Present Address: HM.Clause, rue Louis Saillant, 26802 Portes-lès-Valence, France.

References
1. Cole CN. Choreographing mRNA biogenesis. Nat Genet. 2001;29:6–7.
2. Weber C, Nover L, Fauth M. Plant stress granules and mRNA processing bodies are distinct from heat stress granules. Plant J. 2008;56:517–30.
3. Castello A, Fischer B, Hentze MW, Preiss T. RNA-binding proteins in Mendelian disease. Trends Genet. 2013;29:318–27.
4. Lukong KE, Chang KW, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. Trends Genet. 2008;24:416–25.
5. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell. 2009;138:673–84.
6. Kedersha N, Anderson P. Stress granules: sites of mRNA triage that regulate mRNA stability and translatability. Biochem Soc Trans. 2002;30:963–9.
7. Kedersha N, Cho MR, Li W, Yacono PW, Chen S, Gilks N, Golan DE, Anderson P. Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. J Cell Biol. 2000;151:1257–68.
8. Kedersha NL, Gupta M, Li W, Miller I, Anderson P. RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules. J Cell Biol. 1999;147:1431–42.
9. Ivanov PA, Nadezhdina ES. Stress granules: RNP-containing cytoplasmic bodies springing up under stress. The structure and mechanism of organization. Mol Biol. 2006;40:937–44.
10. Anderson P, Kedersha N. RNA granules. J Cell Biol. 2006;172:803–8.
11. Castello A, Horos R, Strein C, Fischer B, Eichelbaum K, Steinmetz LM, Krijgsveld J, Hentze MW. System-wide identification of RNA-binding proteins by interactome capture. Nat Protoc. 2013;8:491–500.
12. Marondedze C, Thomas L, Serrano NL, Lilley KS, Gehring C. The RNA-binding protein repertoire of *Arabidopsis thaliana*. Sci Rep. 2016;6:29766.
13. Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume AM, Schwarzl T, Curk T, Foehr S, Huber W, Krijgsveld J, Hentze MW. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat Commun. 2015;6:10127.
14. Kwon SC, Yi H, Eichelbaum K, Fohr S, Fischer B, You KT, Castello A, Krijgsveld J, Hentze MW, Kim VN. The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol. 2013;20:1122–30.
15. Liepelt A, Naarmann-de Vries IS, Simons N, Eichelbaum K, Fohr S, Archer SK, Castello A, Usadel B, Krijgsveld J, Preiss T, Marx G, Hentze MW, Ostareck DH, Ostareck-Lederer A. Identification of RNA-binding proteins in macrophages by interactome capture. Mol Cell Proteomics. 2016;15:2699–714.
16. Sysoev VO, Fischer B, Frese CK, Gupta I, Krijgsveld J, Hentze MW, Castello A, Ephrussi A. Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. Nat Commun. 2016;7:12128.

17. Baltz AG, Munschauer M, Schwanhausser 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. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell. 2012;46: 674–90.

18. Reichel M, Liao Y, Rettel M, Ragan C, Evers M, Alleaume AM, Horos R, Hentze MW, Preiss T, Millar AA. In planta determination of the mRNA-binding proteome of Arabidopsis etiolated seedlings. Plant Cell. 2016;28: 2435–52.

19. Zhang Z, Boonen K, Ferrari P, Schoofs L, Janssens E, van Noort V, Rolland F, Geuten K. UV crosslinked mRNA-binding proteins captured from leaf mesophyll protoplasts. Plant Methods. 2016;12:42.

20. Pashev IG, Dimitrov SI, Angelov D. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. Trends Biochem Sci. 1991;16:323–6.

21. Steen H, Jensen ON. Analysis of protein-nucleic acid interactions by photochemical cross-linking and mass spectrometry. Mass Spectrom Rev. 2002;21:163–82.

22. Ambrosone A, Batelli G, Nurcato R, Aurilia V, Punzo P, Bangarusamy DK, Ruberti I, Sassi M, Leone A, Costa A, Grillo S. The Arabidopsis RNA-binding protein AtRGGA regulates tolerance to salt and drought stress. Plant Physiol. 2015;168:292–306.

23. Daszkowska-Golec A, Wojnar W, Rosikiewicz M, Szarejko I, Maluszynski M, Szweykowska-Kulinska Z, Jarmolowski A. Arabidopsis suppressor mutant of abh1 shows a new face of the already known players: ABH1 (CBP80) and ABI4-in response to ABA and abiotic stresses during seed germination. Plant Mol Biol. 2013;81:189–209.

24. Hugouvieux V, Kwak JM, Schroeder JI. An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in Arabidopsis. Cell. 2001;106:477–87.

25. Papp I, Mur LA, Dalmadi A, Dulai S, Koncz C. A mutation in the cap binding protein 20 gene confers drought tolerance to Arabidopsis. Plant Mol Biol. 2004;55:679–86.

26. Thieme CJ, Rojas-Triana M, Stecyk E, Schudoma C, Zhang W, Yang L, Minambres M, Walther D, Schulze WX, Paz-Ares J, Scheible WR, Kragler F. Endogenous Arabidopsis messenger RNAs transported to distant tissues. Nat Plants. 2015;1:15025.

27. Jacob Y, Mongkolsiriwatana C, Veley KM, Kim SY, Michaels SD. The nuclear pore protein AtTPR is required for RNA homeostasis, flowering time, and auxin signaling. Plant Physiol. 2007;144:1383–90.

28. Berry MD, Boulton AA. Glyceraldehyde-3-phosphate dehydrogenase and apoptosis. J Neurosci Res. 2000;60:150–4.

29. Tristan C, Shahani N, Sedlak TW, Sawa A. The diverse functions of GAPDH: views from different subcellular compartments. Cell Signal. 2011;23:317–23.

30. Bae MS, Cho EJ, Choi EY, Park OK. Analysis of the Arabidopsis nuclear proteome and its response to cold stress. Plant J. 2003;36:652–63.

31. Guerrero FD, Jones JT, Mullet JE. Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted. Sequence and expression of three inducible genes. Plant Mol Biol. 1990;15:11–26.

32. Kirch HH, Nair A, Bartels D. Novel ABA- and dehydration-inducible aldehyde dehydrogenase genes isolated from the resurrection plant *Craterostigma plantagineum* and *Arabidopsis thaliana*. Plant J. 2001;28:555–67.

33. Sunkar R, Bartels D, Kirch HH. Overexpression of a stress-inducible aldehyde dehydrogenase gene from *Arabidopsis thaliana* in transgenic plants improves stress tolerance. Plant J. 2003;35:452–64.

34. Kotchoni SO, Kuhns C, Ditzer A, Kirch HH, Bartels D. Over-expression of different aldehyde dehydrogenase genes in *Arabidopsis thaliana* confers tolerance to abiotic stress and protects plants against lipid peroxidation and oxidative stress. Plant Cell Environ. 2006;29:1033–48.

35. Kirch HH, Schlingensiepen S, Kotchoni S, Sunkar R, Bartels D. Detailed expression analysis of selected genes of the aldehyde dehydrogenase (ALDH) gene superfamily in *Arabidopsis thaliana*. Plant Mol Biol. 2005;57:315–32.

36. Hentze MW, Argos P. Homology between IRE-BP, a regulatory RNA-binding protein, aconitase, and isopropylmalate isomerase. Nucleic Acids Res. 1991; 19:1739–40.

37. Murgia I, Delledonne M, Soave C. Nitric oxide mediates iron-induced ferritin accumulation in Arabidopsis. Plant J. 2002;30:521–8.

38. Navarre DA, Wendehenne D, Durner J, Noad R, Klessig DF. Nitric oxide modulates the activity of tobacco aconitase. Plant Physiol. 2000;122:573–82.

39. Marondedze C, Turek I, Parrott B, Thomas L, Jankovic B, Lilley KS, Gehring C. Structural and functional characteristics of cGMP-dependent methionine oxidation in *Arabidopsis thaliana* proteins. Cell Commun Signal. 2013;11:1.

40. Sweetlove LJ, Heazlewood JL, Herald V, Holtzapffel R, Day DA, Leaver CJ, Millar AH. The impact of oxidative stress on Arabidopsis mitochondria. Plant J. 2002;32:891–904.

41. Köster T, Marondedze C, Meyer K, Staiger D. RNA-binding proteins revisited – the emerging Arabidopsis mRNA interactome. Trends Plant Sci. 2017;22:512–26.

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

43. Kim MH, Sasaki K, Imai R. Cold shock domain protein 3 regulates freezing tolerance in Arabidopsis thaliana. J Biol Chem. 2009;284:23454–60.

44. Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. 2008;13:663–70.

45. Shen C, Zhang D, Guan Z, Liu Y, Yang Z, Yang Y, Wang X, Wang Q, Zhang Q, Fan S, Zou T, Yin P. Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. Nat Commun. 2016;7:11285.

46. Jiang SC, Mei C, Liang S, Yu YT, Lu K, Wu Z, Wang XF, Zhang DP. Crucial roles of the pentatricopeptide repeat protein SOAR1 in Arabidopsis response to drought, salt and cold stresses. Plant Mol Biol. 2015;88:369–85.

47. Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G. HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. Plant Cell. 2003; 15:1480–95.

48. Lamberti G, Gugel IL, Meurer J, Soll J, Schwenkert S. The cytosolic kinase STY8, STY7 and STY46 are involved in chloroplast differentiation in *Arabidopsis thaliana*. Plant Physiol. 2011;157:70–85.

49. Marondedze C, Wong A, Groen A, Serrano N, Jankovic B, Lilley K, Gehring C, Thomas L. Exploring the Arabidopsis proteome: influence of protein solubilization buffers on proteome coverage. Int J Mol Sci. 2014;16:857–70.

50. Ordonez NM, Marondedze C, Thomas L, Pasqualini S, Shabala L, Shabala S, Gehring C. Cyclic mononucleotides modulate potassium and calcium flux responses to H_2O_2 in Arabidopsis roots. FEBS Lett. 2014;588:1008–15.

51. Marondedze C, Groen AJ, Thomas L, Lilley KS, Gehring C. A quantitative phosphoproteome analysis of cGMP-dependent cellular responses in *Arabidopsis thaliana*. Mol Plant. 2016;9:621–3.

52. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc. 1995;57:289–300.

53. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428:726–31.

54. Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition profiler: a tool for discovery and visualization of amino acid composition differences. BMC Bioinformatics. 2007;8:211.

55. Sonnhammer EL, Ostlund G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. 2015;43:D234–9.