# RNA interactome capture in Brachypodium reveals a flowering plant core RBPome

**Meixia Li**
Katholieke Universiteit Leuven
**Zhicheng Zhang**
Vlaams Instituut voor Biotechnologie Department of Plant Systems Biology
**Sam Balzarini**
Katholieke Universiteit Leuven Departement Biologie
**Bhavesh Parmar**
Katholieke Universiteit Leuven Departement Biologie
**Boonen Kurt**
VITO NV
**Anja Vandeperre**
Katholieke Universiteit Leuven Departement Biologie
**Koen Geuten** (koen.geuten@kuleuven.be)
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences https://orcid.org/0000-0002-9692-4499

**Research article**

**Keywords:** Brachypodium distachyon, RNA-binding proteins (RBPs), RNA interactome capture (RIC), candidate RBPs, classic RBPs, core RBPome

**Posted Date:** August 25th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-23468/v1

# Abstract

Background

RNA binding proteins regulate gene expression at the post-transcriptional level by controlling the fate of RNA, in processes such as mRNA localization, translation, splicing and stability. The annotation of RNA binding proteins is mainly based on the well-known RNA binding domains and motifs. However, novel RNA binding proteins without such conventional domains have been identified in different species using *in vivo* RNA interactome capture. To find support for novel conserved RNA binding proteins in plants, we applied an optimized RNA interactome capture to the monocot model *Brachypodium distachyon.*

Results

We provide experimental evidence for 203 RNA binding proteins isolated from Brachypodium shoot tissue and leaf mesophyll protoplasts, and grouped these into classic RNA binding proteins with recognizable RNA binding domains and motifs, and candidate RNA binding proteins without such domains. Compared to RNA binding proteins captured in *Arabidopsis thaliana*, candidate RNA binding proteins involved in carbon fixation and carbon metabolic pathways are highly conserved. We tried to validate the RNA binding proteins captured in this research through a silica-based method, but this method appears not efficient for plants. This may indicate that optimized methods to validate high throughout RNA binding proteome are required for plants.

Conclusions

Our results provide classic and candidate RNA binding proteins in *Brachypodium distachyon* and conserved RNA binding proteins in flowering plants. Future functional characterization should point out what the significance of RNA binding is for the function of these proteins.


# Background

Post-transcriptional regulation of gene expression is essential for growth and development as well as for stress responses in eukaryotes [1]. RNA binding proteins are indispensable chaperone proteins during this process. They mainly participate in RNA metabolism, including splicing of pre-RNA, RNA decay and stability, and localization [2]. RNA binding domains (RBDs) are commonly used to annotate RBPs *in silico* based on the understanding of domain structure and function. Classic RBDs include the RNA recognition motif (RRM), the K homology (KH) domain [3, 4], the DEAD box and COLD SHOCK domain [5, 6] etcetera.

However, many RBPs missing classic RBDs, such as metabolic enzymes, have been identified using immobilized RNA probes or arrayed proteins *in vitro* ([7–9] or RNA interactome capture (RIC) *in vivo* [10]. In RIC, oligo d(T) beads are used to capture the mRNA-protein complexes, which are crosslinked by short wavelength UV and identified through mass spectrometry after being purified in denaturing conditions. RIC has been applied to different species and tissues, from mammalian cells [10, 11], yeast [12],

*Drosophila* [13] to the plant Arabidopsis [14, 15]. RIC yielded 797 RBPs in HEK293 cells, but 245 of these RBPs captured proteins were not annotated as RNA binding [10]. In *Drosophila*, 523 RBPs were identified using RIC, but over half the RBPs were not known to bind RNA [13]. Notable RBPs lacking classic RBDs like enzymes ENO1 and SHMT2 in HEK293 cells [10], TRIM proteins in embryonic stem cells [16] and GAPDH in T cells [17] were confirmed to indeed bind RNA. The emerging numbers of unconventional RBPs may reveal a new era in the function of RBPs.

In Arabidopsis, RIC has been applied to different tissues, like suspension cells (913 RBPs) and seedlings (236 RBPs) [14], leaf mesophyll protoplasts (325 RBPs) [18], and etiolated seedlings (746 RBPs) [15]. It is difficult to compare the yield of RBPs between different experiments because different criteria for protein identification were used. In these Arabidopsis experiments, RBPs with classic RBDs were enriched after RIC as expected, but again hundreds of novel RBPs without known RBDs were identified in Arabidopsis, such as DUF domain proteins and LIM domain proteins [15]. In this research, we optimized the UV dosage and captured RBPs using RIC from the monocot model plant *Brachypodium distachyon*. By making the comparison to Arabidopsis, we hope to contribute support for conserved novel RBPs in flowering plants, which may enhance the understanding of the identification and functions of RBPs in the future.

## Results

# High UV dosage does not degrade RNA but yields less RBPs

Though RIC has been successfully applied to Arabidopsis cell suspension, etiolated plants, seedlings and leaf mesophyll protoplasts, the number of identified proteins is lower than that in mammalian cells [24]. This could be because of a difference in crosslinking efficiency between plants and mammalian cells. Plant tissues contain a waxy cuticle and light acceptors in the chloroplast, which can absorb short wavelength UV light [25] and could lower the crosslinking efficiency. The leaf anatomy of the model plant Arabidopsis is different from monocot plants like wheat and rice [26–28], which may indicate the different UV efficiency for different species.

Therefore, we want to optimize the UV dosage in the monocot model Brachypodium to increase the efficiency of RIC in monocot plants (Fig. 1A). We used Brachypodium leaf mesophyll protoplasts and 2-week old seedlings as starting materials. For leaf mesophyll protoplasts, we applied the same method as previously used for the Arabidopsis leaf mesophyll protoplasts [18]. For seedlings, the UV dosage for Brachypodium was optimized in order to increase the efficiency of crosslinking (Fig. 1A). Different UV treatments and dosages were used including 0.9J/cm$^2$, two times of 0.9J/cm$^2$ treatments, continuous 10J/cm$^2$, two times of 5J/cm$^2$ treatments with turning over plants and four times of 5J/cm$^2$ treatments (Fig. 1B). As derived from the RIN numbers obtained by Bioanalyzer analysis, none of the UV treatments or dosages resulted in RNA degradation, as the average RIN number from three independent biological replicates was above 9 for all conditions (Fig. 1C). After RIC were applied to all seedling samples with different UV treatments, the results of visualization of proteins on the silver stained SDS-PAGE gel suggested that the CL samples were enriched in proteins compared to the noCL samples (Fig. 1B). Also, a

higher UV intensity treatment showed more pronounced protein bands on silver staining, which suggests a possible higher concentration of proteins (Fig. 1B).

After mass spectrometry detection and protein analysis, we identified 106 RBPs under 0.5J/cm$^2$ UV two times exposure and 55 RBPs under 5J/cm$^2$ UV four times exposure (Fig. 1D), of which 39 RBPs overlap (Fig. 1D). 39 RBPs showed in both UV treatment conditions and 85% overlapped proteins contains the RNA binding domains or motifs (Table S2). Surprisingly, the number of RBPs identified in the lower UV condition was higher than the number in the higher UV condition, which is not consistent with the staining intensity on the SDS-PAGE gel (Fig. 1B). One possibility is that UV overexposure can activate the repair mechanism for transcripts, which degrades RNA-protein complexes [29–31]. In this case, only the abundant RNA-protein can be captured using RIC, which is consisted with that the overlapped proteins containing the RNA binding domains or motifs.

In addition, we identified 128 RBPs in leaf mesophyll protoplasts. We pooled the RBPs from shoots in different UV dosage conditions (in total 123 RBPs), which are used to compare with the RBPs from leaf mesophyll protoplasts (Fig. 1D). There are 47 RBPs overlapped between shoots and protoplasts and 85% RBPs contain RNA binding domains or motifs. For further analysis, we pooled identified RBPs from different samples and UV treatment and in total, 203 RBPs were detected in Brachypodium (Table S2).

# Conserved RNA binding domains distinguish classic from candidate RBPs

Of the 203 RBPs identified in Brachypodium, 112 proteins (55%) with known RBDs were grouped as classic RBPs. In classic RBP group, RRM domains are the largest group, followed by ribosomal proteins. The other classic RBPs include the DEAD-Helicase, COLD SHOCK domain, zf-CCCH domain, YTH domain, S1 domain, KOW domain, KH domain, zf-RNABP domain and telomerase RNA binding domain (Fig. 2). The other 91 identified protein (45%) lacking these classic domains or motifs were defined as candidate RBPs. In candidate RBPs group, the largest one contains the information in the photosynthesis, such as photosystem I reaction center subunit D and chloral A/B binding proteins (Fig. 2). However, whether these proteins are involved in RNA metabolism is unclear.

### RBPs with conserved RBDs were identified in classic RBPs group and novel RBPs were found in candidate RBPs group

The detailed RBDs in both classic and candidate RBPs groups were listed in Fig. 2 and Table S2. In classic RBPs group (Fig. 3), the largest group was proteins containing RRM domains (41). In Arabidopsis, proteins with RRM domains through RIC were the largest group in canonical RBPs as well [14, 15]. We identified the classic RNA binding zinc finger proteins, such as Znf-CCCH type and Znf-RanBP2 (Fig. 3). The Znf-RanBP2 protein, also captured in Arabidopsis etiolated seedlings [15], has been reported to bind single strand RNA in human [32]. Other emerging and new RBPs in plants were identified in this research, such as two YTH-domain proteins and three alba proteins, which were also found in Arabidopsis [14, 15]. YTH domain proteins (ECT2/3) in Arabidopsis have been proved to recognize the m$^6$A through binding to

target RNA 3'UTR, and regulate the trichome morphogenesis and leaf development through facilitating the degradation of the ECT2 binding transcripts [33, 34]. Alba domain was closely related to the ancient RNA-binding IF3-C fold [35] and was accepted as multifunctional proteins that participate in genome organization, translational process, RNA metabolism [36]. 28 proteins containing YTH domain were predicted in Brachypodium through Interpro (http://www.ebi.ac.uk/interpro/entry/IPR007275/proteins-matched?taxonomy=15368), but the function has not been reported in Brachypodium.

Except for the classic RBPs, 45% of RBPs are novel RNA binding proteins in Brachypodium, for which no RNA binding function has been proposed previously. Eleven proteins involved in involved in photosystems were strongly enriched in candidate RBPs group. In addition, RBPs containing DUF, ATPase and UspA domain proteins were also identified. Actin family proteins were identified in Brachypodium, which were found in Arabidopsis etiolated seedlings as well [15].

Similar to captured RBPs through RIC in other species, enzymes were captured as well in Brachypodium, such as serine fructose-bisphosphate aldolase (FBPase) and transketolase, which are involved in the Calvin cycle for carbon fixation. Combining with the photosystem proteins, it seems the RBPs in chloroplasts are enriched in RNA binding activity. Glutamine synthetase, glycosyltransferase, pectinesterase, transferase and cellulose synthase are also captured in Brachypodium. It is very interesting to know whether these enzymes captured in this research are really binding to RNA or not.

# GO and KEGG enrichment analysis

According to the biological molecular function analysis (GO analysis), the enriched proteins in classic RBPs were annotated to have RNA binding activity ($p < 0.05$), including mRNA binding and Structural constituent of ribosome (Fig. 3A and Table S3). The subcategories of mRNA binding were RNA binding, nucleic acid binding, organic cyclic compound binding and heterocyclic compound binding (Fig. 2A). The enriched RNA binding function among the classic RBPs demonstrated the high efficiency of this interactome capture and the accuracy of the domain search method for analysis. For the candidate RBPs, the most enriched molecular biological functions were molecular function, catalytic activity and chlorophyll binding (Fig. 2B). However, when comparing the GO enrichment in these two groups, the GO enrichment of classic group is much higher than candidate group.

In addition, the KEEG pathway enrichment was analyzed for classic and candidate RBPs (Table S3). Except for the ribosomal proteins, the most enrichment KEEG were mRNA surveillance pathway, spliceosome, RNA transport and RNA degradation. Similar to the result of GO enrichment analysis, the method to classify classic RBPs and candidate RBPs based on the domain information is efficient and accurate. For the candidate RBPs, the most enrichment in KEGG pathway were metabolic pathways, carbon metabolism, photosynthesis, glyoxylate and dicarboxylate metabolism, fructose and mannose metabolism, biosynthesis of secondary metabolites, biosynthesis of amino acids. It appears more enzymes are detected in the candidate RBPs. However, the validation for candidate RBPs are required to test their RNA binding ability.

# A core RBPome identifies RBPs in plant-specific processes

RIC was first applied to the model plant Arabidopsis and RBPs were identified in Arabidopsis leaf mesophyll protoplasts (325) [18], cell suspension (913) and leaves (236) [14], and etiolated seedlings (746) [15] (Summarized in Fig. 4A). Different criteria and methods used to identify RBPs in each article make it difficult to combine these data. Comparing with all RBPs identified in Arabidopsis, the largest RBPs overlap is between leaf mesophyll protoplasts and etiolated seedlings (129 RBPs). Since shoot tissue and leaf mesophyll protoplasts were used as materials to identify RBPs in Brachypodium, it is meaningful to compare with RBPs identified in the same tissue of Arabidopsis. Thus, we pooled the RBPs data from Arabidopsis leaf and mesophyll protoplasts in order to compare with the RBPs identified in Brachypodium.

Using Inparanoid analysis to compare orthologs between Arabidopsis and Brachypodium, 10689 clusters from both species were extracted as fundamental database for further analysis. 343 protein clusters from Arabidopsis RBPs of leaf and leaf mesophyll protoplasts were matched when compared to Arabidopsis fundamental database, while 130 protein clusters of Brachypodium RBPs were matched to the Brachypodium fundamental database. When comparing the orthologs of the subsets, 57 clusters were defined, which include 82 RBPs from Arabidopsis and 62 RBPs from Brachypodium. We termed these RBPs as core RBPs (Fig. 4B and Table S4.

KEGG analysis was used to compare the core RBPs in Brachypodium and Arabidopsis (Fig. 4C and Table S4). The largest RBPs number belong to ribosomes as expected, which is followed by a carbon fixation in photosynthetic organism and carbon metabolism. These enzymes are highly conserved between Arabidopsis and Brachypodium. In addition, the proteins involved in mRNA surveillance pathway, RNA transport and spliceosome are highly conserved in both species as well. It is interesting to explore the roles of these conserved core RBPs in plants. Interestingly, RNA degradation is absent in Arabidopsis core RBPs, which may reveal the RIC method is limited to capture all RBPs in plants.

# Validation methods in the plant RBPs need further optimization

To validate the candidate RBPs identified in this research, we applied silica-based solid-phase extraction followed by a western blot [28]. Firstly, we transformed the selected RBPs fused with green fluorescent protein (GFP) tag (35S: RBP: GFP) into Brachypodium leaf mesophyll protoplasts and UV treatment was applied after incubation. Silica matrix-based columns were used for nucleic acid purification, in which the crosslinked nucleic acid-protein complexes are retained. The RNA-protein interaction can be visualized using western blot with GFP-antibody to detect the GFP-tag. In order to confirm whether the silica columns capture the RNA-protein complexes, we visualized the isolated proteins from CL and noCL treated leaf mesophyll protoplasts with silver staining (Fig. 5A). The CL samples contained enriched proteins compared to noCL samples, which indicates that the combination of silica column and UV crosslinking can be used for validation. In this research, we randomly selected both classic RBPs (Bradi1g20440.1 (RRM-domain protein), Bradi2g61835 (LSM-domaim), Bradi5g17360.1 (PPR-domain)) as positive controls and candidate RBPs (Bradi4g08800.1 (carbon-fixation). As seen on the western blot

(Fig. 5B), the GFP tagged proteins well expressed in protoplasts, but the target RBPs could not be detected after silica purification of the RNA-protein complexes of crosslinked samples. The possible reasons will be discussed as follows.

# Discussion

RNA interactome capture has been widely used in different species to explore their repertoires of RNA-binding proteins ([10–15]. An interesting follow-up question is whether different phylogenetic groups of species, such as plants in this instance, have a derived repertoire of RBPs or not. A second and key question in these studies is also whether the identified RBPs without a canonical RNA binding domain actually bind RNA or whether they concern a sort of contamination. To shed light on these questions, we investigated the conservation of interactome captured proteins between two plants from divergent lineages: Arabidopsis (a eudicot) and Brachypodium (a monocot), and attempted to validate a subset of these proteins *in vivo*.

In Brachypodium, we identified 203 RBPs from leaf mesophyll protoplasts and seedling tissue. The method we applied to classify the identified RBPs into classic and candidate groups is based on the known RNA binding domains, motifs. The high enrichment of GO analysis and KEGG pathway analysis in the RNA binding ability ensure that this method is efficient and accurate. 55% RBPs were appointed to be classic RBPs with known RNA binding domains and motifs, while 45% of RBPs were novel RBPs without these domains and motifs. The highly enriched photosynthesis proteins in candidate RBPs group could indicate that these proteins somehow have a thus far uncharacterized direct interaction with RNA. Alternatively, the applied method could fool us in some ways. One possible way is that under UV exposure, the light reaction process is strongly active and abundant photosystem proteins are easily captured while being translated from mRNA. Another possible way is that these proteins are contamination and illustrate a limitation of UV crosslinking, which could crosslink the free mRNA to the abundant photosynthesis proteins. A comparison of the abundance of proteins in the general proteome to the RBPome could shed some light on these possibilities. In order to avoid this capture of chloroplast proteins during UV treatment, RIC was also applied to Arabidopsis etiolated plants and cell suspension materials, which seems to yield more RBPs [14, 15]. This may indicate that the UV crosslinking efficiency is lower in green tissues. Therefore, the etiolated plants and cell suspension experiments may be useful to capture RBPs more efficiently. However, plants grown in darkness are similar to phytochrome-deficient mutants [37] and show senescence [38]. Therefore, the etiolated plants and cell suspension materials cannot be used for all research questions.

Of 203 RBPs captured in Brachypodium, 62 (30%) RBPs were conserved with RBPs captured in Arabidopsis seedlings and leaf mesophyll protoplasts. In these core RBPs, proteins in carbon fixation are highly conserved in both species. However, it is interesting to validate whether the conserved enzymes identified through RIC and conserved in Arabidopsis and Brachypodium can bind to RNA or not. Though over 50% of RBPs were identified to harbor conserved RBDs or motifs, still 45% of RBPs identified in this research were lacking these known characteristics. We describe potential RNA binding capacity of some

selected proteins. The proteins with CLU domain proteins contain tetratricopeptide repeat-containing domain (TRP domain), which can bind to mRNAs of multiple genes such as *OsPAL* (*OsPAL1–7*) in rice to promote the turnover of these mRNA [39]. This information may indicate the possibility of CLU domain proteins to bind RNA. Another interesting protein is UspA domain protein, which is upregulated by stress to protect DNA damage in bacteria [40]. The identification of UspA domain protein through mRNA interactome capture may reveal its regulation mechanism to enhance cell survival rate under stress. We also captured actin family proteins in Brachypodium, which were identified in Arabidopsis etiolated seedlings as well [15]. Actin proteins, one kind of the most abundant proteins in eukaryotic cells, are well known to assist cell mobility and scaffold in the cytosol [41]. But nuclear actin proteins are well documented to play a role in transcription, chromatin remodeling, histone modification and DNA damage response [42]. Therefore, it is meaningful to validate these captured RBPome proteins.

The failure to validate RBPs *in vivo* in our research makes it difficult to ensure that these novel RBPs have RNA binding ability or not. We attempted to validate the RBPs captured *in vivo* in this research using silica column-based method [23]. The failure to detect the expected bands by western blot from silica samples could be explained by different reasons relating to sensitivity. First, the efficiency of UV crosslinking in leaf mesophyll protoplasts could be too low so that not enough RNA-protein complexes are detectable. Second, the canonical RNA of the RBPs is not abundant enough to be detected in spite of the fact that the tagged proteins are overexpressed. Indeed, in the original article describing the method, the yield of RNA after silica elution reached up to 12.5ug [23]. In comparison, the RNA yield from protoplasts only reaches 2 μg maximum with around $10^8$ cells per sample. To overcome the above problems, a large number of protoplasts is required. However, the low yield of protoplast in Brachypodium (40 ~ 60 seedlings can produce $1.7 \times 10^7$ cells [19] appears not suitable for the purpose of validating several RBPs. Therefore, we need alternative validation strategies. Though Marondedze et al. [14] validated the DEAD-box RBPs via oligo (dT) beads and western blot, this approach is not fully independent because the same method is used in both RBPs capture and validation. Compared to mammalian cells or yeast, which can be transformed and crosslinked easily, it appears more difficult to validate the high throughput proteomics data in plants. Except for the *in vivo* methods, there are possible in vitro methods for validation, such as the electrophoretic mobility shift assay (EMSA). However, it appears time and labor consuming to validate the high throughput proteome.

The obvious limitation of RIC to capture RBPs is that it can only capture the RBPs bound to poly(A). XRNAX, named for protein crosslinked RNA extraction, can characterize proteins bound to coding and noncoding RNA [43]. Applying XRNAX method to plant cells could possibly unravel cellular mRNA interactome more comprehensively beyond the limitation of yielding the mature mRNAs only. Another issue for RBPomes are the data analysis methods, to compare the proteins in CL versus noCL samples. The ratio of CL/noCL has been used to quantify the significance of proteins in both CL and noCL [10, 13, 15]. However, for the proteins only captured in CL, it is impossible to use this method because of missing noCL data. In Arabidopsis etiolated seedlings, proteins only captured in CL were catalogued as candidate RBPs [15]. While a semi-quantitative analysis based on the number of peptide occurrences was used to

characterize these proteins in *Drosophila* [13], the proteins only identified in CL were also recognized as significant ones. In our research, most of the proteins captured were only presented in CL samples and absent in noCL samples. We used the limma t-test to characterize the proteins showing in both CL and noCL, while it is meaningless to do this for the proteins only identified in CL. We grouped proteins based on their domains and motifs. But more detailed and accurate methods are required to quantify these RBPomes.

In addition, discovering the key regulators and understanding their functions in stress signalling pathways is still valuable to improve the crops stress tolerance in the future. As previous studies of individual RBPs have already demonstrated the diverse stress tolerance [44–46], it is interesting to globally detect the shaping of RBPomes under certain stresses and thereby to find out stress-sensitive RBPs. Recently, RIC has been applied to identify the drought stress RBPs and 150 out of 546 significantly responded to drought [47]. While application of RIC in plants has been limited, using RIC in abiotic and biotic stress experiments can reveal novel RBPs, which may disclose new regulatory mechanisms.

## Conclusions

In this study, we identified 203 RNA binding proteins from Brachypodium shoot tissue and leaf mesophyll protoplasts and conserved RNA binding proteins in both Arabidopsis and Brachypodium. These conserved RNA binding proteins were found to be involved in carbon fixation and carbon metabolic pathways. We provided the conserved RNA binding proteins in flowering plants with modified RNA interactome capture, which paved way to explore more functional RNA binding proteins in different plant species.

## Methods

# Plant materials and growth condition

*Brachypodium distachyon* accession Bd21-3 was grown in a 12/12 h 22 °C growth room and 4-week old plant leaves were used to isolate leaf mesophyll protoplasts for the RIC assay. Seedlings from *Brachypodium distachyon* accession Bd21 were grown in a 16/8 h, 22 °C growth room and 2-week old plants were used for the RIC assay. For validation experiment, Bd21 plants were grown in a 16/8 h, 22 °C growth room and the leaf of 2-week old seedlings was used to isolate the leaf mesophyll protoplasts. All replicates used in each experiment were taken at the same time and randomly from different pots.

# UV crosslinking and RNA interactome capture in leaf mesophyll protoplasts

For RIC applied to leaf mesophyll protoplasts, the protoplasts isolation was performed using an optimized protocol [19–21]. Briefly, leaves of 4-week old seedlings were cut into small pieces and incubated in enzyme solution (Mannitol 0.8M, MES 0.1M, PH 5.7, R10 cellulase 1.5% (Duchefa Biochemie), macerozyme 0.75%, BSA 1%, $CaCl_2$ 1M, $\beta$-mercaptoethanol 5%) for 4 hours after 30 minutes

vacuum. Collected protoplasts were exposed to 254 nm UV (UVP Ultraviolet Crosslinkers with 5 × 8-watt UV dual bipin discharge type tubes) and irradiated by 0.13 J/cm$^2$ (about 1 min) before lysis. The detailed method to capture the mRNA binding proteins followed the protocol by Zhang et al. (2016). Three independent biological replicates for non-crosslinked (noCL) samples and crosslinked (CL) samples were used in this research.

# UV crosslinking and RNA interactome capture in seedlings

For RIC applied to seedlings, 2 g of shoot tissue (without the root) per sample were exposed to 254 nm UV and irradiated at different UV intensities, including 5 J/cm$^2$ four times, 5 J/cm$^2$ two times with turning over the plants, continuous 10 J/cm$^2$, and 0.9 J/cm$^2$. The RNA was isolated from different UV treatments using TRIsure™ and RNA quality was tested using a Bioanalyzer. Three independent biological replicates of CL and noCL samples were used for RNA quality control. Four independent biological replicates exposed to 5 J/cm$^2$ four times UV with turning over plants each time and 0.5 J/cm$^2$ two times UV were used for RIC assay. After UV exposure, samples were immediately flash frozen and used for RIC.

A modified RIC method was applied to seedling samples. After UV crosslinking, the samples were ground for 30 minutes in liquid nitrogen and lysed in 30 ml lysis buffer (500 mM LiCl, 0.5% (w/v) Lithium Dodecyl Sulphate (LiDS), 5 mM DTT, 20 mM Tris–HCl, pH 7.5, and 1 mM EDTA, pH 8.0). After centrifugation, the supernatant was incubated with 4 mg Oligo-d(T)$_{25}$ beads at RT for 1 h. The RNA-protein-beads complexes were washed two times with both washing buffer I (500 mM LiCl, 0.1% (w/v) LiDS, 5 mM DTT, 20 mM Tris–HCl, pH 7.5, and 1 mM EDTA, pH 8.0) and washing buffer II (500 mM LiCl, 5 mM DTT, 20 mM Tris–HCl, pH 7.5, and 1 mM EDTA, pH 8.0), and one time with low salt buffer (200 mM LiCl, 20 mM Tris–HCl, pH 7.5, and 1 mM EDTA, pH 8.0). Then the mRNA-protein complexes were collected with 1 ml elution buffer (20 mM Tris–HCl, pH 7.5, and 1 mM EDTA, pH 8.0) after being incubated at 55 °C for 3 minutes. The capture was repeated two times from the same lysate. After pooling all eluates together, 200 µl was used for RNA quality control after 1 µl Proteinase K (NEB) treatment at 37 °C for 1 h. After treating the remaining 1.8 ml eluate with 1 µl RNase Cocktail™ Enzyme mix (ThermoFisher Scientific) for 1 h, samples were concentrated to 75 µl using Amicon® Ultra Centrifugal Filters. 25 µl was loaded on an SDS-PAGE gel for visualization and the rest was analyzed by LC-MS/MS after digestion.

# Protein digestion and mass spectrometry

Protein samples were digested and detected using LC-MS/MS as previously described [18]. Briefly, gel lanes were hydrated and dehydrated with 50 µL 100 mM NH4HCO3 and CH3CN for 10 min, respectively. Proteins in the gel pieces were digested overnight at 37 °C. Then the tryptic peptides were extracted using 80 µL of 50 mM $NH_4HCO_3$ and twice with 80 µL of 50% (w/v) $CH_3CN$ and 5% (v/v) formic acid (FA) for 30 min each. The samples were dried and dissolved in 25µL solution containing 2%(v/v) $CH_3CN$ and 0.1% (v/v) aqueous trifluoroacetic acid (TFA) and desalted with Millipore ZipTip C18 columns. The final eluent containing purified peptides was dissolved in 4 µL 60% (v/v) $CH_3CN$ and 0.1% (v/v) FA and dried again.

The LC–MS analysis was performed on a Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Scientific, Bremen, Germany), coupled online to an Ultimate 3000 ultra-high-performance liquid chromatography (UHPLC) instrument (Thermo Scientific, San Jose, CA). 5 µL sample was loaded on a precolumn (Acclaim PepMap100, C18, 75 µm x 20 mm, 3 µm, 100 Å; Thermo Fisher Scientific) at a flow rate of 5µL/min. Further, the sample was separated at a flow rate of 300 nl/min on an analytical column integrated with the nano-electrospray ion source (EASY-Spray, PepMap RSLC, C18, 75 µm x 150 mm, 3 µm, 100 Å; Thermo Fisher Scientific) over a 60 min gradient. Mobile phase A consisted of 99.9% H2O and 0.1% (v/v) FA and mobile phase B of 19.92% $H_2O$, 80% (w/v) $CH_3CN$ and 0.08% (v/v) FA. The mass spectrometer was operated in positive mode with data-dependent acquisition within the scan range of 350–1750 m/z. For each precursor spectrum, top 10 intense peaks were selected for fragmentation. For precursor spectra, resolving power of 70,000 full width at half maximum (FWHM) was used with automatic gain control (AGC) target of $1 \times 10^6$ ions and a maximum ion injection time (IT) of 100 ms. For fragmentation spectra, resolving power of 17,500 FWHM was used with an AGC target of $5 \times 10^5$ ions and a maximum IT of 180 ms. Dynamic exclusion of 30 s was applied for precursor selection and ions with charge 1 + and > 7 + were excluded from fragmentation.

**Identification of mRNA binding proteins for** Brachypodium distachyon

Four independent biological replicates of seedlings and three independent biological replicates of leaf mesophyll protoplasts for both CL and noCL were used for RIC capture and label-free quantification (LFQ) LC-MS/MS (Table S1). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD018076. We used MaxQuant and applied a FDR < 1% for peptides and protein identification with fixed and variable modifications. *Brachypodium distachyon* protein database from Uniprot Protein were used for protein identification. LFQ intensity was used for statistical analysis. Proteins with at least one peptide in two biological replicates were considered for further analysis. Proteins detected only in noCL were excluded from further statistical analysis and proteins detected only in CL samples were considered as positive ones. For proteins in both noCL and CL, we applied a limma moderated T-Test (R limma package) on LFQ intensities for the statistical analysis and proteins with adjusted P-value < 1% were considered as positive ones.

# Domain analysis and Inparanoid comparison of identified RBPs

The proteins only in CL samples and proteins with P-value < 1% in both CL- and noCL samples were annotated against the Interpro and Pfam database. Based on the known RBDs, motifs and RNA binding characteristics listed on the review by [22], we grouped the RBPs identified to classic group and candidate group. GO analysis and enrichment were performed using PANTHER (http://www.geneontology.org/page/go-enrichment-analysis).

# STRING

(https://string-db.org/cgi/input.pl?sessionId=8BeWlyUSyKaF&input_page_show_search=on) and KEGG mapper (https://www.kegg.jp/) were used to analyze the biological pathways and KEGG enrichment for RBPs identified in this research. The orthologs between *Arabidopsis thaliana* and *Brachypodium distachyon* were downloaded in Inparanoid8 (http://inparanoid.sbc.su.se/cgi-bin/index.cgi). Comparison between RBPs identified and orthologs of these two species were performed.

# Validation for RBPs identified using silica column and Brachypodium leaf mesophyll protoplasts

In order to validate the RBPs *in vivo*, we tried to apply a silica column treatment on Brachypodium leaf mesophyll protoplasts transformed with GFP-tagged RBPs. Firstly, the coding sequence of selected RBPs (showing in both shoot and protoplasts) was PCR-amplified without stop codon from Bd21 cDNA and inserted in the HBT95 expression vector containing a GFP tag in frame. The construct with GFP protein sequence was driven by the 35S promoter and terminated by the NOS terminator. All constructs were confirmed by sequencing (LGC Genomics, Berlin). Then we applied a PEG-$Ca^{2+}$-mediated transfection to protoplasts (Cells x $10^{-7}$=1) with 30 µg GFP-construct plasmid DNA. Protoplasts were incubated in dim light (light intensity around 8 µmol/$m^2s^2$) for 12 hours.

The validation method we applied is based on Asencio et al. [23]. After UV treatment (0.13 J/$cm^2$) to the incubated protoplasts, a commercial RNA extraction kit (InviTrap Spin Plant RNA Mini Kit) was used. The columns in this kit can retain the RNA and RNA-protein complexes [23]. After elution, RNase cocktail enzyme mix (ThermoFisher Scientific) was used to free the proteins and a western blot with anti-GFP antibody was used to detect the GFP tagged target proteins.

# Abbreviations

RBPs
RNA-binding proteins
RIC
RNA interactome capture
RBDs
RNA binding domains
RRM
RNA recognition motif
KH
K homology
TRP
tetratricopeptide repeat-containing
EMSA
electrophoretic mobility shift assay
noCL

non-crosslinked
CL
crosslinked
LiDS
Lithium Dodecyl Sulphate
AGC
automatic gain control
IT
injection time
LFQ
label-free quantification
GFP
green fluorescent protein

# Declarations

### Ethics approval and consent to participate

Not applicable

### Consent to publish

Not applicable

### Availability of data and materials

Data are available via ProteomeXchange with identifier PXD018076.

### Submission details:

Project Name: RNA interactome capture in Brachypodium reveals a flowering plant core RBPome

Project accession: PXD018076

Reviewer account details:

Username: reviewer10999@ebi.ac.uk

Password: 5YLYS0uu

### Competing interests

The authors declare that the research was conducted in the absence of any commercial and financial relationships that could be construed as a potential conflict of interest.

### Funding

We gratefully thank the supportive funding. ML was supported by the China Scholarship Council (CSC) for 4 years of study at KU Leuven. BS was supported by a FWO SB fellowship 1S06517N. Geuten lab were supported by KU Leuven grant C24/17/037. The funders did not play any roles in the design, analysis, and interpretation of this study or relevant data

### Author contributions

ML, ZhZh and KG designed the experiment. ML designed and carried out the shoot RIC assay, analyzed all the data, wrote the manuscript. ZhZh performed the leaf mesophyll protoplasts RIC assay. SB performed the RNA quality assay and VA cloned all the constructs in this experiment. PB and KB carried out the MS Spectrum assay. KG revised the manuscript. All authors contributed to the final manuscript.

### Acknowledgements

We gratefully thank for the SymbioMa KULeuven for the facilities and Filip Roland from KU Leuven for the GFP plasmid.

# References

1. Lee K, Kang H. Emerging Roles of RNA-Binding Proteins in Plant Growth, Development, and Stress Responses. Mol Cells. 2016;39:179–85. doi:10.14348/molcells.2016.2359.
2. Sutherland JM, Siddall NA, Hime GR, McLaughlin EA. RNA binding proteins in spermatogenesis: an in depth focus on the Musashi family. Asian J Androl. 2015;17:529–36. doi:10.4103/1008-682X.151397.
3. Albà MM, Pagès M. Plant proteins containing the RNA-recognition motif. Trends Plant Sci. 1998;3:15–21. doi:10.1016/S1360-1385(97)01151-5.
4. Lorković ZJ. Role of plant RNA-binding proteins in development, stress response and genome organization. Trends Plant Sci. 2009;14:229–36. doi:10.1016/j.tplants.2009.01.007.
5. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol. 2007;8:479–90. doi:10.1038/nrm2178.
6. Nakaminami K, Matsui A, Shinozaki K, Seki M. RNA regulation in plant abiotic stress responses. Biochim Biophys Acta. 2012;1819:149–53. doi:10.1016/j.bbagrm.2011.07.015.
7. Butter F, Scheibe M, Mörl M, Mann M. Unbiased RNA-protein interaction screen by quantitative proteomics. P Natl Acad Sci USA. 2009;106:10626–31. doi:10.1073/pnas.0812099106.
8. Scherrer T, Mittal N, Janga SC, Gerber AP. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. Plos One. 2010;5:e15499. doi:10.1371/journal.pone.0015499.
9. Tsvetanova NG, Klass DM, Salzman J, Brown PO. Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. Plos One 2010;5. doi:10.1371/journal.pone.0012671.

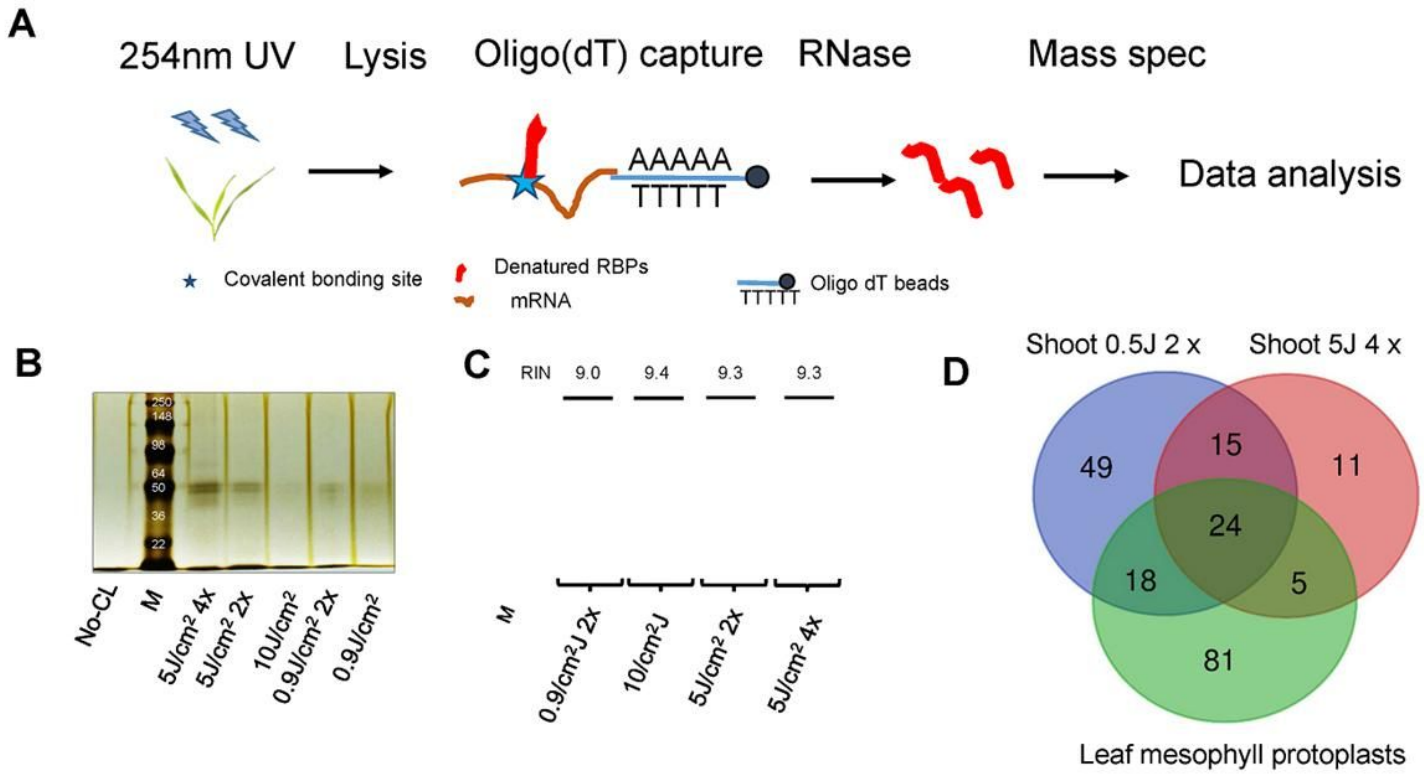
10. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell. 2012;149:1393–406. doi:10.1016/j.cell.2012.04.031.
11. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell. 2012;46:674–90. doi:10.1016/j.molcel.2012.05.021.
12. Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume AM, et al. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat Commun. 2015;6:10127. doi:10.1038/ncomms10127.
13. Sysoev VO, Fischer B, Frese CK, Gupta I, Krijgsveld J, Hentze MW, et al. Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. Nat Commun. 2016;7:12128. doi:10.1038/ncomms12128.
14. Marondedze C, Thomas L, Serrano NL, Lilley KS, Gehring C. The RNA-binding protein repertoire of Arabidopsis thaliana. Sci Rep. 2016;6:29766. doi:10.1038/srep29766.
15. Reichel M, Liao Y, Rettel M, Ragan C, Evers M, Alleaume AM, et al. In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. Plant Cell. 2016;28:2435–52. doi:10.1105/tpc.16.00562.
16. Kwon SC, Yi H, Eichelbaum K, Föhr S, Fischer B, You KT, et al. The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol. 2013;20:1122–30. doi:10.1038/nsmb.2638.
17. Chang CH, Curtis JD, Maggi LB, Faubert B, Villarino AV, O'Sullivan D, et al. Posttranscriptional control of T cell effector function by aerobic glycolysis. Cell. 2013;153:1239–51. doi:10.1016/j.cell.2013.05.016.
18. Zhang Z, Boonen K, Ferrari P, Schoofs L, Janssens E, van Noort V, et al. UV crosslinked mRNA-binding proteins captured from leaf mesophyll protoplasts. Plant Methods. 2016;12:42. doi:10.1186/s13007-016-0142-6.
19. Hong SY, Seo PJ, Cho SH, Park CM. Preparation of leaf mesophyll protoplasts for transient gene expression in Brachypodium distachyon. J Plant Biol. 2012;55:390–7. doi:10.1007/s12374-012-0159-y.
20. Lee JH, Ryu HS, Chung KS, Posé D, Kim S, Schmid M, et al. Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. Science. 2013;342:628–32. doi:10.1126/science.1241097.
21. Yoo SD, Cho YH, Sheen J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat Protoc. 2007;2:1565–72. doi:10.1038/nprot.2007.199.
22. Jankowsky E, Harris ME. Specificity and nonspecificity in RNA-protein interactions. Nature Reviews Mol Cell Biol. 2015;16:533–44. doi:10.1038/nrm4032.
23. Asencio C, Chatterjee A, Hentze MW. Silica-based solid-phase extraction of cross-linked nucleic acid-bound proteins. Life Sci Alliance. 2018;1:e201800088. doi:10.26508/lsa.201800088.

24. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. Nature Reviews Mol Cell Biol. 2018;19:327–41. doi:10.1038/nrm.2017.130.
25. Krauss P, Markstadter C, Riederer M. Attenuation of UV radiation by plant cuticles from woody species. Plant Cell Environ. 1997;20:1079–85. doi:10.1111/j.1365-3040.1997.tb00684.x.
26. Chatterjee J, Dionora J, Elmido-Mabilangan A, Wanchana S, Thakur V, Bandyopadhyay A, et al. The evolutionary basis of naturally diverse rice leaves anatomy. Plos One. 2016;11:e0164532. doi:10.1371/journal.pone.0164532.
27. Donnelly PM, Bonetta D, Tsukaya H, Dengler RE, Dengler NG. Cell cycling and cell enlargement in developing leaves of Arabidopsis. Dev Biol. 1999;215:407–19. doi:10.1006/dbio.1999.9443.
28. Littlejohn GR, Mansfield JC, Christmas JT, Witterick E, Fricker MD, Grant MR, et al. An update: improvements in imaging perfluorocarbon-mounted plant leaves with implications for studies of plant pathology, physiology, development and cell biology. Front Plant Sci. 2014;5:140. doi:10.3389/fpls.2014.00140.
29. Danon A, Rotari VI, Gordon A, Mailhac N, Gallois P. Ultraviolet-C overexposure induces programmed cell death in Arabidopsis, which is mediated by caspase-like activities and which can be suppressed by caspase inhibitors, p35 and Defender against Apoptotic Death. Biol Chem. 2004;279:779–87. doi:10.1074/jbc.M304468200.
30. Gao C, Xing D, Li L, Zhang L. Implication of reactive oxygen species and mitochondrial dysfunction in the early stages of plant programmed cell death induced by ultraviolet-C overexposure. Planta. 2008;227:755–67. doi:10.1007/s00425-007-0654-4.
31. Stapleton AE. Ultraviolet radiation and plants: burning questions. Plant Cell. 1992;4:1353–8. doi:10.1105/tpc.4.11.1353.
32. Nguyen CD, Mansfield RE, Leung W, Vaz PM, Loughlin FE, Grant RP, et al. Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA. J Mol Biol. 2011;407:273–83. doi:10.1016/j.jmb.2010.12.041.
33. Arribas-Hernández L, Bressendorff S, Hansen MH, Poulsen C, Erdmann S, Brodersen P. An m6A-YTH Module Controls Developmental Timing and Morphogenesis in Arabidopsis. Plant Cell. 2018;30:952–67. doi:10.1105/tpc.17.00833.
34. Wei LH, Song P, Wang Y, Lu Z, Tang Q, Yu Q, et al. The m6A Reader ECT2 Controls Trichome Morphology by Affecting mRNA Stability in Arabidopsis. Plant Cell. 2018;30:968–85. doi:10.1105/tpc.17.00934.
35. Aravind L, Iyer LM, Anantharaman V. The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. Genome Biol. 2003;4:R64. doi:10.1186/gb-2003-4-10-r64.
36. Goyal M, Banerjee C, Nag S, Bandyopadhyay U. The Alba protein family: Structure and function. Biochim Biophys Acta. 2016;1864:570–83. doi:10.1016/j.bbapap.2016.02.015.
37. Roldán M, Gómez-Mena C, Ruiz-García L, Salinas J, Martínez-Zapater JM. Sucrose availability on the aerial part of the plant promotes morphogenesis and flowering of Arabidopsis in the dark. Plant J.

1999;20:581–90. doi:10.1046/j.1365-313X.1999.00632.x.

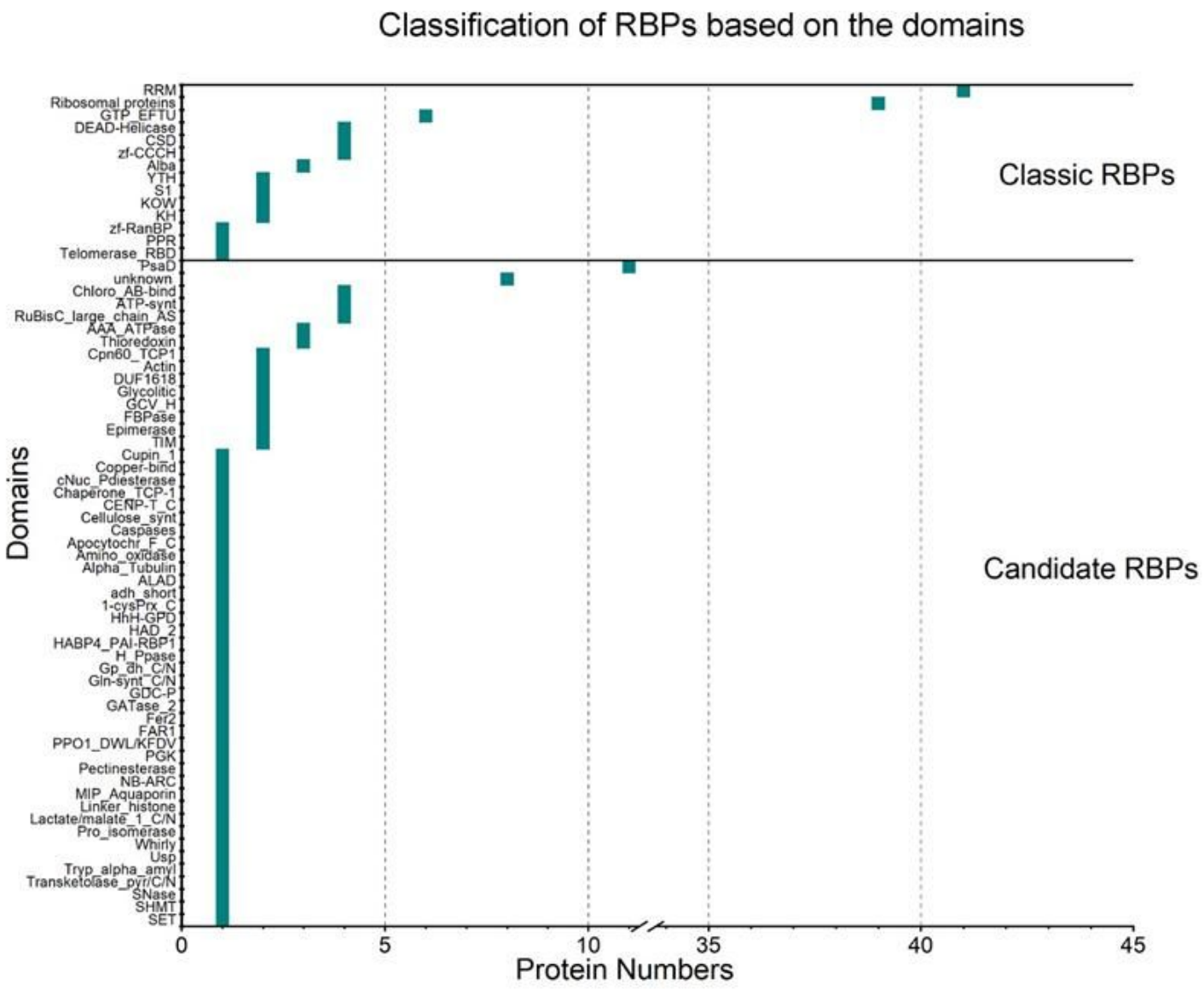
38. Zimmermann P, Zentgraf U. The correlation between oxidative stress and leaf senescence during plant development. Cell Mol Biol Lett. 2005;10:515–34.
39. Zhou X, Liao H, Chern M, Yin J, Chen Y, Wang J, et al. Loss of function of a rice TPR-domain RNA-binding protein confers broad-spectrum disease resistance. P Natl Acad Sci USA. 2018;115:3174–9. doi:10.1073/pnas.1705927115.
40. Siegele DA. Universal stress proteins in Escherichia coli. J of Bacteriol. 2005;187:6253–4. doi:10.1128/JB.187.18.6253-6254.2005.
41. Dominguez R, Holmes KC. Actin structure and function. Annu Rev of Biophys. 2011;40:169–86. doi:10.1146/annurev-biophys-042910-155359.
42. Klages-Mundt NL, Kumar A, Zhang Y, Kapoor P, Shen X. The Nature of Actin-Family Proteins in Chromatin-Modifying Complexes. Fron Genet. 2018;9:398. doi:10.3389/fgene.2018.00398.
43. Trendel J, Schwarzl T, Horos R, Prakash A, Bateman A, Hentze MW, et al. The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell. 2019;176:391–403.e19. doi:10.1016/j.cell.2018.11.004.
44. Paolo C, Warner D, Bensen RJ, Anstrom DC, Harrison J, Stoecker M, Abad M, et al. Bacterial RNA chaperones confer abiotic stress tolerance in plants and improved grain yield in maize under water-limited conditions. Plant physiol. 2008;147:446–55. doi.org/10.1104/pp.108.118828.
45. Ambrosone A, Batelli G, Nurcato R, Aurilia V, Punzo P, Bangarusamy DK, et al. The Arabidopsis RNA-binding protein AtRGGA regulates tolerance to salt and drought stress. Plant physiol. 2015;168:292–306. doi.org/10.1104/pp.114.255802.
46. Ambrosone A, Costa A, Leone A, Grillo S. Beyond transcription: RNA-binding proteins as emerging regulators of plant response to environmental constraints. Plant Sci. 2012;182:12–8. doi.org/10.1016/j.plantsci.2011.02.004.
47. Marondedze C, Thomas L, Gehring C, Lilley KS. Changes in the Arabidopsis RNA-binding proteome reveal novel stress response mechanisms. BMC Plant Biol. 2019;19:139. doi:10.1186/s12870-019-1750-x.

# Figures



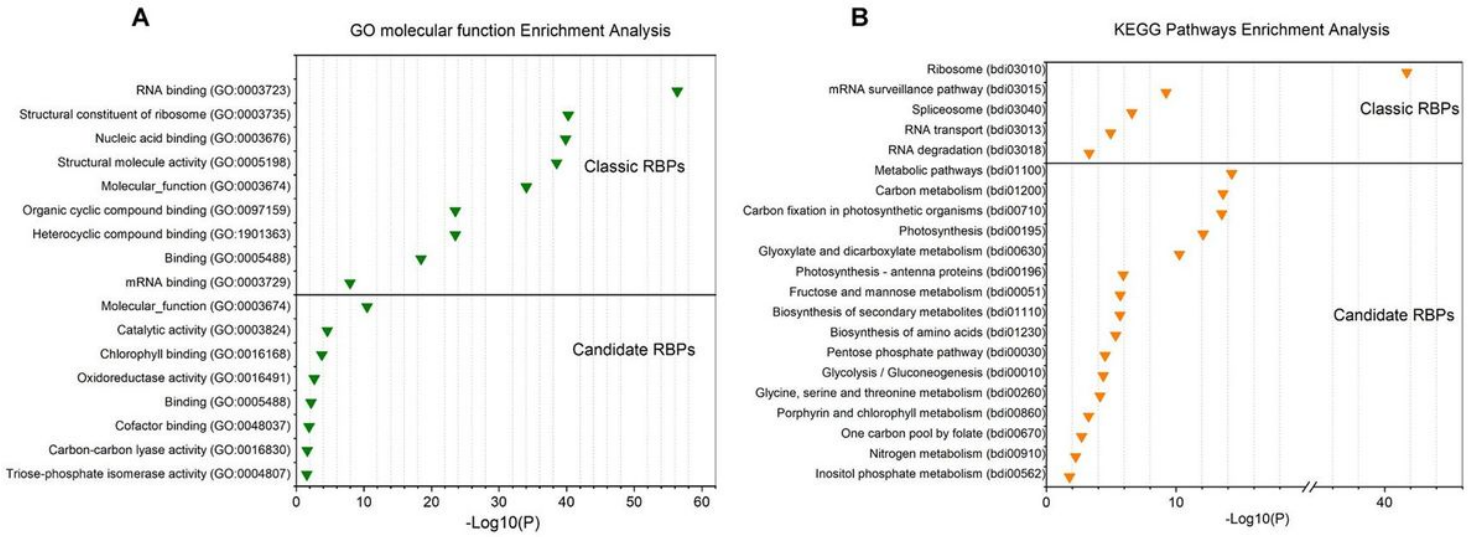

## Figure 1

RNA interactome capture in Brachypodium. (A): workflow of RNA interactome capture in Brachypodium seedlings. (B): silver staining results on different UV treated seedlings from 5J/cm2 UV four times exposure to 0.9J/cm2 UV exposure. (C): RNA quality and average RIN number for different UV dose treated seedlings (3 replicates) by Bio-analyzer. (D): comparison of detected proteins from different tissues (shoot and leaf mesophyll protoplasts) and UV dose treatment (5J/cm2 UV four times and 0.5J/cm2 UV two times).



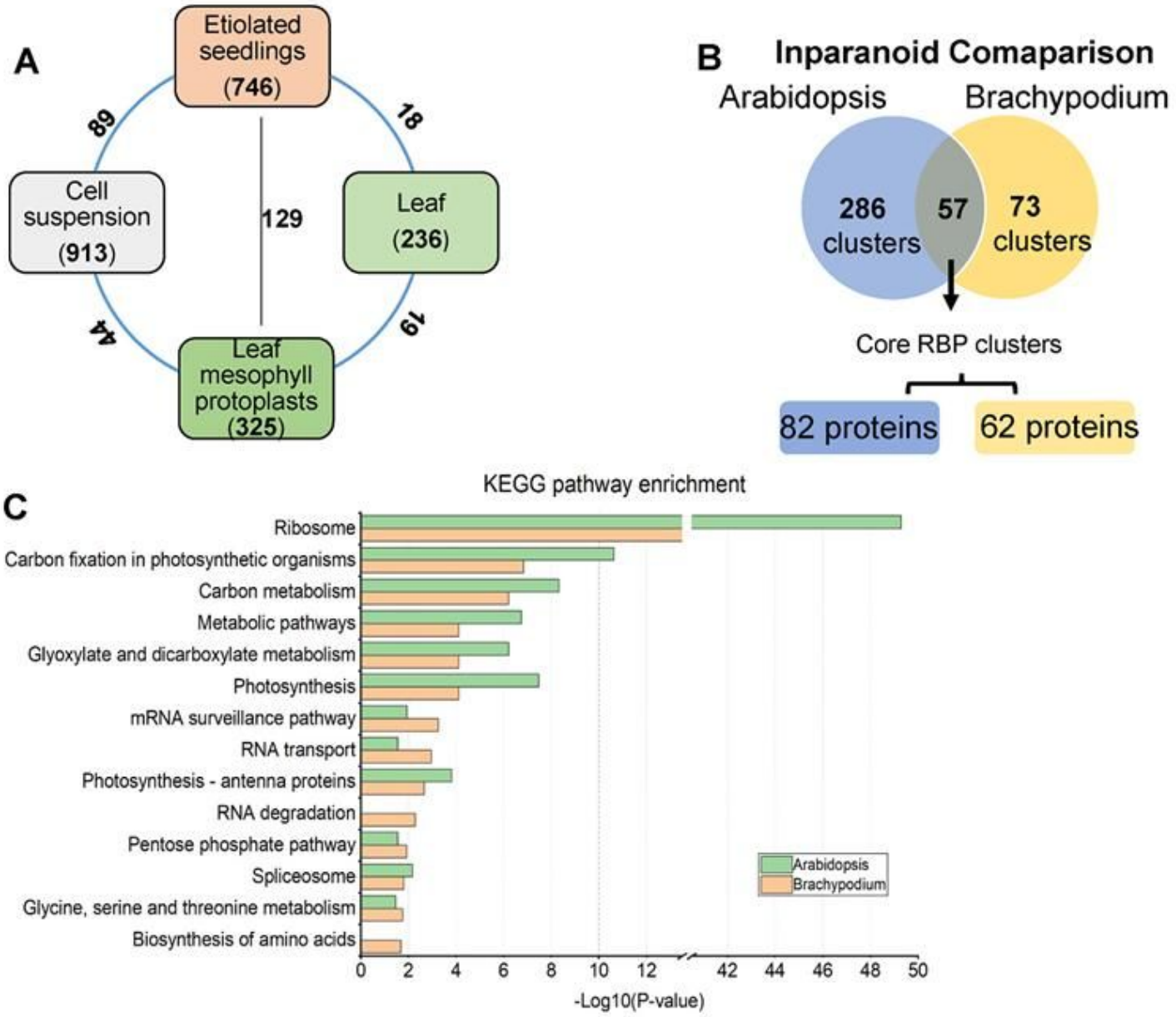

## Figure 2

Domain analysis on the classic RBPs and candidate RBPs.




## Figure 3

Molecular biological function GO enrichment analysis (A) and KEGG pathways enrichment analysis (B) on classic RBPs and candidate RBPs. Visualize data with -log10(P-value).




## Figure 4

Comparison of the RBPs between published data of Arabidopsis and Brachypodium identified in this research. (A): summary of the RBPs in Arabidopsis on published articles. (B): Inparanoid analysis in leaf and leaf mesophyll protoplast between Arabidopsis and Brachypodium. (C): KEGG pathway enrichment analysis on the core conserved RBPs in Arabidopsis and Brachypodium. Visualize data with -log10(P-value).

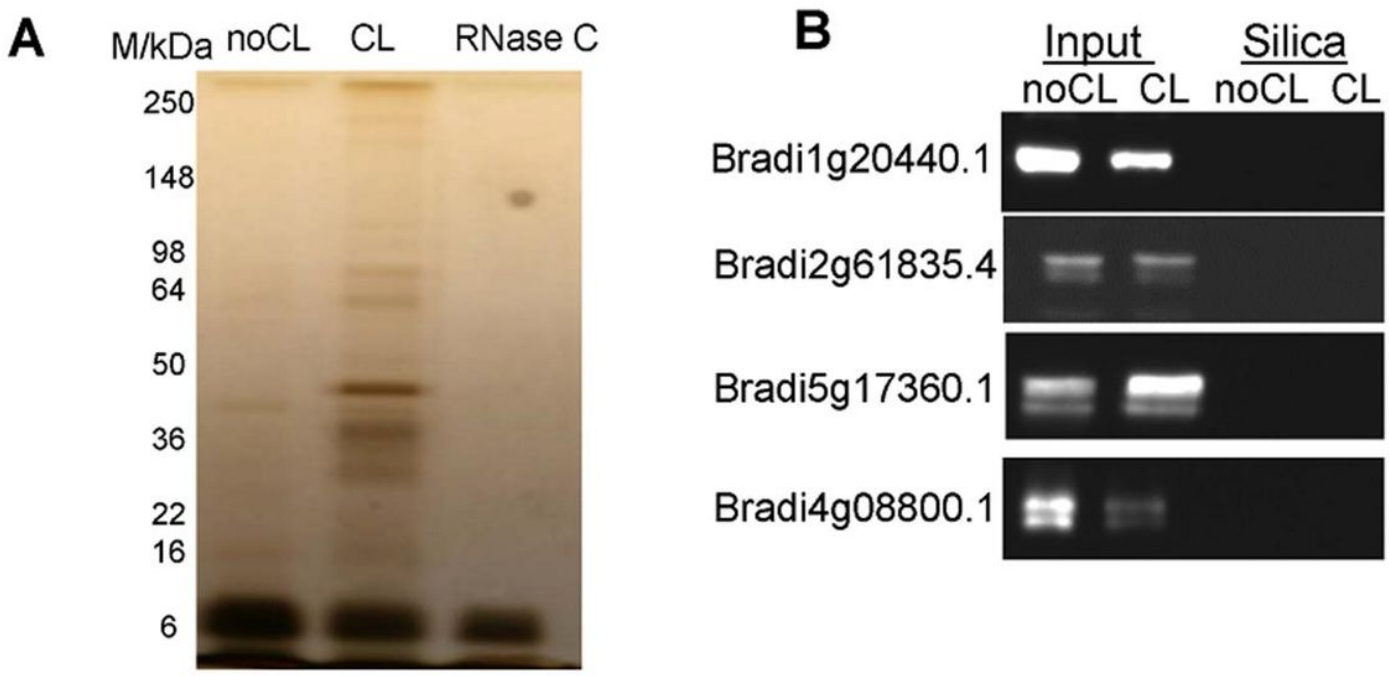



**Figure 5**

Validation of selected RBPs via silica columns. (A): the silver staining result on SDS page of leaf mesophyll protoplasts treated with the silica columns. noCL: non crosslinked samples, CL: crosslinked samples. RNase C: RNase cocktail control. (B): western blot of silica-based solid-phase extraction method on leaf mesophyll protoplast after GFP-tagged RBP transformation and incubation. Input samples: incubated leaf mesophyll protoplasts after GFP-tagged RBP transformation and incubation. Silica: GFP-tagged RBP transformed leaf mesophyll protoplasts treated with silica column extraction.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplymentaryfigures.pdf
- supplymentaryfigurelegend.docx
- SurpportinginformationTables.XLSX