ISOLATION, IDENTIFICATION AND CHARACTERIZATION OF RNA BINDING PROTEINS INVOLVED IN PROLAMINE MRNA LOCALIZATION IN RICE ENDOSPERM CELLS.

By

NAOKO CROFTS

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
Department of Molecular Plant Sciences

AUGUST 2009

To the faculty of Washington State University:

The members of the committee appointed to examine the dissertation of NAOKO CROFTS find it satisfactory and recommended that it be accepted.

Thomas W. Okita, Ph.D., Chair

Lee A. Hadwiger, Ph.D.

ChulHee Kang, Ph.D.

Andrew McCubbin, Ph.D.

B.W. Poovaiah, Ph.D.

ACKNOWLEDGEMENTS

Firstly I would like to thank my Graduate advisor, Dr Thomas W. Okita for welcoming me into his laboratory. I appreciate that he has provided me with the environment which I can try out many experimental techniques from the field of molecular biology, cell biology, and biochemistry. I am grateful for the patience he showed whilst I improved myself, especially my language skills and scientific knowledge, at my own pace. I would also like to thank all current and former members of the Okita laboratory with whom it has been a pleasure to working with, learning from, and just to be friends. I am very grateful for the help and assistance of all the IBC support staff including Sue, Craig and Julia for taking great care of my plants. I thank my Committee members, Dr. Lee Hadwiger, Dr. ChulHee Kang, Dr. Andy McCubbin and Dr. Joe Poovaiah for their many helpful comments and suggestions.

I acknowledge financial support from Washington State University in the form of a Research Assistantship, and awards from the Institute of Biological Chemistry in the form of Loyal H. and Helen Davis Fellowship and the John and Maggie McDougall Memorial Fund. I would also like to acknowledge financial support in the laboratory of Thomas W. Okita through grants from the National Science Foundation.

I would like to mention my undergraduate supervisor, Dr. Masahiro Ogawa who showed me the world of rice starchy endosperm by electron microscopy, and Peter and Catalina Salamone, whose influence lead me to study in Okita laboratory. I am very grateful to my parents in Japan, Toshiko Kawai, and in England, Graham and Sue Crofts for their support. Lastly, but foremost, I am very thankful to my husband, Andrew, for his influence and continual support and for my children, Kenji and Mikio, for the love and joy that they give me.

ISOLATION, IDENTIFICATION AND CHARACTERIZATION OF RNA BINDING PROTEINS INVOLVED IN PROLAMINE MRNA LOCALIZATION IN RICE ENDOSPERM CELLS.

Abstract

By Naoko Crofts, Ph.D.
Washington State University

August 2009

Chair: Thomas W. Okita

RNA localization is an evolutionarily conserved mechanism for targeting gene products to specific subcellular locations. The best-studied example of RNA localization within the plant kingdom is the sorting of the mRNAs encoding the rice seed storage proteins, prolamine and glutelin, to distinct endoplasmic reticulum (ER) subdomains. In developing rice seed endosperm, prolamine mRNAs are highly enriched on ER membranes which bound spherical protein bodies (PB-ER) formed by the controlled aggregation of nascent prolamine polypeptides. In contrast, glutelin mRNAs are enriched on adjacent cisternal ER membranes and, following translation, the encoded glutelin precursors are exported to the protein storage vacuole. RNA localization signals, known as zipcodes, have been identified in both prolamine and glutelin mRNAs and are responsible for their distinct localization patterns. Since it is known that prolamine mRNAs are transported via the actin cytoskeleton in a zipcode-dependent manner, *trans*-acting RNA binding proteins must be required to localize these mRNAs to the PB-ER. Therefore, RNA binding

proteins (RBPs) were isolated from cytoskeleton-enriched seed extracts using a prolamine zipcode affinity column, resulting in the identification of eighteen proteins by mass spectrometry. Affinity purified antibodies were generated against fifteen of these, including three chloroplast RBPs and seven members of the heterogeneous ribonucleoprotein (hnRNP) family which have known roles in RNA targeting. RBP-A, RBP-D and RBP-I, all of which are hnRNPs, show specificity for prolamine zipcode RNA compared to control RNA. The expression profiles of RBP-A and RBP-D show that both proteins are highly expressed only during the most rapid phase of storage protein accumulation. Microscopy data for RBP-A and RBP-D is also consistent with a role in RNA localization. Both are present on ER membranes and in particles which are closely associated with cytoskeletal elements. The development of a novel approach based on RNA immunoprecipitation coupled with microarray analysis allowed the identification of RNA species interacting with RBP-A. These included both prolamine and glutelin, in addition to a number of others. This data therefore supports a role for RBP-A as an RNA binding protein involved in the localization of storage protein RNAs to ER membranes.

TABLE OF CONTENTS

CHAPTER THREE

Development and Use of a Method for the Analysis of RNA Species Interacting with a Rice Cytoskeleton-enriched Prolamine Zipcode RNA Binding Protein

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF FIGURES

Page

DEDICATION

To my mother, Toshiko Kawai, my brother, Ryota Kawai and to the loving memory of my father, Fumio Kawai. To all those who always supported my decisions and encouraged me to explore the world, both scientifically and geographically.

To my family-in-law, Sue and Graham Crofts, Mark Crofts, Luke, Jo, Charlie and Meg Brown who always welcome the sharing of our joys.

To my beautiful sons, Kenji William Crofts and Mikio James Crofts, whose kisses and hugs allowed mummy to accomplish many things.

To my loving husband, Andrew James Crofts, who is always here for me.

CHAPTER ONE

INTRODUCTION

Overview

The endosperm cells of monocots accumulate seed storage proteins as a reserve of carbon, nitrogen, sulfur and energy for subsequent use by germinating seedling. Since cereals are a staple food for much of the world's population, storage proteins represent a vital source of dietary protein. Understanding the mechanisms of storage protein biosynthesis is of fundamental importance since it will allow the manipulation of protein content (both quantity and quality) and the potential use of cereals seeds as phytofactories for the production of high-value pharmaceutically active compounds.

Cereal plants usually store alcohol-soluble prolamin, while dicots accumulate saline-soluble 7S and 11S globulin proteins. Rice endosperm cells are unique in that they synthesize and accumulate high levels of two major classes of seed storage proteins, prolamines and glutelins (11S globulin-like), and sort them to separate subcellular compartments within the endomembrane system. Current evidence indicates that this protein sorting is dependent upon the segregation of prolamine and glutelin mRNAs onto distinct endoplasmic reticulum (ER) subdomains. Prolamine mRNAs are found almost exclusively on the ER membrane (protein body ER; PB-ER), which bounds spherical protein bodies formed by the controlled assembly of newly synthesized prolamine polypeptides. In contrast, glutelin mRNAs are enriched on adjacent cisternal ER (cis-ER) membranes and, following translation, glutelin precursors are transported via the Golgi to the protein storage vacuole where they are deposited and processed. RNA

localization signals, also known as zipcodes, have previously been identified in both prolamine and glutelin RNAs and are responsible for their distinct localizations.

mRNA localization is an important and evolutionarily conserved mechanism for targeting gene products to specific subcellular locations. Although much is known about RNA localization in mammalian systems, only a few examples are known in the plant kingdom. Therefore, understanding the role of RNA localization in storage protein biosynthesis is important. In this chapter, the work done in other systems as well as our current work with rice is reviewed.

The "RNA localization pathway"

The process of RNA localization is well-studied in eukaryotes where it allows gene products to be distributed in a temporally and spatially coordinated manner (reviewed by Van de Bor and Davis, 2004; Fedoroff, 2002; St Johnston, 2005; Crofts et al., 2004; Crofts et al., 2005; Czaplinski and Singer, 2006; Martin and Ephrussi, 2009). Although other mechanisms of delivering RNA to specific sub-cellular locations are known (including site-specific transcription, protection from degradation and diffusion-specific anchoring), this review will focus on the mechanism of active RNA localization which is dependent upon motor-driven movement along cytoskeletal elements (St Johnston, 2005). The most fundamental components required for such an RNA localization mechanism include the localized mRNAs themselves, which contain cis-acting targeting signals (or zipcodes) and the *trans*-acting RNA binding proteins (RBPs) which recognize these sequences and assemble to form ribonucleoprotein (RNP) complexes. Following nuclear export, the assembly and re-modeling of multiple ribosome-containing RNP complexes leads to the formation of a particle or granule that is frequently associated with ER or endosomal sorting complexes (Aronov et al., 2007; Paquin and Chartrand, 2008). This complex of protein

and RNA is then actively transported to its destination along the microtubule and/or actin cytoskeleton (Krichevsky and Kosik, 2001). Targeted RNAs are linked to cytoskeletal elements by adaptor proteins which interact with both RBPs and motor proteins such as myosin or dynein (reviewed by Kloc and Etkin, 2005; King et al., 2005; Lopez de Heredia and Jansen, 2004; Paquin and Chartrand, 2008).

During their transport, mRNAs are frequently present in translation initiation complexes but are translationally repressed (Nakamura et al., 2004). As the complex reaches its destination, a change in the phosphorylation status of a repressor protein is often associated with the onset of translation (Hüttelmaier et al., 2005; Paquin and Chartrand., 2008). This means that protein translation is spatially restricted and can occur immediately following arrival at the final destination. The overall picture of mRNA localization, therefore, is that it is a very dynamic and finely balanced mechanism in which the nature of multiple protein-RNA and protein-protein interactions determines the destination of the targeted RNA.

Examples of RNA localization

There are many known examples of RNA localization, with the best-studied examples being in *Drosophila* and *Xenopus* (reviewed by St Johnston, 2005). Studies in mammalian cells and in yeast are also highly advanced, but have focused on a smaller number of localized RNAs, for example ASH1 in budding yeast and β-actin in human fibroblasts. In *Drosophila*, mRNAs coding for transcription factors are localized to specific regions of the developing oocyte. For example, *bicoid* localizes to the anterior, whilst *gurken* localizes to the anterior-dorsal corner while *oskar* assumes a posterior localization. The spatial regulation of these key developmental signals is thus able to establish the anterior-posterior and dorsal-ventral axes. A number of other

mRNAs are also transported via microtubules through interaction with the motor proteins dynein or kinesin in order to establish cell polarity at different developmental stages (reviewed by Johnstone and Lasko, 2001). High-throughput and high-resolution *in situ* hybridization analysis performed on over 3000 *Drosophila* transcripts (one quarter of all transcripts encoded by the genome), revealed that 71 % are found to have a non-uniform distribution in early embryos (Lecuyer et al., 2007).

In *Xenopus* oocytes there are numerous mRNAs which are known to be localized to either the animal or vegetal side, again to establish correct polarity for further development. For example, maternal Vg1 mRNAs, which encode transforming growth factor beta (TGF-β), are localized at the wedge-shaped region of ER-enriched stage II oocytes (Melton, 1987) and the vegetal pole during stage III of oogenesis (King et al., 2005). Vg1 mRNA contains localization elements in its 3'UTR which are recognized and bound by a number of different RBPs including the zipcode binding protein VERA (Deshler et al., 1998), Prrp (proline rich RNA binding protein; Zhao et al., 2001), VgRBP60 (Cote et al., 1999) and 40LoVe (Czaplinski et al., 2005; Kroll et al., 2009). This complex, whose composition varies during its localization, travels along microtubules using the motor protein kinesin II (Betley et al., 2004). Prrp can also interact with profilin, a promoter of actin polymerization, to allow the actin-dependent anchoring of Vg1 mRNA (Zhao et al., 2001). Following anchoring, translation is initiated by the RNA binding protein VgRBP71 which stimulates cleavage of Vg1 mRNA at a polyadenylation signal and removes a *cis*-element that represses translation (Kolev and Huber, 2003).

RNA binding proteins involved in RNA localization

RNA binding proteins (RBPs) frequently contain one or more domains which function in RNA binding (reviewed by Colegrove-Otero et al., 2005; Singh and Valcarcel, 2005; Lunde et al., 2007; Glisovic et al., 2008). These domains include the RNA recognition motif (RRM), RNA binding domain (RBD) motif, K-homology (KH) domain, Arg-Ser (RS) domain, the Arg-Gly-Gly (RGG) box, the double stranded RNA binding domain (dsRBD). RBPs include a number of different protein families including the multifunctional heterogeneous ribonucleoproteins (hnRNPs). These proteins are involved in RNA maturation in the nucleus, but several family members have been shown to shuttle between the nucleus and cytoplasm where they also function in RNA localization and translational control (Dreyfuss et al., 2002; Hazelrigg, 2004; Kress et al., 2004, Percipalle et al., 2009).

The RNA binding specificity of RBPs often does not depend entirely on the presence of a single domain, but depends instead on the number of domains present and the distance between them (Lunde et al., 2007). RNA binding specificity *in vivo* is often conferred by the simultaneous interaction of multiple distinct RNA binding proteins which may, in turn, also interact with one another or with accessory proteins such as motor protein adaptors (Colegrove-Otero et al., 2005). Frequent re-modeling of transport particles during their maturation and localization is likely to be necessary to allow recycling of nuclear components and recruitment of additional cytosolic factors required for tasks such as the lifting of translational repression.

It is frequently difficult to observe specific interactions between RBPs and target RNAs *in vitro* (Li et al., 1991; Tillich et al., 2009). This is most likely due to the lack of functional complexes, or the inability of these complexes to assemble correctly on an RNA *in vitro*. However, *in vivo*, it could be advantageous to have multiple low affinity interactions since this

would give more flexibility for complex formation and re-modeling. It is therefore important to consider the potential problems of *in vitro* methods for RBP isolation and characterization. The focus of this study is therefore directed towards understanding the RNA localization process *in vivo*. This will be accomplished by analyzing the localization and composition of *in vivo* assembled RNPs using antibodies generated against RBPs initially identified using an *in vitro* approach.

The role of mRNA localization in rice seed storage protein biogenesis.

Rice endosperm cells accumulate seed storage proteins as a source of nitrogen and carbon for growth of the post-germinative seedling. Unlike most other cereal crops, rice synthesizes large amounts of both major classes of storage proteins, prolamines and glutelins, and sorts them to distinct intracellular compartments within the endomembrane system. Mature rice seeds contain (by weight) 18-20 % alcohol-soluble prolamines, 60-80 % weak acid or alkali-soluble glutelins and 2-8 % salt-soluble globulins (Osborne, 1924; Takaiwa, 1998). Owing to the smaller molecular size of prolamines compared to glutelins, these two protein classes are accumulate at essentially equimolar amounts.

The sorting of storage proteins in rice endosperm cells is established prior to their translation. As shown in Figure 1, RNA localization plays a central role in determining the final site of storage protein accumulation by targeting mRNAs for synthesis on morphologically distinct subdomains of the ER (Crofts et al., 2004; Crofts et al., 2005). The ER subdomain to which prolamine mRNAs are specifically targeted is called the protein body ER (PB-ER) and is defined as the ER membrane which surrounds the spherical type I protein body (PB-I). It is within this membrane that prolamine polypeptides accumulate with the assistance of the ER

resident binding protein, BiP (Li et al., 1993b). Prolamines do not contain ER retention signals, KDEL or HDEL, but are instead likely to be retained in the ER lumen by aggregation (Saito et al., 2009).

In contrast to prolamine mRNAs, which are targeted to the PB-ER and accumulate within the same structure, the site of translation and final accumulation of glutelin differs. Glutelin mRNAs are enriched on cis-ER which forms sheet-like structures and, following translation glutelin precursors are exported via the Golgi to the protein storage vacuole (PSV or protein body type II, PB-II) where they are cleaved to acidic and basic subunits (Krishnan, 1986; Krishnan, 1992). The only known exception to this rule is globulin, since its mRNA is localized to the PB-ER (H. Washida and T. Okita personal communication), but the encoded protein accumulates in PB-II (Krishnan, 1992).

RNA localization influences the site of storage protein deposition in rice endosperm cells

The generation of transgenic rice has shown that mis-targeting of RNAs to alternative ER subdomains by the addition of RNA localization signals leads to the mis-localization of the encoded protein (Washida et al., 2009). For example, the expression of the prolamine-type maize storage protein, 10 kD δ-zein, in rice shows that its mRNA is localized to the PB-ER, with the encoded protein accumulating in PB-I. When cis-ER RNA localization signal(s) from the glutelin 3'end are fused to the 10 kD δ-zein, the hybrid zein RNAs are subsequently mis-localized to cis-ER with the resulting protein now targeted to the PSV. In contrast, the expression of sunflower seed albumin (SSA) in rice shows that SSA mRNA is targeted to the cis-ER with the encoded protein stored in the PSV. The addition of cis-localization elements from the 10 kD δ-zein to SSA transcript led to the partial mis-localization of the hybrid SSA RNA to both the

cis-ER and PB-ER with SSA proteins correspondingly found in both the PSV and PB-I. Collectively, this mis-localization data provides additional evidence that the initial site of RNA localization determines the final site of storage protein accumulation (Washida et al., 2009). It also shows that the basic mechanism of RNA localization mechanisms between rice and maize is highly conserved between rice, maize and sunflower.

The importance of faithful RNA localization to specific subdomains of the ER has also been demonstrated using rice endosperm mutants which accumulate abnormally large amounts of glutelin precursor. In *glup2* mutants, prolamine mRNAs are mis-targeted to cis-ER and, as predicted by the studies described in the preceding paragraph, the protein is correspondingly mis-localized to protein storage vacuoles (H. Washida, M. Ogawa, T. Kumamaru, and H. Satoh, unpublished data). In *glup4* and *glup6* mutants, glutelin mRNAs are mis-targeted to the PB-ER rather than cis-ER and glutelin proteins are found in large dilated membrane structures as well as in protein storage vacuoles. Genetic analysis of *glup4* showed mutations to be present in Rab5a genes (M. Sato, M. Ogawa, T. Kumamaru, and H. Satoh, unpublished data). Rab5a is a member of the small GTPase superfamily of molecular switch proteins and, in general, plays an important role in the initiation of endocytic membrane trafficking (Bucci et al., 1992). This then leads to the fusion of vesicles to the endosome (Gorvel et al., 1991), stimulating the movement of early endosomes via microtubules (Nielsen et al., 1999). Rab proteins are active when bound to GTP, but inactive following the hydrolysis to GDP. The rate of endocytic fusion can be directly correlated with the rate of GTP binding and hydrolysis (Stenmark et al, 1994). The fact that the *glup4* and g*lup6* mutants have very similar phenotypes suggests that the gene mutated in *glup6* may also be involved in membrane transport.

Prolamine RNA localization is dependent upon active transport

There are several lines of evidence which indicate that the transport of prolamine mRNA from the nucleus to the PB-ER is an active process that is dependent upon both zipcode localization signals and *trans*-acting RNA binding proteins. Firstly, tracking of prolamine mRNA movement in real-time by observation of GFP fluorescence in living rice endosperm cells shows that they move as particles and that the addition of cytoskeleton disrupting drugs prevents transport, demonstrating that movement is dependent upon the actin cytoskeleton (Hamada et al., 2003a). Secondly, *in situ* RT-PCR observation of transgenic rice seed expressing GFP fusions with both full length and a number of truncated prolamine genes revealed that prolamine mRNAs contain two distinct zipcode sequences. The first lies in the 5'CDS and the second within the 3'UTR; both are required for correct mRNA localization (Hamada et al., 2003b) in addition to the translation initiation codon (Choi et al., 2000). Interestingly these two zipcode are located at opposite ends of the prolamine mRNA structure as predicted by mfold (Matthews et al., 1999; Zuker, 2003). Finally, heterologous expression of prolamine zipcodes in transgenic rice endosperm cells showed that addition of PB-ER targeting zipcodes can redirect the localization of mRNAs from cis-ER to PB-ER (Washida et al., submitted). Collectively, this evidence strongly supports the hypothesis that *trans*-acting RNA binding proteins are responsible for actin-dependent targeting of prolamine mRNA to the PB-ER subdomain.

Identification of *trans*-acting RNA binding proteins involved in prolamine mRNA localization

Previous efforts to isolate mRNA binding proteins involved in the localization of prolamine RNA resulted in the identification of Rp120 (later renamed to *Os*Tudor-SN) as a

potential candidate. Rp120 was shown to bind to the prolamine 3'UTR, but not to its 5'CDS in an *in vitro* UV cross-linking assay using a cytoskeleton-PB enriched fraction (Sami-Subbu et al., 2001). Rp120 has four nucleic acid binding SN (*Staphylococcal* nuclease) domains, a tudor domain and a coiled-coil domain and is homologous to p100, a human transcriptional co-activator and a component of the RNA induced silencing complex (Hacker and Karin, 2002; Caudy et al., 2003). *Os*Tudor-SN exists as large RNase sensitive particles which co-sediment with polysomes and contain both prolamine and glutelin mRNAs (Wang et al., 2008). Immunocytochemistry shows that *Os*Tudor-SN is present as particles in the cytoplasm which are closely associated with both PB- and cis-ER. Observation of GFP-tagged *Os*Tudor-SN in live endosperm cells using a number of different cytoskeleton disrupting drugs demonstrated that movement of *Os*Tudor-SN is dependent on presence of microfilaments. The generation of transgenic rice with reduced expression levels of *Os*Tudor-SN by either RNAi or antisense approaches resulted in lower levels of both prolamine mRNA and protein levels and a reduction in the number of prolamine protein bodies (Wang et al., 2008). This data strongly suggests that *Os*Tudor-SN is involved in prolamine biogenesis and that it may play a role in prolamine mRNA transport, localization and anchoring.

Another rice protein which has prolamine mRNA binding activity is the peroxisomal multi-functional protein (MFP) which was originally purified using microtubule and ssDNA affinity columns (Chuong et al., 2002). MFP is a peroxisomal resident protein which is involved in fatty acid β-oxidation and has the ability to bind prolamine 3'UTR. However, the binding specificity observed *in vitro* was weak (Chuong et al., 2002).

The aim of this study is to isolate and identify RNA binding proteins which are required for the cytoskeleton-mediated transport of prolamine mRNA from the nucleus to the PB-ER. Identified proteins will then be further characterized to shed light on their functions *in vivo*.

Isolation, identification and characterization of trans-acting RNA binding proteins involved in prolamine mRNA transport

To understand the mechanism of prolamine mRNA transport, it is necessary to study prolamine zipcode RNA binding proteins. As described in chapter 2, we have isolated, and identified a large number of such proteins. Following identification, antibodies were generated for use in further analysis. We employed an RNA bait corresponding to the previously characterized 5'CDS prolamine zipcode sequence for the capture of RNA binding proteins (RBPs) from cytoskeleton-enriched fractions of developing rice seed. This work has led to the identification of eighteen distinct RNA binding proteins. Following cloning of cDNA into expression vector, histidine tagged proteins were over-expressed in *Escherichia coli*, purified using nickel immobilized metal affinity chromatography and fifteen antibodies were generated. Using these antibodies, the *in vitro* binding specificity of captured proteins was determined. Confocal immunofluorescence microscopy shows that two of the candidate prolamine mRNA binding proteins, termed RBP-A and RBP-D, both of which are hnRNP family members, co-localize with cytoskeletal elements in addition to being present in the nucleus. RBP-A co-localizes with microtubules and is associated with cortical ER membranes. RBP-D, in contrast, is present in small particles which are in close association with both prolamine protein bodies and the actin cytoskeleton. Protein expression patterns of both RBP-A and RBP-D during seed development strongly correlate with those of the major seed storage proteins.

Since both the localization and expression data described above for RBP-A are consistent with it having a role in storage protein RNA localization, the RNAs bound by RBP-A *in vivo* were further studied in chapter 3 in order to provide additional insight into its potential role in RNA transport. Immunoprecipitation using affinity purified anti-RBP-A antibody followed by RNA isolation and RT-PCR have shown that RBP-A binds to significant amounts of both prolamine and glutelin mRNA *in vivo* compared to controls. Further investigation of the RNA species bound by RBP-A by whole genome microarray analysis suggests that both prolamine and glutelin may be present in same RNP complex. In addition, mRNAs encoding globulin, two glycine rich RNA binding proteins and several plant defense proteins were also found to be present. Computational analysis of the mRNAs bound by RBP-A revealed several consensus sequences which lie within the established prolamine 3'UTR zipcode.

Increasing evidence, from both yeast (Schmid et al., 2006; Irion et al., 2006) and *Drosophila* (Irion and St Johnston, 2007) shows that RNAs are frequently transported as large organelle-derived membrane complexes (reviewed by Cohen, 2005). The data from the *glup4* mutants, where a mutation in Rab5a causes RNA mis-targeting, also supports this relationship between membrane and RNA transport. Since zipcodes for the localization of both prolamine and glutelin have now been identified (Hamada et al., 2003b; Washida et al., submitted), future work will focus on more closely defining the roles of the proteins involved in RNA transport. In particular, it should be possible identify proteins which are present in transport complexes with both prolamine and glutelin mRNAs, and others which directly determine the ER target membrane. The factors involved in repressing translation, and how their activity is regulated, by phosphorylation or otherwise, are also candidates for further study. Collectively, addressing

these important questions should provide further valuable insight into the process of RNA localization.

References

Betley, J. N., Heinrich, B., Vernos, I., Sardet, C., Prodon, F., and Deshler, J. O. (2004). Kinesin II mediates Vg1 mRNA transport in Xenopus oocytes. *Curr. Biol.* **14**, 219-224.

Bucci, C., Parton, R. G., Mather, I. H., Stunnenberg, H., Simons, K., Hoflack, B., and Zerial, M. (1992). The small GTPase rab5 functions as a regulatory factor in the early endocytic pathway. *Cell.* **70**, 715-728.

Caudy, A. A., Ketting, R. F., Hammond, S. M., Denli, A. M., Bathoorn, A. M., Tops, B. B., Silva, J. M., Myers, M. M., Hannon, G. J., and Plasterk, R. H. (2003). A micrococcal nuclease homologue in RNAi effector complexes. *Nature.* **425**, 411-414.

Choi, S.B., Wang, C., Muench, D. G., Ozawa, K., Franceschi, V. R., Wu, Y., and Okita, T. W. (2000). Messenger RNA targeting of rice seed storage proteins to specific ER subdomains. *Nature.* **407**,765-767.

Chuong, S. D., Mullen, R. T., and Muench, D. G. (2002). Identification of a rice RNA- and microtubule-binding protein as the multifunctional protein, a peroxisomal enzyme involved in the beta -oxidation of fatty acids. *J. Biol. Chem.* **277**, 2419-2429.

Cohen, R. S. (2005). The role of membranes and membrane trafficking in RNA localization. *Biol Cell.* **97**, 5-18.

Colegrove-Otero, L. J., Minshall, N., and Standart, N. (2005). RNA-binding proteins in early development. Crit. Rev. *Biochem. Mol. Biol.* **40**, 21-73.

Cote, C. A., Gautreau, D., Denegre, J. M., Kress, T. L., Terry, N. A., and Mowry, K. L. (1999). A Xenopus protein related to hnRNP I has a role in cytoplasmic RNA localization. Mol. *Cell* **4**, 431-437.

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

Crofts A. J. , Washida H., Okita T. W., Satoh M., Ogawa M., Kumamaru T., Satoh H. (2005). The role of mRNA and protein sorting in seed storage protein synthesis, transport, and deposition. *Biochem Cell Biol.* **83**, 728-737.

Czaplinski, K., Kocher, T., Schelder, M., Segref, A., Wilm, M., and Mattaj, I. W. (2005). Identification of 40LoVe, a Xenopus hnRNP D family protein involved in localizing a TGF-beta-related mRNA during oogenesis. *Dev. Cell* **8**, 505-515.

Czaplinski K., and Singer RH. (2006) Pathways for mRNA localization in the cytoplasm. *Trends Biochem Sci*. **12**, 687-693.

Deshler, J. O., Highett, M. I., Abramson, T., and Schnapp, B. J. (1998). A highly conserved RNA-binding protein for cytoplasmic mRNA localization in vertebrates. *Curr. Biol.* **8**, 489-496.

Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* **3**, 195-205.

Fedoroff, N. V. (2002). RNA-binding proteins in plants: the tip of an iceberg? *Curr. Opin. Plant Biol.* **5**, 452-459.

Glisovic, T., Bachorik, J. L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* **582**, 1977-1986.

Gorvel, J. P., Chavrier, P., Zerial, M., and Gruenberg J. (1991). rab5 controls early endosome fusion *in vitro*. *Cell.* **64**, 915-925.

Hacker, H., and Karin, M. (2002). Is NF-kappaB2/p100 a direct activator of programmed cell death? *Cancer Cell.* **2**, 431-433.

Hamada, S., Ishiyama, K., Choi, S. B., Wang, C., Singh, S., Kawai, N., Franceschi, V. R., and Okita, T. W. (2003a). The transport of prolamine RNAs to prolamine protein bodies in living rice endosperm cells. *Plant Cell.* **15**, 2253-2264.

Hamada, S., Ishiyama, K., Sakulsingharoj, C., Choi, S. B., Wu, Y., Wang, C., Singh, S., Kawai, N., Messing, J., and Okita, T. W. (2003b). Dual regulated RNA transport pathways to the cortical region in developing rice endosperm. *Plant Cell.* **15**, 2265-2272.

Hazelrigg, T. (2004). Lost in translation gets an oskar. Dev. Cell ***6***, 611-613.

Irion, U. and St Johnston, D. (2007). bicoid RNA localization requires specific binding of an endosomal sorting complex. *Nature* **445**, 497-499.

Johnstone, O., and Lasko, P. (2001). Translational regulation and RNA localization in Drosophila oocytes and embryos. *Annu. Rev. Genet.* **35**, 365-406.

King, M. L., Messitt, T. J., and Mowry, K. L. (2005). Putting RNAs in the right place at the right time: RNA localization in the frog oocyte. *Biol. Cell* **97**, 19-33.

Kloc, M., and Etkin, L. D. (2005). RNA localization mechanisms in oocytes. *J. Cell Sci.* ***118***, 269-282.

Kroll, T. T., Swenson, L. B., Hartland, E. I., Snedden, D.D., Goodson, H.V., and Huber, P.W. (2009). Interactions of 40LoVe within the ribonucleoprotein complex that forms on the localization element of Xenopus Vg1 mRNA. *Mech Dev.* [Epub ahead of print]

Kolev, N. G., and Huber, P. W. (2003). VgRBP71 stimulates cleavage at a polyadenylation signal in Vg1 mRNA, resulting in the removal of a cis-acting element that represses translation. *Mol. Cell* **11**, 745-755.

Kress, T. L., Yoon, Y. J., and Mowry, K. L. (2004). Nuclear RNP complex assembly initiates cytoplasmic RNA localization. *J. Cell Biol.* **165**, 203-211.

Krichevsky, A. M., and Kosik, K. S. (2001). Neuronal RNA granules: a link between RNA localization and stimulation-dependent translation. *Neuron* **32**, 683-696.

Krishnan, H. B., Franceschi, V.R., and Okita, T.W. (1986). Immunocytochemical studies on the role of the Golgi complex in protein body formation in rice seed. *Planta* **169**, 471-480.

Krishnan, H. B., White, J.A., and Pueppke, S.G. (1992). Characterization and localization of rice (Oryza sativa L.) seed globulins. *Plant Science* **81**, 1-11.

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T. R., Tomancak, P., and Krause, H. M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell.* **131**, 174-187.

Li, X., Franceschi, V. R., and Okita, T. W. (1993a). Segregation of storage protein mRNAs on the rough endoplasmic reticulum membranes of rice endosperm cells. *Cell* **72**, 869-879.

Li, X., Wu, Y., Zhang, D. Z., Gillikin, J. W., Boston, R. S., Franceschi, V. R., and Okita, T. W. (1993b). Rice prolamine protein body biogenesis: a BiP-mediated process. *Science* **262,** 1054-1056.

Lopez de Heredia, M., and Jansen, R. P. (2004). mRNA localization and the cytoskeleton. Curr. Opin. *Cell Biol.* **16**, 80-85.

Lunde, B. M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol.* **8**, 479-490.

Martin, K. C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. *Cell.* **136**, 719-730

Matthews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999). Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. J. *Mol. Biol.* **288**, 911-940.

Melton, D. A. (1987). Translocation of a localized maternal mRNA to the vegetal pole of Xenopus oocytes. *Nature* **328**, 80-82.

Nakamura, A., Sato, K., and Hanyu-Nakamura, K. (2004). Drosophila cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis. *Dev. Cell* **6**, 69-78.

Nielsen, E., Severin, F., Backer, J. M., Hyman, A. A., and Zerial M. (1999). Rab5 regulates motility of early endosomes on microtubules. *Nat Cell Biol.* **1**, 376-382.

Osborne, T. B. (1924). The vegetable proteins, 2nd edn).

Paquin, N., and Chartrand, P. (2008). Local regulation of mRNA translation: new insights from the bud. *Trends Cell Biol.* **18**, 105-111.

Percipalle, P., Raju, C. S., and Fukuda, N. (2009). Actin-associated hnRNP proteins as transacting factors in the control of mRNA transport and localization. *RNA Biol.* **6,** [Epub ahead of print]

Sami-Subbu, R., Choi, S. B., Wu, Y., Wang, C., and Okita, T. W. (2001). Identification of a cytoskeleton-associated 120 kDa RNA-binding protein in developing rice seeds. *Plant Mol. Biol.* **46**, 79-88.

Schmid, M., Jaedicke A., Du, T.G., and Jansen, R. P. (2006). Coordination of endoplasmic reticulum and mRNA localization to the yeast bud. *Curr. Biol.* **16**, 1538-1543.

Singh, R., and Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. *Nat. Struct. Mol. Biol.* **12**, 645-653.

St Johnston, D. (2005). Moving messages: the intracellular localization of mRNAs. *Nat. Rev. Mol. Cell Biol.* **6**, 363-375.

Takaiwa, F. (1998). Structure and expression of rice storage protein genes. (Tokyo, Japan, Springer-Verlag).

Van de Bor, V., and Davis, I. (2004). mRNA localisation gets more complex. *Curr. Opin. Cell Biol.* **16**, 300-307.

Wang, C., Washida. H., Crofts, A. J., Hamada, S., Katsube-Tanaka, T., Kim, D., Choi, S. B., Modi, M., Singh, S., and Okita, T. W. (2008). The cytoplasmic-localized, cytoskeletal-associated RNA binding protein OsTudor-SN: evidence for an essential role in storage protein RNA transport and localization. *Plant J.* **55**,443-454.

Washida, H., Sugino, A., Kaneko, S., Crofts, N., Sakulsingharoj, C., Kim, D., Choi, S. B., Hamada, S., Ogawa, M., Wang, C., Esen, A., Higgins, T. J., and Okita, T. W. (2009). Identification of cis-localization elements of the maize 10 kD delta-zein and their use in targeting RNAs to specific cortical ER subdomains. *Plant J.* (Accepted)

Washida, H., Kaneko, S., Crofts, N., Sugino, A., Wang, C., and Okita, T. W. (2009). Identification of *cis*-localization elements that target glutelin RNAs to a specific subdomain of the cortical endoplasmic reticulum in rice endosperm cells. *Plant Cell Physiol.* (Accepted)

Zhao, W. M., Jiang, C., Kroll, T. T., and Huber, P. W. (2001). A proline-rich protein binds to the localization element of Xenopus Vg1 mRNA and to ligands involved in actin polymerization. *EMBO J*. **20**, 2315-2325.

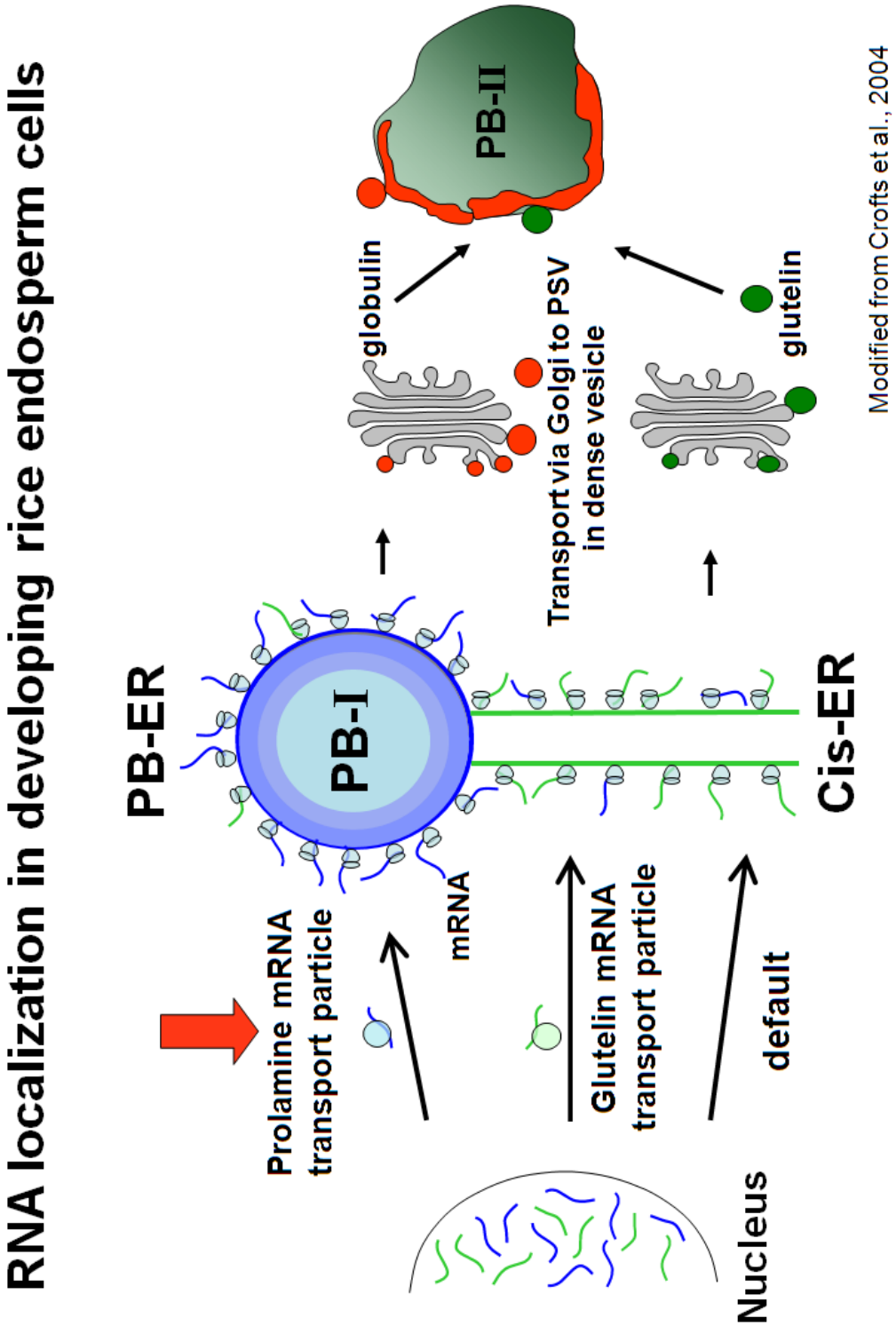
Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* ***31***, 3406-3415.

Figure legends

Figure 1. Storage protein biosynthesis in developing rice seed

This schematic diagram depicts RNA-dependent seed storage protein targeting in developing rice endosperm. After transcription in the nucleus, mRNAs encoding seed storage proteins are transported to the cortical ER via the actin cytoskeleton where they segregate onto either the protein body ER (PB-ER; prolamine and globulin) or the cisternal ER (glutelin). This asymmetric mRNA localization on the ER determines the final destination of the encoded proteins. Prolamine polypeptides are translocated into the ER lumen where they form spherical intracisternal inclusion granules known as protein body type I (PB-I). In contrast, globulin and glutelin polypeptides are transported via the Golgi to protein storage vacuole (PSV or PB-II). Since the RNA localization signals of prolamine have been determined, the next step is to identify RNA binding proteins likely to be present in prolamine RNA transport particles (large arrow).

Figure 1. Storage protein biosynthesis in developing rice seed



CHAPTER TWO

ISOLATION, IDENTIFICATION AND CHARACTERIZATION OF CYTOSKELETON-ASSOCIATED PROLAMINE MRNA BINDING PROTEINS FROM DEVELOPING RICE SEEDS

Abstract

The messenger RNA of the rice seed storage protein prolamine is known to be targeted to the endoplasmic reticulum (ER) membranes surrounding prolamine protein bodies via a mechanism which is dependent upon both RNA localization signals and the actin cytoskeleton. In this study we have used an RNA bait corresponding to the previously characterized 5'CDS prolamine zipcode sequence for the capture of RNA binding proteins (RBPs) from cytoskeleton-enriched fractions of developing rice seed. In comparison to a control RNA, the zipcode RNA bait sequence led to the capture of a much larger number of proteins, eighteen of which have been identified by tandem mass spectrometry. cDNAs corresponding to most of these proteins were then cloned and expressed in order to generate antibodies. Western blots demonstrate that almost all of the candidate proteins analyzed to date show good to excellent specificity for binding to zipcode over non-zipcode RNA bait. Protein expression analysis of two promising candidates, RBP-A and RBP-D, during seed development shows that their protein levels are elevated during the most rapid phase of storage protein accumulation. Confocal immunofluorescence microscopy shows that RBP-A and RBP-D, which are both hnRNP family members, co-localize with cytoskeletal elements in addition to being present in the nucleus. RBP-A is associated with microtubules and is also present on cortical ER membranes. RBP-D, in contrast, is present in small particles which are found in close association with both prolamine

protein bodies and the actin cytoskeleton. Collectively, these results indicate that employing a combination of *in vitro* binding, *in vivo* localization and expression studies is a valid strategy for the identification of putative prolamine mRNA binding proteins, which play a role in controlling expression of prolamine RNA in the cytoplasm.

Introduction

In recent years it has become clear that mRNA localization is a widespread and efficient means of targeting gene products to specific intracellular regions in a timely manner (reviewed by Martin et al., 2009; Du et al., 2007; St Johnston, 2005; Kloc et al., 2002; Jansen, 2001 and for plants by Okita and Choi, 2002; Fedoroff, 2002). A number of localized RNAs from both animal and plant species have now been described, and in some model systems nearly 20 % of all mRNA species are localized (Lecuyer et al., 2007; Dubowy and Macdonald, 1998). In addition to its primary function of generating localized high concentrations of protein in somatic cells, RNA targeting also plays a central role in cell fate determination during embryonic development and in the generation of cell polarity (Martin et al., 2009; Ma et al., 2008; Du et al., 2007; St Johnston, 2005).

The general mechanism of RNA targeting ('the RNA localization pathway') involves the interaction of *trans*-acting RNA binding factors with *cis*-acting localization elements present in the targeted RNA, followed by transport via the cytoskeleton and its associated motor proteins (reviewed by Wilhelm and Vale, 1993; Bullock, 2007). These *cis*-acting localization elements are typically found in the 3' untranslated region (UTR) of targeted gene products (Bullock and Ish-Horowicz, 2001), although they are also present in the coding sequence (CDS) and 5'UTR (Pyhtila et al., 2008). It is thought that the interaction of binding proteins with the targeted RNA influence its conformation and thus trigger the binding of additional proteins and the assembly of a large ribonucleoprotein particle or granule (Gonsalvez et al., 2004; Anderson and Kedersha, 2009). This RNA transport particle is likely to contain many different localized RNAs in addition to factors which mediate their correct targeting and anchoring to the destination site as well as to maintain transported RNAs in a translationally silent state (Cáceres and Nilson, 2009;

Du et al., 2008). The RNA transport particle may also contain components for translation including ribosomes, elongation factors, and aminoacyl tRNA synthetases along with vesicles or organellar membranes derived from the ER (Krichevsky and Kosik, 2001; Cohen, 2005; Gerst 2008). The co-transport of many, if not all, components required for protein synthesis enables the RNA to be translated immediately after upon reaching its destination site. RNA localization therefore allows for the rapid, site-specific, expression of protein; a process which is also very efficient in terms of energy since the transport of a single mRNA can give rise to multiple protein molecules while avoiding unnecessary interaction (Martin and Ephrussi., 2009). Recently, progress has been made regarding the mechanism by which translational arrest is lifted upon completion of localization. Studies in both yeast and mammalian systems have identified phosphorylation to be important in regulating this key step (Landers et al., 2009; Hüttelmaier et al., 2005).

In addition to *trans*-acting RNA binding proteins, the cytoskeleton has also been shown to be an essential component for the targeting of mRNAs in many different species, providing a framework for the delivery of RNAs to their final destinations (Jansen, 1999; Stebbings, 2001; López de Heredia and Jansen, 2004). For example, mating type switching in yeast is restricted to the mother cell by the myosin-dependent transport of *ASH1* mRNA into the daughter cell along the actin cytoskeleton (Bertrand et al., 1998; reviewed by Muller et al., 2007; Paquin and Chartrand, 2008). In general, however, RNA transport is more commonly accomplished using microtubules which can form longer structures and are thus better suited for long distance RNA localization in polarized, differentiated, vertebrate cells and oocytes (Stebbings, 2001; Kloc et al., 2002; Messitt et al., 2008).

In plants, RNA localization is known to be important in the plasmodesmatal trafficking of viral movement particles (Boyko et al., 2000), in the control of root hair elongation via the mRNA localization of actin regulators (Baluska et al., 2000), differential localization of different classes of expansin mRNAs to either the apical or basipetal end of xylem cells (Im, et al., 2000) and in targeting to organelles including chloroplasts and mitochondria (Okita and Choi, 2002). RNA localization has also been demonstrated for mRNAs encoding maize storage proteins (Washida et al., 2004) and for calreticulin mRNA targeting to callus protein bodies in maize (Samaj et al., 2008). Most recently, a cytoskeleton-associated RNP was identified that stores and transport mRNAs during pollen tube growth in tobacco male gametophyte (Honys et al., 2009).

The best-studied example of RNA localization in plants is the targeting of RNAs encoding storage proteins to different subdomains of the ER in developing endosperm cells of rice (Li et al., 1993; reviewed by Crofts et al., 2004, 2005). Rice is unique amongst the cereals in that developing seeds synthesize large amounts of both classes of seed storage proteins, prolamines and glutelins, and store them in different compartments of the endomembrane system (Yamagata and Tanaka, 1986; Li et al., 1993; Choi et al., 2000). Both prolamine and glutelin messenger RNAs are targeted via an RNA-based mechanism to distinct subdomains of the endoplasmic reticulum (ER; Choi et al., 2000; reviewed by Crofts et al., 2004, 2005). Prolamine mRNAs are transported to the protein body ER (PB-ER) within large particles, which are then translated to eventually form an intracisternal inclusion. In contrast, glutelin mRNAs are transported to the cisternal ER before being translated and then transported as precursors via the Golgi to protein storage vacuoles where they are cleaved into acidic and basic subunits (Krishnan et al., 1986; Krishnan et al., 1992).

Previously, our laboratory has shown that there are two distinct RNA localization signals in prolamine mRNA, both of which are required for faithful targeting to the PB-ER in addition to the presence of start codon (Choi et al., 2000; Hamada et al., 2003a). The first prolamine localization signal, or zipcode, resides in the 5'CDS and the other lies in the 3'UTR. It has also been demonstrated that this RNA targeting process is dependent upon the actin cytoskeleton (Hamada et al., 2003b). However, an RNA containing multiple localization signals can be redirected to the cisternal ER by the addition of glutelin 3' end sequences indicating presence of dominant RNA localization signals in glutelin mRNA (Hamada et al., 2003a). A recent study demonstrated that glutelin mRNA indeed contains three RNA localization signals, which are located in both the 5' and 3' end of the CDS as well as in 3'UTR (Washida et al., submitted). Based on the analysis of hybrid RNAs containing prolamine and/or glutelin sequences and of rice mutants defective in RNA targeting, there are at least two regulated RNA transport pathways and a default pathway operating in rice endosperm (Hamada et al., 2003a; Crofts et al., 2004, 2005). This raises the question of whether any of the components involved in these pathways are conserved, whilst others, perhaps those which determine the target site, differ.

Much of the prior research on the sorting of the RNAs encoding rice storage proteins has focused on the identification of targeting signals rather than the *trans*-acting RNA binding proteins which recognize them. In this study we have used the 5'CDS prolamine mRNA targeting signal (zipcode) as bait for the capture of RNA binding proteins from cytoskeleton-enriched developing rice seed extracts. Using this broad experimental approach, we have been able to isolate a large number of RNA binding proteins. The identity of 18 of these proteins was determined by tandem mass spectroscopy and the majority were found to contain a pair of RNA

recognition motifs (RRMs) with several sharing significant homology with members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family.

The generation of antibodies against almost all of the candidate prolamine mRNA binding proteins has allowed us to demonstrate that many of the proteins identified show strong and specific binding for the prolamine zipcode sequence. In addition, immunofluorescence localization experiments and protein expression analysis by western blotting have provided additional evidence supporting a role for at least two of these proteins in the localization of prolamine mRNA.

Materials and Methods

Rice growth conditions

Oryza sativa was grown under controlled environmental conditions in the growth chambers at Washington State University. *Oryza sativa*, cultivar Kitaake was used for all experiments except for microarray analysis and plants were grown with cycles of light periods at 26 ^{o}C for 12 hours and dark periods at 22 ^{o}C for 12 hours with daily watering until seeds reached maturity. *Oryza sativa*, cultivar Taichung 65 was grown under the same conditions as Kitaake for vegetative growth. In order to induce flowering, plants were then moved to a shorter light period (10 hours) for two weeks, and then to cycles of light periods at 27 ^{o}C for 12 hours and dark periods at 23 ^{o}C for 12 hours with daily watering.

Preparation of cytoskeleton-enriched fractions from developing rice seeds

Developing wild-type rice seeds (*Oryza sativa,* cultivar *Kitaake*; grown as described above) were harvested 10-15 days after flowering. In a typical extraction, 20 g of seeds were

frozen in liquid nitrogen and ground to a fine powder prior to extraction with 100 mL cytoskeleton stabilizing buffer (CSB; 5 mM Hepes-KOH, pH 7.5, 10 mM $Mg(OAc)_2$, 2 mM EGTA, 0.5 % PTE, 10 % glycerol, 200 mM sucrose) supplemented with 0.1 mM PMSF, 0.5 mM DTT, 1 μM pepstatin, 100 μM leupeptin. The slurry was then filtered through cheesecloth and the filtrate was centrifuged at 100 *g* for 1 min. The resulting supernatant was then centrifuged at 3000 *g* for 10 min and the pellet was extracted in 20 mL cytoskeleton destabilizing buffer (CDB; 200 mM Tris-HCl, pH 8.5, 450 mM KOAc, 25 mM $Mg(OAc)_2$, 2 % PTE, 10 % glycerol) supplemented with 0.1 mM PMSF, 0.5 mM DTT, 1 μM pepstatin, 100 μM leupeptin. This material was then placed on a 60 % (w/v) sucrose cushion and centrifuged in a Beckman SW41 rotor at 28,000 rpm for 1.5 hrs at 4 ^{o}C. The material above the sucrose was collected and desalted into protein binding buffer (PBB; 20 mM Hepes-KOH, pH 7.5, 50 mM KCl) containing protease inhibitors using a Sephadex G-25 column.

Capture of RNA binding proteins with streptavidin magnetic beads and biotinylated RNA

For routine experiments, 75 μL of 10 mg/mL Streptavidin magnetic beads (Roche) were washed twice with 1 mL of oligo binding buffer (OBB; 10 mM Tris-HCl, pH 7.5, 150 mM NaOH). 5' Biotinylated RNA oligos with sequence from either the prolamine 5'CDS zipcode (GAGUUUGAUGUUUUUAGGUCAAAGUUAUUAGGCAAUA) or a non-zipcode region of the prolamine open reading frame (GUUCAGGCCAUAGCGUACGAGCUACAACUCCA-GCAA) were purchased from Integrated DNA Technologies (Coralville, IA). These were added in excess (150 pmol per mg beads) to 250 μL of OBB and the beads were incubated on ice for 20 mins. Unbound RNA was then removed by washing twice with 1 mL of oligo wash buffer (OWB; 10 mM Tris-HCl, pH 7.5, 1 M NaCl). Before the protein extract was added, the beads

were washed twice with PBB and heparin was added to a final concentration of 5 mg/mL. RNase inhibitor was also added to a final concentration of 10 U/mL. Extract and beads were incubated on a rotating platform at 4 ^{o}C for one hour. Beads were then washed three times with 1 mL PBB prior to elution with 0.3 M $MgCl_2$ in PBB. Eluted proteins were loaded on 13 % acrylamide SDS-PAGE gels and visualized by silver staining. In order to produce sufficient material for mass spectroscopy analysis, up to 7 captures were made in parallel from a total volume of 1.5 mL per tube. After elution, the same beads were washed twice with PBB prior to recapture from fresh extract. This recapture and elution step was repeated twice, to give a total capture volume of approximately 30 mL.

***In vitro* UV cross-linking assays**

pBML1, which contains four copies of zipcodes of prolamine 7 under T7 promoter, was linearized by BamHI and used for the template for run-off transcription using T7 MAXIscript kit (Ambion) in order to generate internally ^{32}P-UTP labeled probe. Run-off transcription reactions contained 1 μg of linearized DNA template, 0.5 mM each of ATP, CTP, GTP , 3 μM UTP, 50 μCi of ^{32}P-UTP (3000 Ci/mmol), 20 unit of RNasIn (Ambion), 1X MAXIscript buffer and 2 μL of T7 MAXIscript enzyme in a 20 μL scale reaction. Reaction mixtures were incubated at 37 ^{o}C for 20 min and the DNA template was removed by digestion with 2 units of RNase-free DNaseI at 37 ^{o}C for 15 min. RNA probe was extracted first with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) mixture and then with an equal volume of chloroform:isoamyl alcohol (24:1) and finally ethanol precipitated by addition of 1/10 volume of 5 M ammonium acetate and 3 volumes of ethanol in the presence of 1 μg glycogen. The resulting RNA pellet was rinsed by 70 % ethanol, dried and resuspended in 20 μL of water. Probe quality

was confirmed by running a small amount on a denaturing acrylamide gel and exposing to X-ray film. The incorporation ratio was assessed by scintillation counting of samples taken before and after purification and was generally around 10 %.

For the use in competition assay, unlabeled RNAs were produced from following plasmid DNA templates: prolamine 3'zipcode (pNK15 digested with XhoI), prolamine control CDS (pNK26 digested with EcoRI), glutelin CDS (pSAMG2 digested with HincII), glutelin 3'UTR (pHWAS11 digested with EcoRI), vector (pBSIIKS$^-$ digested with Acc65I). The reaction mixture contained 1 μg of linearized DNA template, 0.5 mM each of ATP, CTP, GTP, UTP, 20 unit of RNasIn (Ambion), 1X MAXIscript buffer and 2 μL of T7 MAXIscript enzyme in a 20 μL scale reaction. Synthesis was performed using the same condition as for labeled run-off transcription except that duration of incubation was increased to 2 hours.

UV cross-linking assays was performed as described by Sammi-subbu et al., (2001). The 20 μL reaction mixture typically contained 20 mM Hepes-KOH, pH 7.5, 50 mM KCl, 0.5 mM DTT, 5 μL of captured protein (approximately 1 μg), 1.5 X 10^5 cpm (approximately 1 pg) of internally labeled prolamine zipcode probe along with other ingredients and competitors as indicated in the figure. The mixture was incubated at room temperature and cross-linked for 7 min on ice using an Ultralum UV cross-linker. The unbound portion of RNA was digested by incubation with 10 μg RNaseA and 10 units of RNase T1 at 37 ^oC for 30 min. After addition of 5X SDS sample buffer, samples were boiled for 3 min and separated by 15 % SDS-PAGE. Gels were fixed in 7.5 % acetic acid, 10 % glycerol solution and dried prior to exposure to X-ray film at -80 ^oC for 12 to 24 hours.

Sample preparation for mass spectroscopy

Samples were prepared for mass spectrometry by in gel trypsin digestion according to the protocol given on the Larkins lab website (http:/ag.arizona.edu/research/larkinslab/) with slight modification. In brief, SYPRO Ruby stained bands were excised, chopped into 0.5 mm^3 pieces and dehydrated by incubation with acetonitrile. Gel pieces were incubated with 200 µL of 10 mM DTT, 100 mM ammonium bicarbonate (ambic) at 56 ^oC for 1 hour to reduce cysteines. Gel pieces were then incubated with 200 µL of 55 mM iodoacetoamide, 100 mM ambic in the dark for 45 min. After washing with 100 mM ambic followed by acetonitrile, gel pieces were dried. 50 µL of trypsin solution (12.5 ng/µL of trypsin in 50 mM ambic, 5 mM $CaCl_2$) was added to the gel piece and it was incubated on ice for 45 min to swell prior to digestion at 37 ^oC for 16 hours. Tryptic peptides were extracted twice with 100 µL of 5 % formic acid, 50 % acetonitrile after 20 min and finally with acetonitrile. Extracted peptides were dried in a vacuum centrifuge and subjected to tandem mass spectrometry analysis.

Analysis of tryptic peptide sequence tags by micro-liquid chromatography tandem mass spectrometry (µLC-MSMS)

Mass spectrometry analysis was done in collaboration with Julian P. Whitelegge, at the University of California. Samples were analyzed by µLC-MSMS with data-dependent acquisition (LCQ-DECA, ThermoFinnigan, San Jose, California) after dissolution in 5 µL 70 % acetic acid (v/v). A reverse-phase column (200 µm x 10 cm; PLRP/S 5 µm, 300 Å; Michrom Biosciences, San Jose) was equilibrated for 10 minutes at 1.5 µL/min with 95 % A, 5 % B (A, 0.1 % formic acid in water; B, 0.1 % formic acid in acetonitrile) prior to sample injection. A linear gradient was initiated 10 min after sample injection ramping to 60 % A, 40 % B after 50

minutes and 20 % A, 80 % B after 65 minutes. Column eluent was directed to a coated glass electrospray emitter (TaperTip, TT150-50-50-CE-5, New Objective) at 3.3 kV for ionization without nebulizer gas. The mass spectrometer was operated in 'triple-play' mode with a survey scan (400-1500 m/z), data-dependent zoom scan and MSMS with exclusion of singly-charged ions. Individual sequencing experiments were matched to a custom rice sequence database that included genomic open-reading frames (TIGR; www.tigr.org), chloroplast open reading frames from rice and mitochondrial open reading frames from *A. thaliana* NCBI (www.ncbi.nlm.nih.gov) using Sequest software (ThermoFinnigan, San Jose). The search was run under the 'no enzyme' mode to identify non-tryptic peptides. The results of Sequest searches were carefully scrutinized. MSMS spectra of doubly charged ions with cross correlation scores (Xcorr) greater than 2.8 and triply charged ions with scores over 3.2, and some others, were examined manually. Where two or more peptides were matched reliably, a strong hit was reported. Where a single good quality peptide hit was returned, a potential hit was reported.

cDNA cloning

RBP-A and RBP-E cDNA was synthesized by reverse transcription of total RNA from mid-stage developing rice seed using oligo(dT) primers or from rice cDNA library, respectively, using established techniques. Whilst the rest of the RBP cDNAs were obtained from the Rice Genome Resource Center (http://www.rgrc.dna.affrc.go.jp/).

RBP-A (AK105751) cDNA was amplified using RBP-A-253f 5'-GATCCAACCCGATCATCAATC-3' and RBP-A-1785r 5'-GATGCACCTGTGGTTCTCTCC-3' primers, phosphorylated and blunt cloned into pBS IIKS^{-}. The fragment was then sub-cloned into pET30a by EcoRI/SacI digestion. RBP-B (AK061072) was cut from the cDNA vector by

NcoI/NotI digestion and sub-cloned into pET30a, whilst RBP-C (AK065955) was sub-cloned into pET30a by NcoI/BamHI digestion. RBP-D (AK059225) was first cloned into pBSII KS$^-$ by EcoRI/NotI digestion and then sub-cloned into pET30c by EcoRI/NotI digestion. RBP-E (AK069121) cDNA was amplified by PCR from a kitaake developing seed cDNA library using pTRGf 5'-TGGCTGAACAACTGGAAGCT-3' and RBP-E nested-1039r 5'-AAACCAGTATCAGACGTGCTATGA-3'. The resulting product was then re-amplified by nested PCR using RBP-E-Nco1-17f 5'-CATTCCCAACCATGGCCTC-3 and RBP-E-HindIII-990r 5'-AATAAGCTTAAAGCCTGCTATCCCATTAAAAG-3' primers. This product was then phosphorylated and blunt cloned into EcoRV cut pBSII KS$^-$ prior to sub-cloning into pET30a by NcoI/BamHI digestion. RBP-I (AK100904) cDNA was amplified by PCR using AK100904-NcoI-217f 5'-GTTGTTGTGTGCCATGGCG-3' and AK100904-BamHI-1411r primers and blunt cloned into pBSII KS which had been previously digested with EcoRV. AK100904 cDNA was then sub-cloned into pET30a by NcoI/BamHI digestion. RBP-J (AK120962) was amplified by PCR using AK120962-BamHI-1f 5'-TAGGATCCATGGAGTCGGATCAGGG-3' and AK120962-XhoI-1401r 5'-CTCCTCGAGCCTCATTTACTGAGCAAC-3' primers. This product was then phosphorylated and blunt cloned into EcoRV cut pBSII KS$^-$ prior to sub-cloning into pET30a by BamHI/XhoI digestion. RBP-K (AK121802) Partial NcoI/BamHI fragment from cDNA clone was sub-cloned into pET30a. RBP-L (AK120834) Not I/partial BamHI fragment from cDNA clone was first cloned into pBSII KS$^-$ and the resulting NotI/XbaI fragment was then cloned into pET30b. RBP-P (AK099896) NcoI fragment from cDNA clone was sub-cloned into pET30b. RBP-Q (AK120636) BamHI/EcoRI fragment from cDNA was sub-cloned into pET30a.

Protein expression and purification

Prior to expression, all plasmids were transformed into an *E. coli* host strain which includes the pLacI-RARE plasmid (Novagene) which encodes rare tRNA synthetases. Typically, a 1 L LB culture was induced with 1 mM IPTG for 3 hours when OD_{600} reached 0.6. Cells were collected by centrifugation, stored at -80 ^{o}C, and prepared for immobilized metal ion affinity chromatography (IMAC) purification. In general, cells were sonicated and extracted in denaturing binding buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 to 15 mM immidazole and 7 M urea, 1 % tritonX-100). Following centrifugation, supernatants were bound to nickel-charged iminodiacetic acid Sepharose resin and washed with binding buffer followed by with 60 mM imidazole containing binding buffer. Proteins were finally eluted by 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 250 mM imidazole or 50 mM EDTA, 4 M urea). Final purification of eluted material was performed by band excision from multiple 3 mm thick SDS-PAGE minigels followed by electroelution. Purified proteins were dialyzed against 1X PBS prior to preparation for injection.

Antibody production and affinity purification

Approximately 200 μg of purified protein in PBS was vortexed with 1 mL of Ribi adjuvant and injected intra-dermally (6X 50 μL), intra-muscularly (2X 300 μL) and subcutaneously (100 μL) to female New Zealand White rabbits to generate polyclonal antibodies. Boost injections were given 28 days after the primary immunization and titer was checked 10 days later by western blot. Terminal bleeds were then collected and partially purified by collecting 30-50 % ammonium sulfate precipitates and dialyzing against 1X PBS. Partially purified polyclonal antibodies were further affinity purified using a technique based on the

immobilization of antigen using irreversibly oxidized Co(III)-IDA resin (Smith et al., 1992). Excess antigen was bound to cobalt-charged iminodiacetic acid Sepharose resin under denaturing conditions and oxidized by incubation with final concentration of 0.03 % hydrogen peroxide for at least 2 hrs. Resin was washed with sequentially lower concentration of urea and finally any remaining unoxidized cobalt was chelated and washed off with 50 mM EDTA. Antigen resin was equilibrated with 100 mM Tris-HCl, pH 8.0, 150 mM NaCl and incubated with partially purified antibody previously diluted 5 fold with equilibration buffer overnight at 4 ^{o}C with rotation. Resin was washed initially by equilibration buffer and then with 100 mM Tris-HCl, pH 8.0, 1 M NaCl. Antibodies were eluted with 100 mM glycine-HCl, pH 2.5, 150 mM NaCl and neutralized using 1M Tris-HCl, pH 8.0 prior to dialysis against PBS and concentrated by ultrafiltration.

Protein expression analysis by western blotting

For protein expression analysis, developing seeds from 1 to 23 DAF (odd days only) were collected in liquid nitrogen and stored at -80 ^{o}C until use. Five developing seeds were directly extracted in 2 mL of 1X SDS-sample buffer with urea (6.4 M urea, 125 mM Tris-HCl, pH 6.8, 2 % SDS, 10 % glycerol, 0.5 mg/mL bromophenol blue, β-mercaptoethanol 50 µL/mL) and centrifuged. 20 µL were loaded on 13 % SDS-PAGE and proteins were transferred to nitrocellulose membrane. Membranes were blocked with 4 % non-fat milk powder in 1X PBS and incubated with primary antibody at 1/1000 overnight at 4 ^{o}C. Membranes were washed 3 times with 1X TBST containing 0.05 % Tween 20 and incubated with anti-rabbit HRP at 1/5000 for 1 hour at room temperature. After washing 3 times with 1X TBST, chemiluminescence

substrates (Super Signal west pico kit, Pierce) were applied and membranes were exposed to X-ray film.

For analysis of tissue specific expression, leaf, root, mid-developing seed and mature seed were ground to powder with liquid nitrogen using a mortar and pestle, extracted in 1X SDS-sample buffer lacking bromophenol blue and 50 μg total protein was applied to 13 % SDS-PAGE gels. The rest of western blot procedures used for this and all other experiments were performed as described above.

Amplification of full length cDNA for RBP-D

Total RNA preparation:

10 grains of dehulled 10-14 DAF developing seeds (yielding 0.20 g material) were ground to powder with liquid nitrogen using a mortar and pestle, extracted with 2 mL of TRIzol (Invitrogen) and centrifuged to remove starch and cell debris. 600 μL of supernatant were then extracted with 120 μL of chloroform: isoamyl alcohol (24:1) and centrifuged. 200 μL of water-phase were mixed with 700 μL of RLT buffer, 500 μL ethanol and applied onto an RNeasy column (QIAGEN). Columns were washed with 500 μL each of RPE buffer followed by 80 % ethanol and RNAs were eluted in 14 μL of nuclease-free water which generally yielded 40 μg of total RNA.

Reverse transcription-polymerase chain reaction (RT-PCR):

Oigo(dT) primed cDNA was produced using Moloney murine leukemia virus (M-MLV) reverse transcriptase in a 50 μL reaction containing 1X MMLV buffer, 0.5 mM dNTPs, SUPERaseIn 10 U (Ambion), oligo(dT) 100 ng, 15 μg of total RNA, MMLV 40 U by incubating at 37 ^{o}C for 1 hour. RBP-D was initially amplified from oligo(dT) primed cDNA using RBP-D-

outside-F: CGAAGCTCCGGTTAGCTCGG and RBP-D-outside-R: CTGCAGCAGCTGAC-CGATAGC primers with the following conditions. 94 ^{o}C 3min, 4 cycles of 94 ^{o}C 30 sec, 47 ^{o}C 30 sec, 72 ^{o}C 1 min 30 sec, 26 cycles of 94 ^{o}C 30 sec, 53 ^{o}C 30 sec, 72 ^{o}C 1 min 30 sec, 72 ^{o}C 5 min. The product was then re-amplified by nested PCR using RBP-D-BglII-F: CTC*AGATCT*ATGGCGGGGTACGGGGAGG and RBP-D-BamHI-R: CAGAT*GGATCC*GAA-AATGTAAC with the following conditions. 94 ^{o}C 3min, 4 cycles of 94 ^{o}C 30 sec, 45 ^{o}C 30 sec, 72 ^{o}C 1 min 30 sec, 26 cycles of 94 ^{o}C 30 sec, 58 ^{o}C 30 sec, 72 ^{o}C 1 min 30 sec, 72 ^{o}C 5 min. PCR products were phosphorylated and sub-cloned into the EcoRV site of pBSIIKS^{-}. Sequencing reactions were performed to confirm the homology of the cloned cDNA with genomic DNA.

Indirect fluorescence labeling of developing rice seeds for confocal microscopy

Mid-developing rice seeds were hand-sectioned as thinly as possible and placed in 100 μL microtubule stabilizing buffer (MTSB; 50 mM Pipes-KOH, pH 6.9, 10 mM EGTA, 10 mM $MgSO_4$, 1 % DMSO, 0.1 % Triton X-100) supplemented with 1 mM DTT and 0.2 μM PMSF and incubated for 10 min at room temperature. Fixation was performed using 4 % paraformaldehyde in MTSB (with DTT/PMSF) for 30 min followed by three washes with MTSB. Cell wall digestion was performed by incubation with 100 μL pre-warmed 1 % cellulase solution (Cellulose Onoduka R-10, SERVA) at 37 ^{o}C for 5 min. Samples were then treated with PBS, 0.5 % Triton X-100 for 15 min. Prior to incubation with antibodies, seed sections were treated with blocking buffer (2 % BSA in PBS, 0.1 % Triton X-100) for 1 hr. Antibodies were used at a dilution of 1:50 to 1:200 in blocking buffer and sections were incubated for 3 hrs or overnight before washing three times using PBS containing 0.01 % Triton-X100. Secondary antibody,

either Alexa Fluor 594 or Fluorescein conjugated goat anti-rabbit or anti-mouse antibodies (Molecular Probes), was added at a dilution of 1:200 in blocking buffer and sections were incubated for 1 hr, with Rhodamine B hexyl ester at a final concentration of 1 μM or $DiOC_6$ at a final concentration of 5mg/L being added to the sample volume (when required) after 40 mins. Following three 5 min washes with PBS, 0.1 % Triton X-100 and a final wash with PBS alone, Vectashield anti-fade mounting medium (Vector Laboratories, Inc. Burlingham) was added to each well followed by a number one coverslip which was sealed with nail polish.

Confocal microscopy was performed on a Zeiss 510 series laser scanning confocal microscope (Jena, Germany) and a Bio-Rad View Scan DVC-250 laser scanning confocal microscope using the fluorescein and rhodamine filter sets. Image processing was performed using Zeiss LSM Image browser (Munich, Germany), Adobe Photoshop (Mountain View, CA), NIH Image, or Microsoft PowerPoint (Redmond, WA).

Microarray analysis of developing rice seed mRNA levels

Methods for probe preparation were adapted from the manufacture's instruction for TRIzol (Invitrogen), RNeasy kit (QIAGEN) and Amino Allyl MessageAmp^{TM} aRNA amplification kit (Ambion). The hybridization protocol was adapted from the website of NSF Rice Oligonucleotide Array Project (http://www.ricearray.org/rice_protocols.shtml).

Total RNA preparation:

10 grains of dehulled 10-14 DAF developing seeds (generally yielding 0.2 g) were ground to powder with liquid nitrogen using a mortar and pestle, extracted with 2 mL of TRIzol and centrifuged to remove starch and cell debris. 600 μL of supernatant were then extracted with 120 μL of chloroform: isoamyl alcohol (24:1) and centrifuged. 200 μL of water-phase were mixed

with 700 μL of RLT buffer, 500 μL ethanol and applied onto RNeasy column. Columns were washed with 500 μL each of RPE buffer followed by 80 % ethanol and RNAs were eluted in 14 μL of nuclease-free water which generally yielded 40 μg of total RNA.

First strand cDNA synthesis:

2 μg of total RNA were used for first strand cDNA synthesis in a 10 μL reaction. Total RNA was first primed with T7 Oligo(dT) primer and cDNA was synthesized using ArrayScript by incubation at 42 ^{o}C for 2 hours.

Second strand cDNA synthesis:

Second strand cDNA synthesis was performed to generate double stranded transcription template by adding 40 μL of second strand synthesis reaction mixture, which contained DNA polymerase, and was incubated at 16 ^{o}C for 2 hours. Double stranded DNA was then purified by mixing with 250 μL of cDNA binding buffer, applied onto the supplied column, washed and then eluted twice using 2 X 10 μL of nuclease-free water. Purified dsDNA was concentrated to 7 μL using a vacuum centrifuge.

In vitro transcription:

13 μL of *in vitro* transcription reaction mixture was mixed with dsDNA template and antisense RNA (aRNA) was then amplified in order to incorporate amino allyl UTP (aaUTP). The reaction was incubated at 37 ^{o}C for 16 hours and was terminated by the addition of 80 μL nuclease-free water. aRNA was then mixed with 350 μL of aRNA binding buffer, 250 μL of ethanol and purified using the supplied columns. After washing, aRNA was eluted in 100 μL of nuclease-free water.

Dye coupling reaction:

40 μg of purified aRNA was dried to completion in a vacuum centrifuge and resuspended in 9 μL of coupling buffer. 40 nmol *N*-hydroxysuccinimide (NHS) ester Cy3 dye (GE Health care) was dissolved in 11μL of DMSO and mixed with aRNA coupling buffer in the dark for 1 hour. The reactions were stopped by incubation with 4.5 μL of 4M hydroxylamine and further for 15 min followed by dilution to 30 μL total volume with water.

Labeled aRNA purification and fragmentation:

aRNA was mixed with 105 μL of aRNA binding buffer, 75μL of ethanol and applied onto the supplied columns. After washing, labeled aRNA was eluted twice using 2X 14 μL of nuclease-free water and dried down to 18 μL. 2 μL of 10X fragmentation buffer (Ambion) was added and samples were incubated at 70 ^{o}C for 15 min before the addition of 2 μstop solution. Probes were then dried to completion in a vacuum centrifuge.

Hybridization:

Rice NSF45K arrays were purchased through the NSF Rice Oligonucleotide Array Project (http://www.ricearray.org/index.shtml) described by Jung et al., (2008). These paired array slides contain 43,311 50 to 70-mer oligos based on 45,116 gene models, each of which is supported by either an EST and/or full-length cDNA in TIGR rice annotation release 3. Array slides were pre-hybridized with 5X SSC, 0.1 % SDS, 1 % BSA at 42 ^{o}C for at least 2 hours with rotation and then rinsed twice, each with 10 dipping cycles in water before drying a stream of nitrogen gas. The probe was resuspended in 58 μL of hybridization buffer (50 % formamide, 5X SSC, 0.1 % SDS, 10 mM DTT) and supplemented with 20 μg of salmon sperm DNA. Probe was then heat denatured for 3 min, snap cooled on ice and centrifuged to remove any precipitates before 28 μL was applied onto the pre-hybridized A and B slides. Coverslips were placed on the slides, which

were then sealed in the hybridization chambers and incubated at 48 ^{o}C for 16 hours in the dark. After hybridization, slides were placed in staining dish with low stringency buffer (2X SSC, 0.1 % SDS) at 48 ^{o}C and washed for 10 min twice with stirring. Washing buffer was exchanged with high stringency buffer (0.1X SSC, 0.1 % SDS) and slides were washed for 5 min twice with stirring at room temperature. Two final washes were performed with 0.1X SSC for 10 min each at room temperature. Slides were then dried under a stream of nitrogen gas.

Slide scanning and data analysis:

Slides were scanned with at 10 μm resolution using a ScanArray 4000XL (Packard BioScience, now PerkinElmer) and the resulting TIF images were analyzed using GenePix Pro 6.1 (Molecular Devices, Sunyvale, CA). Average of total intensities from nine experiments including six biological replicates was used as factor to calculate normalized intensity value from the following oligonucleotide IDs (in brackets) for each candidate RBPs. RBP-A (TR069611), RBP-B (TR050180), RBP-C (TR073107), RBP-D (TR072362), RBP-F (TR072020), RBP-G (TR049850), RBP-H (TR035250), RBP-I (TR035964) RBP-J (TR055170), RBP-K (TR051953), RBP-L (TR071546), RBP-M (TR043139), RBP-O (TR064934), RBP-P (TR031129), RBP-Q (TR032642). The resulting data was then analyzed using Microsoft Excel (Redmond, WA).

Analysis of RBP-A tissue specific expression pattern

Expression levels of RBP-A in a variety of tissues as indicated in the Figure 16. were analyzed using experimental data sets available from Jiao et al., 2009. Data was analyzed using Microsoft Excel (Redmond, WA).

Results

Capture of RNA binding proteins from a cytoskeleton-enriched fraction of developing rice seeds using zipcode and control RNA bait sequences

In rice endosperm cells, the majority of protein synthesis occurs in the cortical region located close to the plasma membrane (Okita and Choi, 2002). This region is rich in microtubules and actin filaments (Muench et al., 2000), the latter of which are required for the targeting of prolamine mRNA (Hamada et al., 2003b) and may act as a scaffold for polysome attachment (Davies et al., 1991; Muench et al., 1998). Our strategy, therefore, was to generate a cytoskeleton-enriched fraction from which putative prolamine RNA binding proteins (RBPs) could be captured using prolamine zipcode sequences as bait.

Figure 1A shows a schematic representation of the prolamine gene including the position and length of the biotinylated bait RNAs which were used for the RBP capture. Prolamine contains two distinct zipcodes, the first lies within a 60 nucleotide region of the 5'CDS, and the second within a 40 nucleotide region of the 3'UTR. Both zipcodes are required for RNA localization to the protein body ER (Hamada et al., 2003a). RNA structure modeling using mfold (Matthews et al., 1999; Zuker, 2003) indicates that both prolamine zipcodes form stem-loop structures at opposite ends of an essentially linear structure (Figure 1B). This could potentially allow the RNA molecule to form a loop, bringing both zipcodes together and allowing a single RBP with multiple domains (or multiple RBPs within a complex) to bind them simultaneously. Interestingly, many genes within the prolamine superfamily, including prolamine 7 under investigation here, lack introns (Kim and Okita, 1988). This means that the mature mRNA does not contain any exon-junction complexes, complexes which are known to influence RNA stability and localization (Hachet and Ephrussi, 2004).

A cytoskeleton-enriched extract was prepared from mid-stage developing rice seeds by initial extraction with a low salt, cytoskeleton stabilizing buffer (CSB) and differential centrifugation (Muench et al., 1998) followed by sucrose density gradient centrifugation to remove any residual chromatin and nuclei. After desalting, extracts were incubated with streptavidin magnetic beads to which biotinylated bait RNA had previously been bound. Bait sequences included the 36 nucleotide prolamine 5'CDS zipcode and a control oligo of the same length (Figure 1A) randomly selected from a non-zipcode region of the prolamine open reading frame to assess binding specificity. Heparin was included in the binding buffer at a concentration of 5 mg/mL to reduce non-specific binding of proteins to the RNA. Following capture, the beads were washed with binding buffer and bound proteins were eluted by the addition of 0.3 M $MgCl_2$ prior to analysis by SDS-PAGE. An initial assessment of the population of proteins captured by the 5'CDS and 3'UTR zipcodes revealed that they were essentially the same (data not shown). This is not surprising given the high degree of homology between the two zipcode bait sequences which are both predicted to form a hairpin structure (Figures 1B and 1C; see discussion). From this point on, only the control and 5'CDS oligos were used in protein capture experiments.

The 5'CDS prolamine zipcode bait sequence captures significantly more proteins than the control RNA sequence

Figure 2A shows a comparison of proteins captured using the prolamine 5'CDS zipcode RNA and the control RNA. Only three major polypeptide species are captured by the control oligo, whilst many more proteins are captured by the prolamine 5'CDS zipcode bait (Figure 2A). The large difference between the number and abundance of proteins captured using the zipcode and control sequences is most likely attributable to the ability of the zipcode-containing bait

RNA to form a more stable stem-loop structure, a favored target for RBPs (Svoboda and Cara, 2006).

An assessment of the strength of the protein-RNA interactions was made by performing sequential elution of captured material using solutions of increasing ionic strength (Figure 2B). For both control and zipcode capture experiments, only a minor amount of protein was released using 0.2 M NaCl, but sequential elution with 0.5 M NaCl and 0.3 M $MgCl_2$ released much more of the total captured protein. The three prominent polypeptide bands observed following capture with the control RNA also appear to bind, presumably non-specifically, to the zipcode RNA.

In vitro UV cross-linking assays show that some of the captured proteins retain RNA binding activity

In order to assess the RNA binding activity and specificity of eluted candidate RBPs, UV cross-linking experiments were undertaken. Internally ^{32}P-labeled prolamine zipcode RNA was incubated with 5'CDS zipcode captured material in the presence or absence of competitors as indicated in Figure 3. Following UV cross-linking, free RNA and unprotected regions of bound RNAs were digested by RNase, resolved by SDS-PAGE and exposed to X-ray film. The results show that at least three polypeptides in the eluted material retained RNA binding activity following high salt elution and dialysis. These activities were all sensitive to high concentrations of KCl. The inhibition of cross-linking seen with the addition of magnesium chloride was surprising since it has often been reported that magnesium stabilizes protein-RNA interactions (Lambert and Draper, 2007). Chelating residual magnesium ions by the addition of EDTA slightly increased the binding activity of RBPs. As expected, the observed RNA binding activities were not affected by the presence of the non-specific competitor heparin, but could be

competed for by yeast tRNA (Figure 3A), perhaps reflecting a preference for highly structured RNAs.

In order to further explore the RBPs specificity, competition assays was performed between prolamine zipcodes and polyribonucleotide homopolymers or unlabelled RNAs as indicated in Figure 3B. RNA binding activities present in the analyzed material bound to poly (G), but not to other RNA homopolymers. Binding to a variety of prolamine and glutelin zipcode sequences was also observed, but there no significant was between them and a vector control in the competition assay. Similar results, high binding specificity to polyribonucleotide homopolymers but not to zipcodes, have been reported in *in vitro* binding studies performed by others (Li and Sugiura, 1991; Lambermon et al., 2002; Dreyfuss et al., 1993). Although individual RBPs are able to bind to prolamine zipcodes RNA, the lack of sequence specificity suggests a loss of function during the high salt elution step following affinity chromatography, most likely due to dissociation of protein complexes. In support of this view, very little of the material captured by the prolamine zipcode bait could be re-captured following elution and dialysis to reduce the salt concentration of the eluted material.

Many of the captured binding proteins are members of the hnRNP protein family

Having demonstrated that the majority of the proteins captured using the 5'CDS prolamine zipcode bait exhibit both strong and specific binding, we proceeded with their identification using tandem mass spectroscopy. After elution with 0.3 M $MgCl_2$, proteins were separated by SDS-PAGE and visualized using SYPRO Ruby staining. Excised bands were subjected to in gel trypsin digestion and were submitted for microbore capillary liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS).

Table 1 shows the 18 distinct proteins that were positively identified by tandem mass spectroscopy. The observed tryptic peptides are given in Table 2. These proteins were arbitrarily designated RBP-A through RBP-R. Many of the proteins identified share significant homology with members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family, proteins which have known roles in RNA maturation, localization and translation (reviewed by Krecic and Swanson, 1999; Bolognani and Perrone-Bizzozero, 2008).

In addition to the seven hnRNPs, other proteins identified include two isoforms of malate dehydrogenase, a cold-shock DNA binding protein, ribosomal L3, a hypothetical protein, a putative RNA binding protein (RBP45), an oligouridylate binding protein, a glycine-rich RNA binding protein and three chloroplast RNA binding proteins. The presence of these latter three nuclear encoded chloroplast-targeted proteins in our cytoskeleton-enriched fraction could be the result of contamination from chloroplast-containing rice seed husks, from amyloplasts within the seed tissue, or due to capture of these proteins prior to import into plastids. Clearly though, our *in vitro* capture experiments have led to the identification of a significant number of genuine RNA binding proteins. The homology between the identified putative prolamine zipcode binding proteins is indicated by the phylogram in Figure 4.

cDNA cloning, expression and antibody production

The absence of significant RNA binding specificity following initial capture indicated that alternative strategies would be required in order to further assess the possible roles of the identified RNA binding proteins. The approach taken was to determine their intracellular location, since RBPs associated with RNA transport and localization would likely co-localize with cellular features related to these functions. Accordingly, cDNAs corresponding to several of

the identified proteins were either cloned from developing rice seed or obtained from the Rice Genome Resource Centre (RGCR; www.rgrc.dna.affrc.go.jp). These cDNAs were then expressed in *E. coli* in order to generate antigen for the production of antibodies in rabbits. Figure 5 summarizes the process by which His-tagged fusion proteins underwent initial purification using immobilized metal affinity chromatography (IMAC) followed by final purification using single slot preparative protein gels. High titer rabbit antibodies were generated and affinity purified by immobilizing antigen using irreversibly oxidized Co(III)-IDA resin (Smith et al., 1992). Efforts were made to purify *E. coli* expressed RBPs in order to assess their RNA binding specificity individually by UV cross-linking assay, but unfortunately over-expressed proteins were found almost exclusively in inclusion bodies and failed to refold after extraction despite repeated attempts using a variety of different expression and refolding conditions.

Several of the identified RNA binding proteins show strong binding specificity for the prolamine 5'CDS bait sequence during initial capture

Following the generation of antibodies against several of the identified proteins, RBP capture experiments were repeated using protein gel blotting to assess binding specificity for the 5'CDS zipcode and the control sequence. Figure 6 shows that of the six proteins analyzed to date, the majority showed either good or excellent specificity. In particular, the hnRNPs RBP-A, RBP-D and RBP-I are readily detected following capture using the zipcode RNA but are not captured by the control RNA. RBP-E, a chloroplast cp31 RNA binding protein, showed very little binding specificity, suggesting that this protein has a more general RNA binding activity.

One concern with this result is the apparent protein degradation observed following capture. This is most clearly seen in the case of RBP-A which runs at significantly less than its predicted molecular weight. This degradation is probably a result of the lengthy protocol used for extraction and capture since direct extraction of seeds in urea-containing SDS sample buffer gave bands of the appropriate molecular weight when western blots were probed with anti-RBP-A through E (Figure 7). The fact that degradation is occurring during the course of the binding experiments may mean that the observed specificity is potentially different than it would be for the full length protein.

Analysis of RBP protein expression in different tissues

In order to assess the quality of our antibodies and to obtain additional information regarding the tissue-specific protein expression levels of the identified RBPs, western blotting was performed using crude antibodies and extracts prepared from developing seeds, mature seeds, leaf and root tissue. Figure 7 shows the results indicate that all of the RBPs analyzed are significantly more abundant in developing seeds than in fully dried, mature seeds. Whether or not this is due to protein degradation during the latter stages of seed maturation is not currently clear. All of the RBPs tested are expressed in leaf tissue and only RBP-B is absent from roots. The presence of multiple reactive bands in the RBP-C (malate dehydrogenase) western indicates that our antibody is cross-reacting with a number of related isozymes. Consequently, this antibody cannot easily be used in future localization studies.

This western analysis confirms that, as hoped, only a single major polypeptide band of the appropriate molecular weight is seen for RBP-A, -B, and -E. Crude RBP-D antiserum detected an additional high molecular weight band in dry seed, but this problem was resolved by

affinity purification of the antibody. What did not change, however, was the fact that the size of the protein detected in developing seed, leaf and root tissue was larger than predicted from the cDNA we obtained.

Isolation of full length RBP-D cDNA

Western blot analyses of RBP-D from the extract using 6M urea containing SDS-sample buffer using purified antibody revealed that only a single band, of around 40 kDa, is recognized (Figure 7). This is considerably larger than the estimated size of 29 kDa predicted from the cDNA clone (AK059225) shown in Table 1. Additionally, the molecular weight of the largest band which reacts with anti-RBP-D following captured with the 5'CDS zipcode also has a molecular weight of approximately 40 kDa. As shown in Figure 8, the AK059225 has a truncated N-terminal RRM domain. Analysis of the genomic clone database indicated that the reading frame of RBP-D extends considerably upstream (Figure 9A and 9B). Efforts were thus made to isolate a full length RBP-D cDNA to establish whether our western blots were correct in the detection of a 40 kDa protein. Subsequently, a full length cDNA clone, which contains the first RRM in its entirety was isolated by RT-PCR using developing rice seed total RNA (Figure 9C). However, due to the extremely G-rich nature of the DNA sequence at the 5' end, the sequence of the 120 bp immediately following the start codon of the putative full length cDNA clone could not be confirmed. This work did, however, confirm that the molecular weight of the protein reacting with our antibody was appropriate.

With this knowledge in hand, we wanted investigate the role of RBP-D and the other identified proteins *in vivo*. The subcellular localizations of candidate prolamine RNA binding proteins within developing rice seed sections was therefore investigated to gain additional

evidence to allow us to determine whether individual proteins were worthy of further study. For this purpose, antibodies bearing high titer and good specificity were used to localize candidate RBPs.

RBP-A is closely associated with cortical ER membranes and microtubules

Immunolocalization of the hnRNP family member RBP-A using affinity purified antibodies reveals that in addition to being present in the nucleus (confirmed by co-localization with propidium iodide as shown in Figure 10A), RBP-A is also present on the cortical ER network. Interestingly, as can be seen in Figure 10B, RBP-A signal (red) does not merge with green $DiOC_6$ signal (an ER vital dye which preferentially stains the PB-ER membrane) as shown in Figure 10B. At higher magnification (Figure 10C), RBP-A is observed as small regions of intense staining on cisternal ER membranes, suggesting that its distribution on the surface of this ER subdomain is heterogeneous. In addition to this characteristic ER membrane staining pattern, the majority of small RBP-A particles also co-localize with microtubules (Figure 10D), structures which are present exclusively in the cortical region of endosperm cells. Collectively, this protein localization data is consistent with extra-nuclear RBP-A having a role in either mRNA anchoring, translational regulation, or in the modulation of RNA stability on microtubules and/or cisternal ER membranes.

RBP-D is present in particles which co-localize with actin filaments and are closely associated with prolamine protein bodies

In light of the excellent binding specificity for the 5'CDS prolamine zipcode exhibited by RBP-D in our *in vitro* binding assay, we were interested in localizing this protein in order to gain

additional insight into its possible cellular function. Confocal immunofluorescence microscopy was performed using developing rice seed sections to localize RBP-D with respect to either the actin cytoskeleton (the mode of transport of prolamine mRNA) or to prolamine protein bodies (the destination site). In contrast to RBP-A, which is present on cisternal ER membranes and associated with microtubules, RBP-D has a very different distribution. Although, primarily nuclear, significant numbers of cytosolic particles (with a mean diameter of ~0.2 μm), were readily detected, primarily in the outer, cortical region of developing endosperm cells. Figure 11A shows that these RBP-D-containing particles were found in close proximity to actin filaments and adjacent to, and occasionally contacting, prolamine protein bodies (Figure 11B). Although not conclusive, when taken together with its excellent *in vitro* binding specificity, the localization data obtained for RBP-D strengthens the possibility that this protein is indeed involved in actin-mediated prolamine mRNA localization.

RBP-I is present exclusively in the nucleus

Another candidate protein which exhibited strong binding specificity for the 5'CDS prolamine zipcode in our *in vitro* binding assay is RBP-I. Figure 12 shows that in contrast to both RBP-A and RBP-D, RBP-I is found almost exclusively in the nucleus, with very little cytoplasmic protein detected, This may possibly be due to differences in the status of complex formation such that RBP-I is accessible to antibody when it is in the nucleus and inaccessible when masked by partner proteins present within a cytoplasmic complex. Further fractionation studies should shed light on this possibility.

RBP-B and RBP-E localize to discrete spots surrounding intracellular structures

The single cp29 and two cp31 chloroplast RNA binding proteins identified in this study are nuclear-encoded and are known to have roles in RNA editing, stabilization and in binding telomeric DNA (Tillich et al., 2009; Nakamura et al., 2001). Biochemical analysis has suggested that cp29 and cp31 are present in the chloroplast stroma, but a GFP fusion with cp31 localizes to the nucleus (Nakamura et al., 2001; Krause and Krupinska, 2009). Although both RBP-B (cp29) and RBP-E (cp31) exhibit little or no specificity for the prolamine 5'CDS zipcode bait (Figure 6), we were interested localize these proteins to assess their possible role in endosperm cells and to establish why these chloroplast RNA binding proteins were captured. Are they present as contaminants from chloroplast containing seed coats or do they actually have a role in endosperm cells which have plastids instead of chloroplasts? The results of our localization studies, shown in Figures 13A and B, demonstrate that RBP-B and RBP-E surround irregularly shaped structures (2-5 μm in diameter) and are present at high concentrations in discrete regions. Although RBP-E localizes to these structures, weak signals can also be found on PB-ER as indicated by yellowish protein bodies seen in the merged image. This suggests RBP-E may be acting as a general RNA binding protein present in the cortical ER including PB-ER, however co-localization with ER markers such as BiP and PDI would be needed to confirm this. Whilst RBP-F is also a cp31, it seems to be less abundant and is not found on PB-ER membranes (Figure 13C). Collectively, this localization data suggests that the chloroplast RNA binding proteins identified in this study (RBP-B, RBP-E and RBP-F) may play interesting roles in developing rice seeds.

Other classes of RBPs were also localized using confocal microscopy to gain additional information regarding their potential roles in prolamine mRNA localization. RBP-N (glycine rich

RBP) shows some bright particles as well as thread like patterns (Figure 13D) whilst RBP-Q (hnRNP) is localized to small particles on cisternal ER membranes (Figure 11E and at higher magnification in 11F). RBP-Q therefore appears to be a good candidate for a factor involved in RNA localization to cisternal ER membranes. Further co-localization experiments with cytoskeletal elements and additional biochemical analysis should shed light on this hypothesis.

RBP-A and RBP-D protein expression profiles are consistent with a role in storage protein RNA localization

In order to correlate the temporal expression patterns of the newly identified RBPs protein levels with those of the rice seed storage proteins, developing rice seeds were collected from 1 to 23 days after flowering (DAF), extracted with 6M urea containing SDS-sample buffer and analyzed by western blotting using antibodies raised against RBP-A, RBP-D, RBP-E, RBP-I, RBP-Q, 13 kDa prolamine and glutelin. As can be seen in Figure 14, RBP-A expression levels closely match those of prolamine and glutelin, rapidly increasing from 11 DAF and remaining high during the latter stages of seed development. RBP-D expression peaks at 11 DAF before dropping significantly as seeds mature, a profile that closely matches prolamine mRNA expression (Su et al., 2001). The expression profile of RBP-E (which is representative of the other chloroplast RBPs, RBP-B and RBP-F), is almost the inverse of RBP-A. It is readily detected early in seed development and then disappears at 13 DAF, just as storage proteins begin to accumulate. The expression levels of RBP-I remain relatively constant during seed development, suggesting that it plays a different role to RBP-A and RBP-D, whilst RBP-Q protein levels increasing from 1-9 DAF and then remain at a high level throughout the later

stages of seed development. Whilst not outstanding on their own, these results provide additional evidence supporting roles for both RBP-A and RBP-D in the localization of prolamine mRNA.

Tissue-specific RNA expression levels of candidate RBPs

Despite the wide array of tissue-specific RNA expression data from the rice atlas project (Jiao et al., 2009), there is no data available on RNA expression levels in rice endosperm tissue. An attempt was therefore made to address this deficiency using whole genome microarray analysis. This would allow the expression profiles of the candidate prolamine zipcode binding proteins to be compared between rice seeds and the many other tissues and data points (totaling 42) for which expression data is readily available.

10-14 DAF developing seeds were harvested and total RNA isolated for probe preparation to hybridize onto rice 45K microarray slides. A total of nine replicates including six unique biological samples were analyzed and the average value for each candidate RBP's signal intensities are shown in Figure 15. RBP-A (hnRNP), RBP-H (cold-shock protein), RBP-O (ribosomal L3) are highly expressed compared to the other candidate proteins. Cold-shock proteins are known to have a role as RNA chaperones to unwind RNA secondary structure and are involved in developmental processes as well as cold acclimation (Sasaki et al., 2007). Examination of RBP-A expression patterns in a variety of tissues using the data obtained from the work of Jiao et al., (2009) along with our microarray data from developing seeds reveals that RBP-A is highly expressed in root tips, axilliary primodia and apical meristems as well as in developing endosperm tissue. This suggests that RBP-A may function as an RNA binding protein, potentially involved in localizing RNA, in rapidly dividing polarized cells.

Discussion

Our laboratory has previously shown that there are two distinct regulated RNA localization pathways which target prolamine and glutelin messenger RNAs to the PB-ER and cisternal ER respectively (Hamada et al., 2003a). Furthermore, the movement of particles containing GFP-tagged prolamine mRNA in living cells was shown to be dependent on intact actin microfilaments (Hamada et al., 2003b). The cytoskeleton, therefore, clearly plays a key role in the targeting of rice storage protein RNAs to the cortical ER and is also likely to be involved in a third constitutive pathway to the cisternal ER (Choi et al., 2000; Hamada et al., 2003a).

We have previously demonstrated that targeting of prolamine RNA to the PB-ER requires two partially redundant zipcode sequences (Hamada et al., 2003a). In an effort to identify *trans*-factors that may be involved in the recognition of these prolamine RNA *cis*-elements, we employed these zipcode sequences as baits to isolate and identify RBPs from cytoskeleton-enriched extracts of developing rice seeds. Preliminary experiments demonstrated that both the prolamine 5'CDS and 3'UTR zipcode sequences captured essentially the same population of proteins (data not shown). This is perhaps not surprising since analysis of both the sequence and predicted structure of the prolamine 5'CDS and 3'UTR bait sequences used in this study reveal that the stem-loop structures predicted for both molecules are remarkably similar. Figure 1C shows that both contain a stem of 8 (3'UTR) or 9 (5'CDS) nucleotides in length and a loop containing either GUUUUAG (5'CDS) or GUUUA (3'UTR). Given the similarity of these two zipcode sequences, we decided to focus only on the 5'CDS prolamine zipcode and a non-zipcode control RNA of the same length.

Biotinylated RNA baits were presented on streptavidin magnetic beads during protein capture and, after washing and elution, provided a rapid and efficient single step isolation

procedure. The inclusion of a high concentration of heparin, a highly charged polyanionic molecule, was used to prevent non-specific binding to RNA present on the beads. Performing capture experiments in the absence of heparin led to the detection of many more proteins (data not shown), including the RNA binding protein Tudor-SN which is known to be efficiently captured on a heparin column (Sami-Subbu et al., 2001).

The extensive RNA binding specificity observed when using the prolamine zipcode as a bait was not evident when *in vitro* RNA binding assays were conducted with the eluted RNA binding activities. This loss of binding specificity was likely due to the high ionic strength conditions required for elution of the RNA binding proteins from the zipcode containing magnetic beads. Although the exact basis for this loss of RNA binding specificity is not known, a likely reason is the disruption of higher order protein structure (discussed in more detail below). In light of the lack of observable specificity following elution of captured material, we chose to identify the binding proteins captured by the 5'CDS zipcode using mass spectrometry so that antibodies could be generated for use in further characterization of the more promising candidates.

Scaling-up of the capture experiment allowed sufficient material to be generated for reliable identification of eighteen distinct proteins. These RNA binding proteins ranged in size from 14 kDa to 50 kDa, with the vast majority containing one or more RNA recognition motifs (RRMs), a domain commonly found in RNA binding proteins (reviewed by Shamoo et al., 1995; Maris et al., 2005). The most prominent homology amongst the identified proteins was with members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family. These proteins are known to be involved in a wide range of cellular processes within both the nucleus and the cytosol, ranging from transcription and pre-RNA processing to cytoplasmic mRNA transport,

translation and turnover (reviewed in Krecic and Swanson, 1999; Dreyfuss et al., 2002; Hazelrigg, 2004; Kress et al., 2004; Percipalle et al., 2009). It has been proposed that after mRNA splicing, many hnRNPs remain associated with mRNAs and accompany them to the cytosol where they provide a molecular history to assist in mRNA localization and translation (Martinez-Contreras et al., 2007; reviewed by Dreyfuss et al., 2002).

In addition to the RRM-containing proteins, several other proteins were identified including a cold shock DNA binding protein, ribosomal L3, three chloroplast RNA binding proteins and two glyoxysomal malate dehydrogenases (MDH). MDH is one of many proteins which use adenine and its derivatives as ligands and are therefore frequently able to bind to adenine in both DNA and RNA (reviewed in Denessiouk et al., 2001). For example, nucleocytoplasmic MDH can act as a transcriptional regulator in response to glucose starvation (Lee et al., 2009) and lactic dehydrogenase has been identified as an AU-rich element (ARE) binding protein which serves as a signal for mRNA degradation and translational repression (Pioli et al., 2002). Whether the MDH enzymes identified in this study also have a cellular function in cytosolic RNA translation, targeting, or degradation remains to be determined, but increasingly research is showing that many RNA binding proteins can act as multifunctional proteins (Fedoroff, 2002; Guil and Cáceres, 2007; Sawicka et al., 2008). One such example is a multifunctional peroxisomal protein that has been shown by immunofluorescence microscopy to be present on both peroxisomes and microtubules (Chuong et al., 2002; Muench and Mullen, 2003). In addition to binding RNA *in vitro* with high affinity, this protein, like almost all of those identified here, lacks a conventional nuclear localization signal. This is not surprising since it is known that many nuclear proteins (including hnRNPs), often do not contain this motif, but

instead have an M9 domain which mediates interaction with other NLS-containing proteins (Siomi and Dreyfuss, 1995; Iijima et al., 2006).

It has frequently been reported that *in vivo* RNA binding specificity is achieved by the interaction of multiple individual binding proteins (Fedoroff, 2002; Gonsalvez et al., 2004; Anderson and Kedersha, 2009). For example, a complex of multiple RBPs has been identified that specifically recognizes the RNA localization signal of the Drosophila *bicoid* RNA (Arn et al., 2003). Hence, failure to demonstrate meaningful binding specificity *in vitro* is often attributed to the dissociation of these protein complexes or their inability to form *in vitro*. However, there are exceptions to this, and RNA binding specificity has been demonstrated *in vitro* for hnRNP A2 from oligodendrocytes (Hoek et al., 1998) and for the homologous proteins ZBP-1 (Ross et al., 1997) and Vg1RBP (Vera; Deshler et al., 1998) which recognize zipcodes in β-actin and Xenopus *Vg1* mRNA respectively. The fact that we cannot demonstrate sequence specific binding following elution of the prolamine zipcode binding fraction by UV cross-linking and that this same fraction cannot be re-bound to zipcode coated magnetic beads suggests the occurrence of an RBP complex which is disrupted during the isolation of these proteins.

The generation of polyclonal antibodies against several of the identified RNA binding proteins allowed for the assessment of *in vitro* binding specificity following initial capture, for the analysis of expression levels and, perhaps most importantly, for the determination of sub-cellular localization within developing rice endosperm tissue. Clearly, *in vitro* data alone is insufficient to allow meaningful conclusions to be made regarding the possible role of a protein in prolamine RNA localization. Therefore, it was important to analyze the *in vitro* capture data, protein expression data and *in vivo* localization data for each particular protein as a whole.

The hnRNPs RBP-D and RBP-I showed a very clear preference for binding to the 5'CDS prolamine zipcode over the control sequence (Figure 6), again suggesting that failure of these proteins to re-bind following desalting of eluted material was the result of complex dissociation. The specificity of RBP-A for the zipcode sequence was significant, but less than that seen for RBP-D and RBP-I, whilst the cp31, RBP-E, showed only poor specificity for the zipcode sequence.

Despite almost all of the hnRNPs for which localization data is presented here (RBP-A, RBP-D, RBP-I, RBP-N and RBP-Q) being present at high levels in nuclei, RBP-A and RBP-D were also found in close proximity to cortical ER membranes as well closely associated with the cytoskeleton. Interestingly, these two proteins were found to associate with different cytoskeletal elements; RBP-A with microtubules and RBP-D with actin microfilaments. This suggests that these proteins may perform different functions within the cytosol, with RBP-A perhaps having a more general role in RNA biology than RBP-D due to the higher specificity of the latter for the 5'CDS prolamine zipcode (Figure 6). Further experiments will be required to establish whether RBP-N and RBP-Q also co-localize with either actin filaments or microtubules.

In addition to the localization data summarized above, the level of RBP mRNA and protein expression observed during seed development also provides valuable insight. For example, RBP-A protein levels are elevated during the most rapid phase of storage protein accumulation (Figure 14), lending further support to the possibility that this protein plays a role in storage protein biosynthesis. Similarly, RBP-D protein levels are also maximal during the period of most rapid storage protein synthesis, but then drop during the final stages of seed maturation. Once again, the cp31, RBP-E, provided a contrasting result, since this protein is only detected during the earliest stages of seed development. Comparison of the available protein and

RNA expression data (Jiao et al., 2009; Figure 15 and 16) shows RBP-A to be highly expressed in many tissues, particularly in polarized and rapidly divided meristematic cells. In contrast, RBP-D is expressed only at low levels in all the tissues assessed by Jiao et al., (2009) in the Rice Atlas project. However, our own microarray and western data indicates that its RNA and proteins levels are comparatively much higher in developing rice endosperm (Figures 7 and 15). This suggests that RBP-D may have a specialized role in RNA localization during the most rapid period of storage protein accumulation.

Interestingly, RBP-A was not present on PB-ER membranes, a site where prolamine mRNA is enriched (Li et al., 1993; Choi et al., 2000). Without additional evidence, it is only possible to speculate about the possible role of this protein in prolamine mRNA transport and translation (see Chapter 3). However, our data suggests that RBP-A may play a general role in cytoplasmic RNA sorting and processing, perhaps associating with RNA either during and/or after transport. Additionally, the co-localization of RBP-A with microtubules suggests a possible role in anchoring RNA on microtubules ready for translation in the cortical region of the cell. Since RBP-A is present in both nuclear and cytoplasmic compartments, there must be a mechanism by which it can shuttle between them. It will be interesting to establish how this shuttling is regulated and whether RBP-A is bound to prolamine mRNA when it exits the nucleus.

The presence of RBP-D in small cytosolic particles which are closely associated with both actin filaments and prolamine protein bodies strongly supports a direct role for this protein in the transport of prolamine mRNA from the nucleus to the cortical ER. This possibility is strengthened by the fact that this protein shows excellent specificity for the prolamine zipcode

over the control RNA in our *in vitro* affinity capture assay (Figure 6) and the temporal and tissue-specific expression patterns described in the previous paragraph.

Further analysis of other candidate prolamine mRNA binding proteins by co-localization with cytoskeletal elements may prove fruitful since majority of the candidates are present as particles in the cytoplasm and all were isolated from cytoskeleton-enriched fractions. Since many of the candidate RBPs appear to be present in particles, this leads to the question of what other proteins or RNAs are being co-transported. It is possible, and highly likely, that these RBPs bind to multiple distinct RNA species. Since glutelin zipcode sequences have now been identified (Washida, Okita submitted), these will provide another direction from which to dissect the mechanism of mRNA localization in rice seeds. It is likely that both prolamine and glutelin RNA localization pathways involve the interaction of proteins having both general and specific RNA binding activities. Presumably, the same general RNA binding proteins may be required for both prolamine and glutelin RNA transport, whilst those with the more specific RNA binding activities may be unique to each pathway.

Acknowledgements

We thank Ron Nugen for fabricating high capacity magnetic bead capture apparatus, Gwen Anderson for handling rabbits during antibody generation, and members of the Franceschi Microscopy & Imaging Center for providing training and assistance in microscopy. Mass spectrometry analysis was done in collaboration with Julian P. Whitelegge at University of California. Data for RBP-A tissue specific expression was obtained from the RICEATLAS Project website. This work was supported by grants from the National Science Foundation.

References

Anderson, P., and Kedersha, N. (2009). RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nat Rev Mol Cell Biol.* **10**, 430-436.

Arn, E. A., Cha, B. J., Theurkauf, W. E., and Macdonald, P. M. (2003). Recognition of a bicoid mRNA localization signal by a protein complex containing Swallow, Nod, and RNA binding proteins. *Dev Cell.* **4**, 41-51.

Baluska F., Salaj, J., Mathur, J., Braun, M., Jasper, F., Samaj, J., Chua, N,-H, Barlow, P.W., and Volkmann, D. (2000). Root hair formation: F-actin-dependent tip growth is initiated by local assembly of profilin-supported F-actin meshworks accumulated within expansin-enriched bulges. *Dev. Biol.* **227**, 618-632.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. *Mol. Cell.* **2,** 437-445.

Bolognani, F., and Perrone-Bizzozero, N. I. (2008). RNA-protein interactions and control of mRNA stability in neurons. *J Neurosci Res.* **86**, 481-489.

Boyko, V., Ferralli, J., Ashby, J.A., Suslova, Schellenbaum, P., and Heinlein, M. (2000). Function of microtubules in intercellular transport of plant virus RNA. *Nat. Rev. Cell Biol*. **2**, 836-832.

Bullock, S. L., and Ish-Horowicz, D. (2001). Conserved signals and machinery for RNA transport in Drosophila oogenesis and embryogenesis. *Nature.* **414**, 611-616.

Bullock, S. L. (2007). Translocation of mRNAs by molecular motors: think complex? *Semin Cell Dev Biol.* **18**,194-201.

Cáceres, L., and Nilson, L. A. (2009). Translational repression of gurken mRNA in the Drosophila oocyte requires the hnRNP Squid in the nurse cells. *Dev Biol.* **326**, 327-334.

Choi, S. B., Wang, C., Muench, D. G., Ozawa, K., Franceschi, V. R., Wu, Y., and Okita, T.W. (2000). Messenger RNA targeting of rice seed storage proteins to specific ER subdomains. *Nature.* **407**, 765-767.

Chuong, S. D. X., Mullen, R. T., and Muench, D. G. (2002). Identification of a Rice RNA- and Microtubule-binding Protein as the Multifunctional Protein, a Peroxisomal Enzyme Involved in the b-Oxidation of Fatty Acids. *J. Biol. Chem.* **277**, 2419-2429.

Cohen, R. S. (2005). The role of membranes and membrane trafficking in RNA localization. *Biol Cell*. **97**, 5-18.

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

Crofts, A. J., Washida, H., Okita, T. W., Satoh, M., Ogawa, M., Kumamaru, T., and Satoh, H. (2005). The role of mRNA and protein sorting in seed storage protein synthesis, transport, and deposition. *Biochem. Cell Biol.* **83**, 728-737.

Davies, E., Fillingham, B. D., Oto, Y., and Abe, S. (1991). Evidence for the existence of cytoskeleton-bound polysomes in plants. *Cell Biol. Int. Rep.* **15**, 973-981.

Denessiouk, K. A., Rantanen, V. V., and Johnson, M. S. (2001). Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. *Proteins* **44**, 282-291.

Deshler, J. O., Highett, M. I., Abramson, T. and Schnapp, B. J. (1998). A highly conserved RNA-binding protein for cytoplasmic mRNA localization in vertebrates. *Curr. Biol.* **8**, 489–496.

Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nature Rev. Cell Biol.* **3,** 195-205.

Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., and Burd, C. G. (1993). hnRNP proteins and

the biogenesis of mRNA. *Annu Rev Biochem.* **62**, 289-321.

Du, T. G., Jellbauer, S., Müller, M., Schmid, M., Niessing, D., and Jansen, R. P. (2008). Nuclear transit of the RNA-binding protein She2 is required for translational control of localized ASH1 mRNA. *EMBO Rep.* **9**, 781-787.

Dubowy, J., and Macdonald, P. M. (1998). Localization of mRNAs to the oocyte is common in Drosophila ovaries. *Mech Dev.* **70**, 193-195.

Fedoroff, N.V. (2002). RNA-binding proteins in plants: the tip of an iceberg? *Curr. Opin. Plant Biol.* **5**, 452-459.

Gehring, N. H., Lamprinaki, S., Hentze, M. W., and Kulozik, A. E. (2009).The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol.* **7**, e1000120

Gerst, J.E. (2008). Message on the web: mRNA and ER co-trafficking. *Trends Cell Biol*. **18**, 68-76.

Gonsalvez, G. B., Little, J.L., and Long, R. M. (2004). ASH1 mRNA anchoring requires reorganization of the Myo4p-She3p-She2p transport complex. *J Biol Chem.* **279**, 46286-46294.

Guil, S., and Cáceres, J. F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol.* **14**, 591-596.

Hachet, O., and Ephrussi, A. (2004). Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature.* **428**, 959-963.

Hamada, S., Ishiyama, K., Choi, S. B., Wang, C., Singh, S., Kawai, N., Franceschi, V. R., and Okita, T. W. (2003a). The transport of prolamine RNAs to prolamine protein bodies in living rice endosperm cells. *Plant Cell* **10**, 2253-2264.

Hamada, S., Ishiyama, K., Sakulsingharoj, C., Choi, S. B., Wu, Y., Wang, C., Singh, S.,

Kawai, N., Messing, J. and Okita, T. W. (2003b). Dual Regulated RNA Transport Pathways to the Cortical Region in Developing Rice Endosperm. *Plant Cell* **15,** 2265-2272.

Hazelrigg, T. (2004). Lost in translation gets an oskar. *Dev. Cell* **6,** 611-613.

Hoek, K.S., Kidd, G. J., Carson, J. H., and Smith, R. (1998). hnRNP A2 Selectively Binds the Cytoplasmic Transport Sequence of Myelin Basic Protein mRNA. *Biochemistry* **37**, 7021-7029.

Honys, D., Re, K. D., Fecikov, J., Jedelsk, P., Nebes, Ov. J., Dobrev, P., and Apkov, V. R. (2009). Cytoskeleton-associated large RNP complexes in tobacco male gametophyte (EPPs) are associated with ribosomes and are involved in protein synthesis, processing and localization. *J Proteome Res.* **8**, 2015–2031

Hüttelmaier, S., Zenklusen, D., Lederer, M., Dictenberg, J., Lorenz, M., Meng, X., Bassell, G. J., Condeelis, J., and Singer RH. (2005). Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. *Nature* **438**, 512-515.

Iijima, M., Suzuki, M., Tanabe, A., Nishimura, A., and Yamada, M. (2006). Two motifs essential for nuclear import of the hnRNP A1 nucleocytoplasmic shuttling sequence M9 core. *FEBS Lett.* **580**, 1365-1370.

Im, K. H., Cosgrove, D. J., and Jones A. M. (2000). Subcellular localization of expansin mRNA in xylem cells. *Plant Physiol.* **123**, 463-470.

Jansen, R. P. (1999). RNA-cytoskeletal associations. *FASEB J.* **13,** 455-466.

Jansen, R. P. (2001). mRNA Localization: message on the move. *Nat. Rev. Mol. Cell Biol.* **2,** 247-256.

Jiao, Y., Tausta, S. L., Gandotra, N., Sun, N., Liu, T., Clay, N. K., Ceserani, T., Chen, M., Ma, L., Holford, M., Zhang, H. Y., Zhao, H., Deng, X. W., and Nelson, T. (2009). A

transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat Genet.* **41**, 258-263.

Jung, K. H., Dardick, C., Bartley, L. E., Cao, P., Phetsom, J., Canlas, P., Seo, Y. S., Shultz, M., Ouyang, S., Yuan, Q., Frank, B. C., Ly, E., Zheng, L., Jia, Y., Hsia, A. P., An, K., Chou, H. H., Rocke, D., Lee, G. C., Schnable, P. S., An, G., Buell, C. R., and Ronald PC. (2008). Refinement of light-responsive transcript lists using rice oligonucleotide arrays: evaluation of gene-redundancy. *PLoS One.* **3**, e3337.

Kim, W. T., and Okita, T.W. (1988). Structure, Expression, and Heterogeneity of the Rice Seed Prolamines. *Plant Physiol.* **88**, 649-655

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

Krause, K., and Krupinska, K. (2009). Nuclear regulators with a second home in organelles. *Trends Plant Sci.* **14**, 194-199.

Krecic, A. M., and Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* **11**, 363-371.

Kress, T. L., Yoon, Y. J., and Mowry, K. L. (2004). Nuclear RNP complex assembly initiates cytoplasmic RNA localization. *J. Cell Biol.* **165**, 203-211.

Krichevsky, A.M. and Kosik, K.S. (2001). Neuronal RNA granules: A link between RNA localization and stimulation-dependent translation. *Neuron* **32**, 683-696.

Krishnan, H. B., Franceschi, V. R. and Okita, T. W. (1986). Immunocytochemical studies on the role of the Golgi complex in protein body formation in rice seed. *Planta* **169**, 471-480.

Krishnan, H.B., White, J.A. and Pueppke, S.G. (1992). Characterization and localization of rice (*Oryza sativa* L.) seed globulins. *Plant Sci.* **81**, 1–11.

Lambermon, M. H., Fu, Y., Wieczorek Kirk, D. A., Dupasquier, M., Filipowicz, W., and Lambert, D., and Draper, D. E. (2007). Effects of osmolytes on RNA secondary and tertiary structure stabilities and RNA-Mg2+ interactions. *J Mol Biol.* **370**, 993-1005.

Landers, S. M., Gallas, M. R., Little, J., and Long, R. M. (2009). She3p possesses a novel activity required for *ASH1* mRNA localization in *Saccharomyces cerevisiae*. *Eukaryot. Cell*. [Epub ahead of print].

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T. R., Tomancak, P., and Krause, H. M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell.* **131**,174-187.

Lee, S. M., Kim, J. H., Cho, E. J. (2009). A nucleocytoplasmic malate dehydrogenase regulates p53 transcriptional activity in response to metabolic stress. *Cell Death Differ.* **16**, 738-748.

Li, X., Franceschi, V. R. and Okita, T. W. (1993). Segregation of storage protein mRNAs on the rough endoplasmic reticulum membranes of rice endosperm cells. *Cell* **72**, 869-879.

Li, Y. Q. and Sugiura M. (1991). Nucleic acid-binding specificities of tobacco chloroplast ribonucleoproteins. *Nucleic Acids Res*. **19**, 2893-2896.

López de Heredia, M. and Jansen, R. P. (2004). mRNA localization and the cytoskeleton. *Curr Opin Cell Biol.* **16**, 80-85.

Lopez de Heredia, M., and Jansen, R. P. (2004). mRNA localization and the cytoskeleton. Curr. Opin. *Cell Biol.* **16**, 80-85.

Ma, L., Xie, B., Hong, Z., Verma, D.P., and Zhang, Z. (2008). A novel RNA-binding protein associated with cell plate formation. Plant Physiol. **148**, 223-234.

Maris, C., Dominguez, C. and Allain, F. H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118-2131.

Martin, K. C., and Ephrussi, A. (2009) mRNA localization: gene expression in the spatial dimension. Cell. **136**,719-730

Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J. F., Revil, T. and Chabot, B. (2007). hnRNP proteins and splicing control. *Adv Exp Med Biol.* **623**, 123-147.

Matthews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999). Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure *J. Mol. Biol.* **288**, 911-940.

Messitt, T. J., Gagnon, J. A., Kreiling, J. A., Pratt, C. A, Yoon, Y. J. and Mowry, K. L. (2008). Multiple kinesin motors coordinate cytoplasmic RNA transport on a subpopulation of microtubules in Xenopus oocytes. *Dev Cell.* **15**, 426-436.

Muench, D. G., Chuong, S. D., Franceschi, V. R., and Okita, T. W. (2000). Developing prolamine protein bodies are associated with the cortical cytoskeleton in rice endosperm cells. *Planta.* **211**, 227-238.

Muench, D. G. and Mullen, R. T. (2003). Peroxisome dynamics in plant cells: a role for the cytoskeleton. *Plant Sci.* **164**, 307-315.

Muench, D. G., Wu, Y., Coughlan, S. J., and Okita, T. W. (1998). Evidence for a Cytoskeleton-Associated Binding Site Involved in Prolamine mRNA Localization to the Protein Bodies in Rice Endosperm Tissue. *Plant Physiol.* **116**, 559-569.

Müller, M., Heuck, A., and Niessing, D. (2007). Directional mRNA transport in eukaryotes: lessons from yeast. *Cell Mol. Life Sci.* **64**, 171-180.

Nakamura, T., Ohta, M., Sugiura, M., and Sugita, M. (2001). Chloroplast ribonucleoproteins function as a stabilizing factor of ribosome-free mRNAs in the stroma. *J Biol Chem.* **276**,147-152.

Okita, T. W. and Choi, S. B. (2002). mRNA localization in plants: targeting to the cell's cortical region and beyond. *Curr. Opin. Plant Biol.* **5**, 553-333.

Paquin, N., and Chartrand, P. (2008). Local regulation of mRNA translation: new insights from the bud. *Trends Cell Biol.* **18**, 105-111.

Percipalle, P., Raju, C. S., and Fukuda, N. (2009). Actin-associated hnRNP proteins as transacting factors in the control of mRNA transport and localization. *RNA Biol.* **6** [Epub ahead of print]

Pioli, P. A., Hamilton, B. J., Connolly, J. E., Brewer, G., and Rigby, W. F. (2002). Lactate dehydrogenase is an AU-rich element-binding protein that directly interacts with AUF1. *J. Biol. Chem.* **277**, 35735-35745.

Pyhtila, B., Zheng, T., Lager, P. J., Keene, J. D., Reedy, M. C., and Nicchitta C. V. (2008). Signal sequence- and translation-independent mRNA localization to the endoplasmic reticulum. *RNA.***14**, 445-453.

Ross, A. F., Oleynikov, Y., Kislauskis, E. H., Tenja, K. L. and Singer, R. H. (1997). Characterization of a *β-actin* mRNA zipcode binding protein. *Mol. Cell. Biol.* **17**, 2158-2165.

Samaj, J., Salaj, J., Obert, B., Baluska, F., Menzel, D., and Volkmann D. (2008). Calreticulin mRNA and protein are localized to protein bodies in storage maize callus cells. *Plant Cell Rep.* **27**, 231-239.

Sami-Subbu, R., Choi, S. B., Wu, Y., Wang, C., and Okita, T. W. (2001). Identification of a cytoskeleton-associated 120 kDa RNA-binding protein in developing rice seeds. *Plant Mol Biol.* **46**,79-88.

Sasaki, K., Kim, M. H., and Imai, R. (2007). Arabidopsis COLD SHOCK DOMAIN PROTEIN2 is a RNA chaperone that is regulated by cold and developmental signals. *Biochem*

Biophys Res Commun. **364**, 633-638.

Sawicka, K., Bushell, M., Spriggs, K. A. and Willis, A. E. (2008). Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochem Soc Trans.* **36**, 641-647.

Shamoo, Y., Abdul-Manan, N. and Williams, K. R. (1995). Multiple RNA binding domains (RBD's) just don't add up. *Nucleic Acids Res.* **23**, 725-728.

Siomi, H., and Dreyfuss, G. (1995). A Nuclear Localization Domain in the hnRNP A1 Protein. *J. Cell Biol.* **129**, 551-560.

Smith, M. C., Cook, J. A., Smanik, P. A., Wauklchik, M., and Kasher, M. S. (1992) Cobalt affinity chromatography using exchange-inert coupling. *Methods: A companion to methods in enzymology* **4**, 73-78.

St Johnston, D. (2005). Moving messages: the intracellular localization of mRNAs. *Nat. Rev. Mol. Cell Biol.* **6**, 363-375.

Stebbings, H. (2001) Cytoskeleton-dependent transport and localization of mRNA. *Int. Rev. Cyto.* **211**, 1-31.

Su, P. H., Yu, S. M., and Chen, C. S. (2001). Spatial and temporal expression of a rice prolamin gene *RP5* promoter intransgenic tobacco and rice. *J.Plant Physiol.* **158**, 247-254.

Svoboda, P., and Cara, A.D. (2006). Hairpin RNA: a secondary structure of primary importance. *Cell Mol Life Sci.* **63**, 901-908.

Tillich, M., Hardel, S. L., Kupsch, C., Armbruster, U., Delannoy, E., Gualberto, J. M., Lehwark, P., Leister, D., Small, I. D., and Schmitz-Linneweber C. (2009). Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A.* **106**, 6002-6007.

Washida, H., Kaneko, S., Crofts, N. , Sugino, A. , Wang, C., and Okita, T. W. (2009). Identification of *cis*-localization elements that target glutelin RNAs to a specific subdomain of the cortical endoplasmic reticulum in rice endosperm cells. *Plant Cell Physiol.*(Accepted)

Washida, H., Sugino, A., Messing, J., Esen, A., and Okita, T.W. (2004). Asymmetric localization of seed storage protein RNAs to distinct subdomains of the endoplasmic reticulum in developing maize endosperm cells. *Plant Cell Physiol*. **45**, 1830-1837.

Wilhelm, J. E., and Vale, R. D. (1993) RNA on the Move: The mRNA Localization Pathway. *J. Cell Biol.* **123,** 269-274.

Yamagata, H., and Tanaka, K. (1986) The Site of Synthesis and Accumulation of Rice Storage Proteins. *Plant Cell Physiol*. **27**, 135-145.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415.

Figure legends

Figure 1. Prolamine RNA zipcodes; position, structure and sequence

(A) Schematic representation of the prolamine gene showing the two RNA localization signals, both of which are required for faithful transport to the PB-ER. Sequences were selected from both zipcode regions based on their homology and predicted structure for use in RNA binding protein capture experiments. Proteins were captured with either the 5'CDS sequence, or with a control sequence randomly selected from a non-zipcode region within the open reading frame. The lengths of RNA baits are indicated in addition to the signal peptide (SP), open reading frame (ORF) and both 3' and 5' untranslated regions (UTR).

(B) Predicted lowest free energy structure of full length prolamine mRNA using mfold. Sequence homology between the 5'CDS and 3'UTR zipcodes is shown in the inset in the lower left hand corner. Notice that the 5'CDS and 3'UTR zipcodes fall at opposite ends of an essentially linear structure.

(C) Conservation of sequence and structure between prolamine 5'CDS and 3'UTR zipcodes. Shown are the lowest energy structures predicted by mfold for the two prolamine mRNA zipcodes. Notice that both structures have a single stem containing a run of 8 or 9 nucleotides and a loop containing either GUUUUUA (5'CDS) or GUUUUA (3'UTR).The regions outlined by solid line were used for the bait for RNA isolation and share high homology between them. Dotted lines indicate the full extent of the zipcode previously identified by deletion analysis.

Figure 2. Capture of RNA binding proteins using biotinylated prolamine zipcode RNA

(A) Comparison of RNA binding proteins captured in the absence of RNA (-), or by either the control (CON), non-zipcode sequence, or the prolamine 5'CDS zipcode bait RNA (CDS).

(B) A multi-step salt elution (0.2 M NaCl, 0.5 M NaCl and 0.3M $MgCl_2$), indicating the relatively high strength of the interactions with the zipcode bait RNA.

Figure 3. *In vitro* UV cross-linking assay of captured RBPs

(A) *In vitro* UV cross-linking assays of captured RBPs were performed using prolamine zipcodes as probe with increasing concentrations of KCl, $MgCl_2$ or EDTA. Note that RNA binding activity decreases in the presence of high salt and magnesium ions. RNA binding activity was competed for by tRNA but not by heparin. The silver stained gel on the left shows the captured proteins used for the experiments. MW: molecular weight in kDa.

(B) Competition assays were performed to assess substrate preference. Upper panels: captured proteins were incubated with 0, 0.5, 1 or 2 μg of nucleotide homopolymers, Poly(A), Poly(U), Poly(G), Poly(C), Poly(A・U) or Poly(I・C) prior to addition of probe. Note that captured proteins show a strong preference for binding to poly(G). Lower panels: captured proteins were incubated with 0, 0.01, 0.1 or 1 μg of unlabeled prolamine (Prol.) zipcodes, prolamine 3' zipcode, prolamine CDS, glutelin (Glu.) CDS, glutelin 3' UTR or vector sequence from pBSIIKS$^-$ as indicated.

Figure 4. Phylogram of candidate RBPs

Phylogram showing the homology between the identified RNA binding proteins using ClustalW analysis of amino acid sequences.

Figure 5. cDNA cloning, protein expression and generation of affinity purified antibodies

cDNA clones encoding identified RBPs were either requested or generated by reverse transcription from total RNA using oligo(dT) followed by PCR using gene-specific primers. Cloned cDNAs were expressed in pET30 vector to generate fusion proteins with an N-terminal His tag. Following initial purification by immobilized metal affinity chromatography (IMAC), pure protein was obtained by protein gel purification and electro-elution. Polyclonal rabbit antibodies were generated and affinity purified using an antigen affinity column prepared using irreversibly oxidized Co(III) coupled to iminodiacetic acid (IDA) resin.

Figure 6. Specificity of RBPs for prolamine zipcode and control bait sequences

Following protein capture using either control (CON) or zipcode (CDS) oligos, eluted proteins were analyzed by protein gel blotting using the newly generated RBP antibodies as indicated. The left panel shows a higher loading of captured material (from Figure 3) for the purpose of clarity. Notice that whilst many of the proteins show specificity for the zipcode over the control oligo, RBP-D and RBP-I exhibit the strongest preference.

Figure 7. Protein expression analysis of candidate RBPs in developing seeds, mature dry seeds, leaf and root material

50 μg of total protein extract was separated by SDS-PAGE and analyzed by western blotting using antibodies against RBP-A (hnRNP), RBP-B (cp29), RBP-C (Malate dehydrogenase: MDH), RBP-D (hnRNP) or RBP-E (cp31). Numbers on the left indicate molecular weights in kDa. Note that majority of antibodies reacted with a single band in developing rice seed extract.

Figure 8. Domain structure of RBP-A, RBP-D and RBP-I

The domain structures of RBP-A, RBP-D and RBP-I are shown as predicted from their protein sequences using the NCBI protein BLAST algorithm. There are two RNA recognition motifs (RRM) at their N-terminus. Note that the N-terminal RRM of RBP-D is truncated.

Figure 9. Genomic DNA analysis for RBP-D

(A) Upper panel: Homologs of RBP-D were shown by comparing Arabidopsis, maize and rice genomic DNA using http://ricegaas.dna.affrc.go.jp/. The red box indicates that rice gene potentially has two exons upstream of the AK059225 cDNA. Bottom panel: alignment of the

AK059225 cDNA and the theoretical full length cDNA predicted for RBP-D using ClustlW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Translation of the newly predicted gene resulted in the addition of 122 extra amino acids.

(B) Genomic DNA sequence of RBP-D from BAC clone (AP005456). The shaded region indicates putative coding sequence, the start and stop codons are highlighted. The BamHI which can be used for cloning is highlighted in yellow.

(C) Upper panel: Sequence of a newly isolated full length RBP-D cDNA isolated from kitaake developing rice seeds. The red region indicates sequence shared with the AK059225 cDNA present in the KOME (Knowledge-based Oryza Molecular Biological Encyclopedia) database. The blue region represents confirmed sequence within upstream exons and the gray region is sequence predicted from genomic DNA. Gray region is sequence unconfirmed, blue region indicates sequence has confirmed. The bottom panel indicates the domain structures predicted for truncated and theoretical full length RBP-D cDNAs. Note that the sequence of the entire first RRM has been confirmed by sequencing.

Figure 10. RBP-A is present on cisternal ER membranes and co-localizes with microtubules

(A) Localization of RBP-A (green) and nucleus (red) and using affinity purified anti-RBP-A antibody and the nucleic acid binding dye, propidium iodide. Shown is a single slice image. Scale bar = 5 μm. Note that in the middle of the cell, majority of RBP-A is localized to the nucleus and some small particles indicated with asterisk. Small particles maybe RNA transport particles since both of signals are co-localized.

(B) Localization of RBP-A (red) and PB-ER (green) and using affinity purified anti-RBP-A antibody and the ER vital dye $DiOC_6$. Shown is a projection image from a stack of ten adjacent 1 μm optical sections. Scale bar = 5 μm.

(C) Higher magnification single channel and merged images of (B) showing that RBP-A signals are non-uniformly distributed on cisternal ER membranes and that these signals do not merge with prolamine protein bodies.

(D) Single channel and merged projection images showing co-localization of RBP-A (red) and tubulin (green) using anti-RBP-A and anti-tubulin antibodies. RBP-A is present in nuclei (N) at high concentrations, but is also associated with ER membranes and co-localizes with microtubules.

Figure 11. Co-localization of RBP-D with actin filaments and protein bodies

(A) Indirect immunofluorescence localization experiment showing that although primarily nuclear (N), RBP-D is also present in discrete particles (red) which are closely associated with actin filaments (green). Scale bar = 5 μm, N = nucleus. The white rectangle defines the area shown at higher magnification in the right hand panel. Arrows indicate the localization of RBP-D particles which are in very close proximity to a single actin microfilament.

(B) RBP-D (green) can be seen in close proximity to the rhodamine-stained prolamine protein bodies (red). The white rectangle defines the area shown at higher magnification in the right hand panel. Scale bar = 5 μm, N = nucleus.

Figure 12. RBP-I localizes in nuclei

(A) Localization of RBP-I (green) and tubulin (red) using affinity purified anti-RBP-I and anti-tubulin antibodies shows RBP-I is localized in the nucleus. Shown is a projection image. Scale bar = 5 μm.

(B) RBP-I (green) can only be seen in the nucleus and do not merge with rhodamine-stained prolamine protein bodies (red). Shown is a projection image. Scale bar = 5 μm.

Figure 13. Localization of RBP-B, RBP-E, RBP-F, RBP-N, and RBP-Q in developing rice endosperm cells

(A) Localization of RBP-B (cp29 in green) using affinity purified anti-RBP-B antibody and rhodamine-stained prolamine protein bodies (red). Scale bar = 5 μm. Note that RBP-B is localized as particles tracing irregularly shaped balloon structure.

(B) Localization of RBP-E (cp31 in green) using affinity purified anti-RBP-E antibody and rhodamine-stained prolamine protein bodies (red). Scale bar = 5 μm. Note that RBP-E is localized as particles and some of them traces irregularly shaped balloon structure.

(C) Localization of RBP-F (cp31 in green) using affinity purified anti-RBP-F antibody and rhodamine-stained prolamine protein bodies (red). Scale bar = 5 μm. Note that RBP-F is localized as particles.

(D) Localization of RBP-N (GRP: glycine rich protein in green) using affinity purified anti-RBP-N antibody and rhodamine-stained prolamine protein bodies (red). Scale bar = 5 μm. Note that RBP-N is localized as particles or thread like structure.

(E) Localization of RBP-Q (hnRNP in green) using affinity purified anti-RBP-Q antibody and rhodamine-staining of prolamine protein bodies (red). Scale bar = 5 μm. Note that RBP-Q is

localized as particles present on cisternal ER membrane which are lightly stained by rhodamine. The highlighted region is enlarged in (F).

(F) Enlargement of area indicated in (E).

Figure 14. RBP-A expression correlates with storage protein accumulation

Developing rice seed extract of odd days from 1 to 23 day after flowering (DAF) were analyzed by western blot using anti-RBP-A (hnRNP), anti-RBP-E (cp31), anti-RBP-I (hnRNP), anti-glutelin (storage protein) or anti-prolamine (storage protein). Each well contains 1/20 seed worth total proteins. Notice that RBP-A expression pattern has strong correlation to glutelin and prolamine accumulation patterns. Whilst RBP-E is expressed before storage protein start accumulating and RBP-I is present constant amount throughout the seed development.

Figure 15. RNA expression levels of RBPs in developing rice seeds

Expression levels of RBPs were analyzed by microarray analysis using probes generate with total RNA extracted from10-14 DAF developing rice seeds. Note that RBP-A (hnRNP), RBP-H (cold-shock protein), RBP-O (ribosomal L3) are highly expressed. Experiments are from 9 replicates including 6 biological replicates. Average of total intensity values were used as factor to calculate target's signal intensities and their average value were calculated. Error bar used is standard error.

Figure 16. RBP-A is highly expressed in rapidly dividing cells

RNA expression levels of RBP-A in different tissue samples as labeled in the graph were analyzed using experimental data sets available through website

http://bioinformatics.med.yale.edu/riceatlas/overview.jspx as described by Jiao et al., 2009. Note that RBP-A is highly expressed in the actively dividing cells.

Table legends

Table 1. Identification of candidate RBPs by tandem mass spectrometry

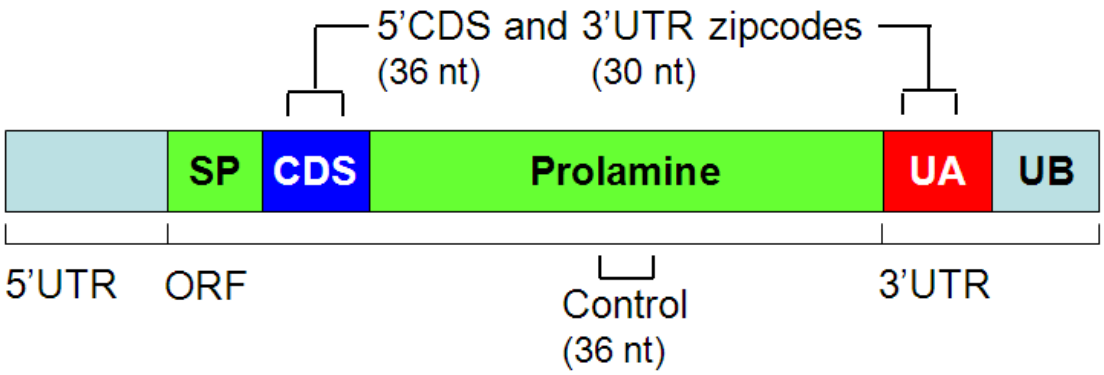
RNA binding proteins captured using prolamine zipcode bait RNA in a scaled-up experiment similar to that shown in Figure 2A were subjected to in gel trypsin digestion followed by LC-MS/MS analysis. Eighteen proteins were identified using Mascot software and are listed in the table together with their Genbank accession numbers (gi), cDNA accession number (AK), homology, predicted molecular weight and isoelectric point. Predicted domains are also given.

Table 2. List of peptides identified by tandem mass spectrometry

Peptides identified by tandem mass spectrometry are shown along with their identification number (AK), peptide coverage (%) and start position of peptides in amino acids

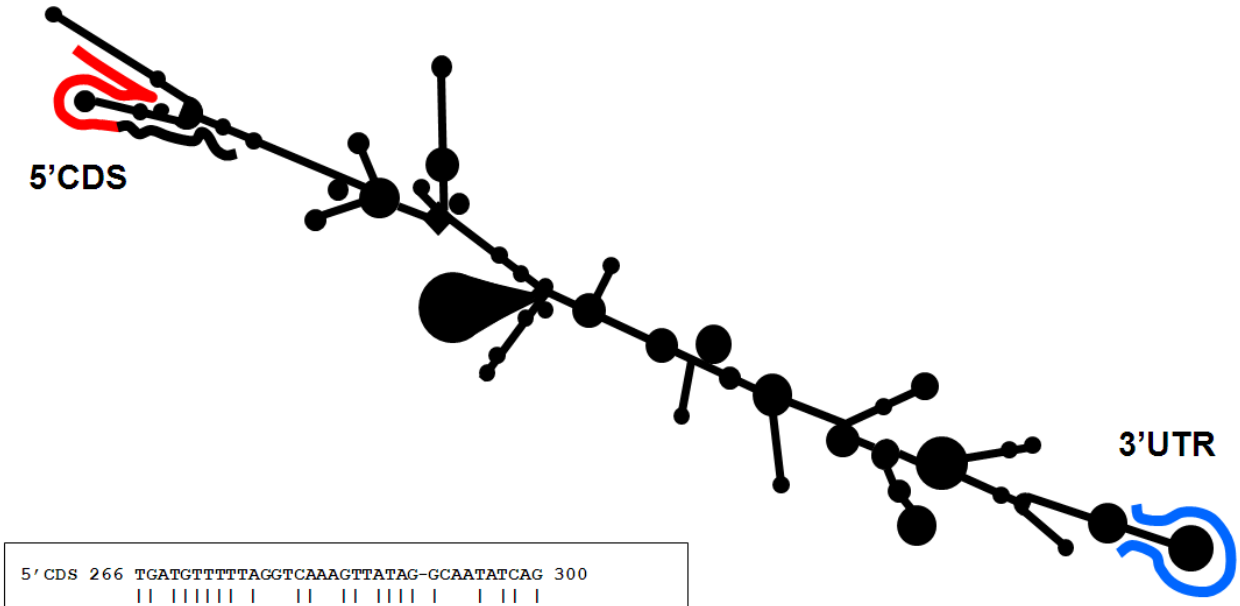
Figure 1. Prolamine zipcodes; position, structure and sequence

(A)



UTR: untranslated region SP: signal peptide ORF: open reading frame

(B)




```
5'CDS 266 TGATGTTTTTAGGTCAAAGTTATAG-GCAATATCAG 300
          || |||||| |   ||  || |||| |   | || |
3'UTR 652 TG-TGTTTTAA---CA--GT-ATAGTGG--T-TC-G 676

5'CDS 301 CTGCAGTCGCCTGTCC-T 317
            | |||    |     |
3'UTR 678 --GAAGT----TAAAAAT 689
```

(C)

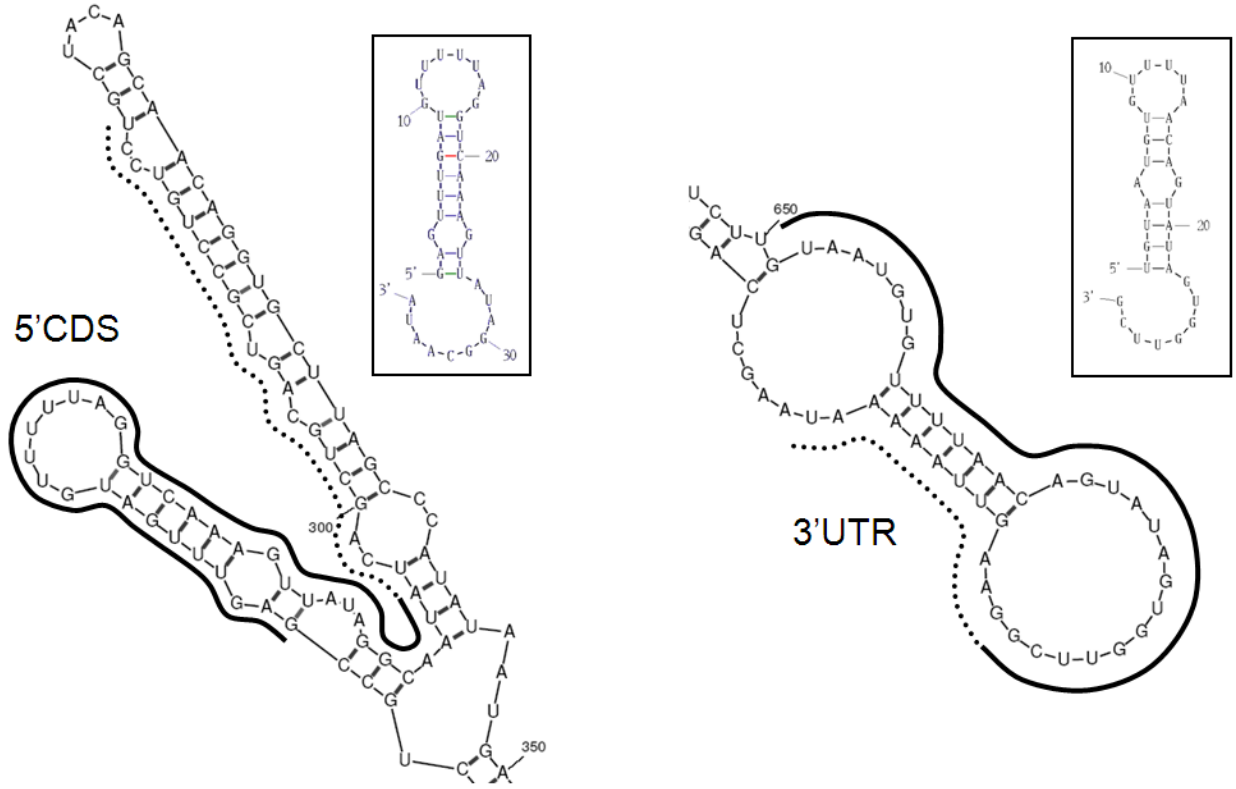
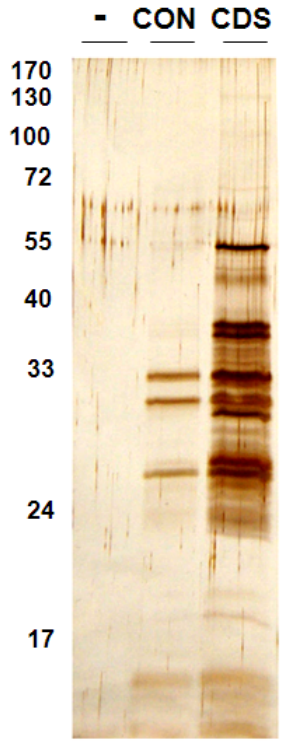


Figure 2. Capture of RNA binding proteins using prolamine zipcode RNA

(A)



(B)

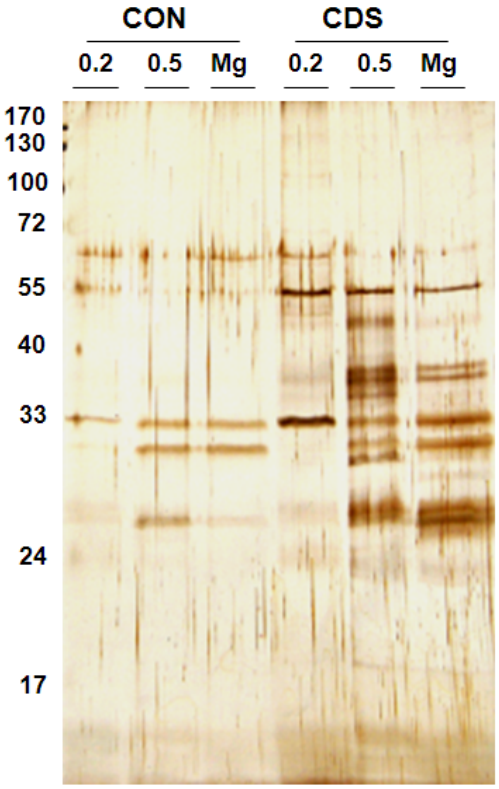
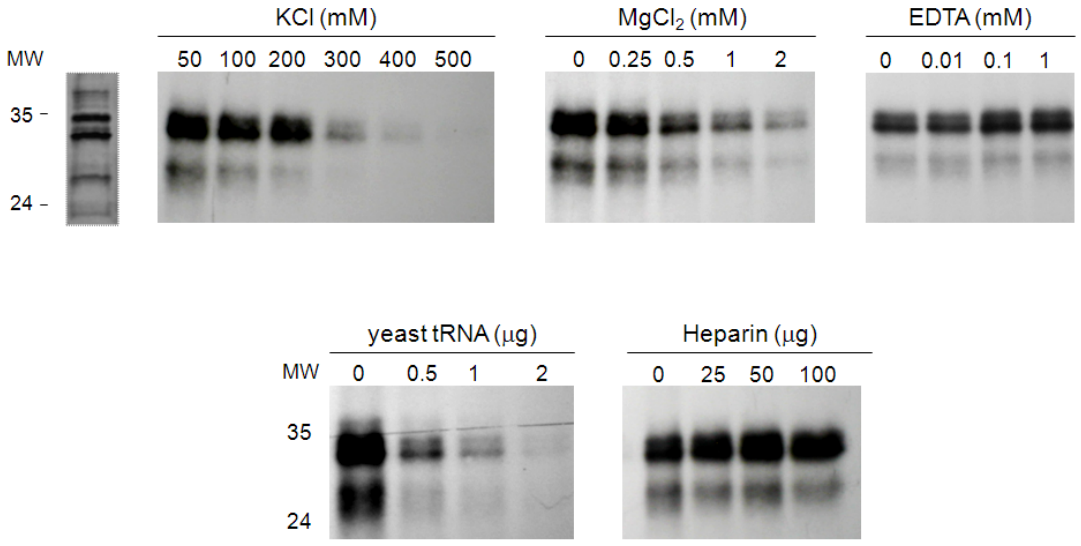


Figure 3. *In vitro* UV cross-linking assay of captured RBPs

(A)



(B)

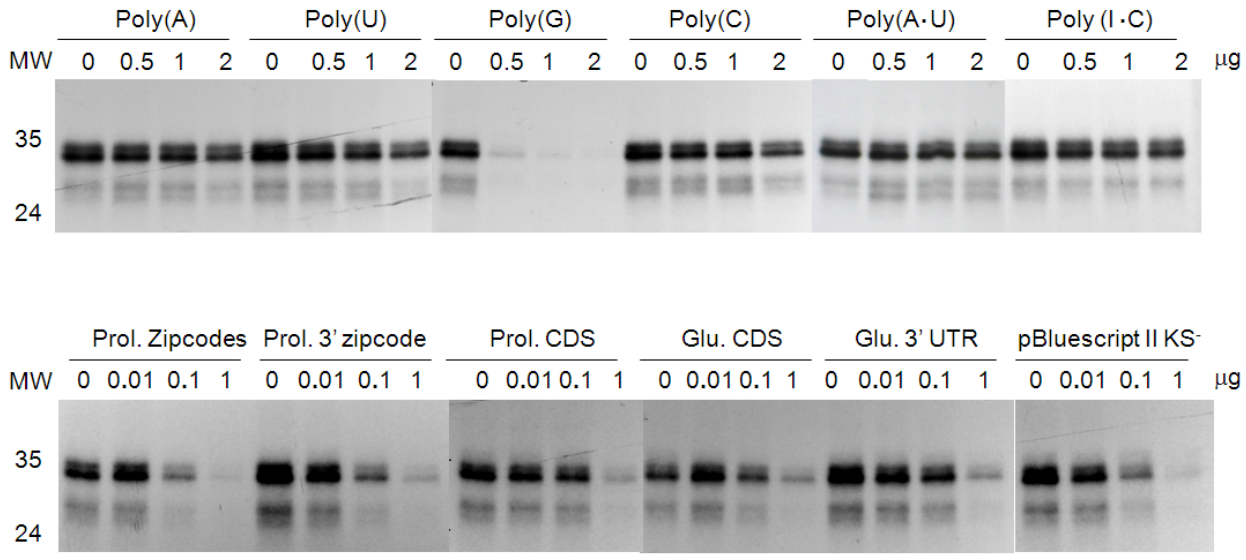


Figure 4. Phylogram of candidate RBPs

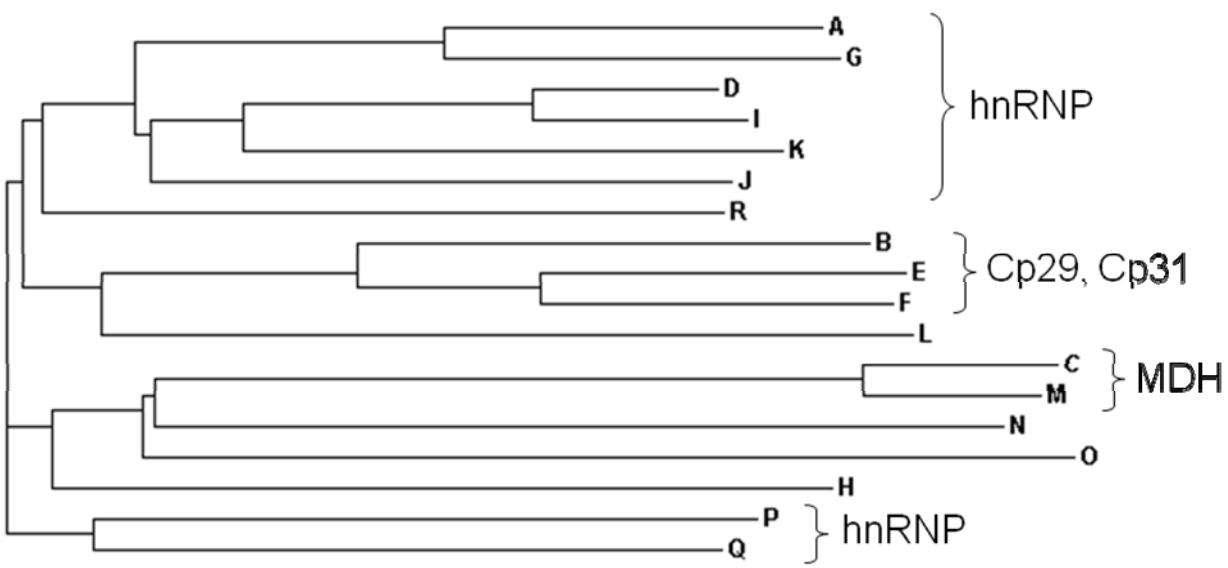


Figure 5. cDNA cloning, protein expression and generation of affinity purified antibodies

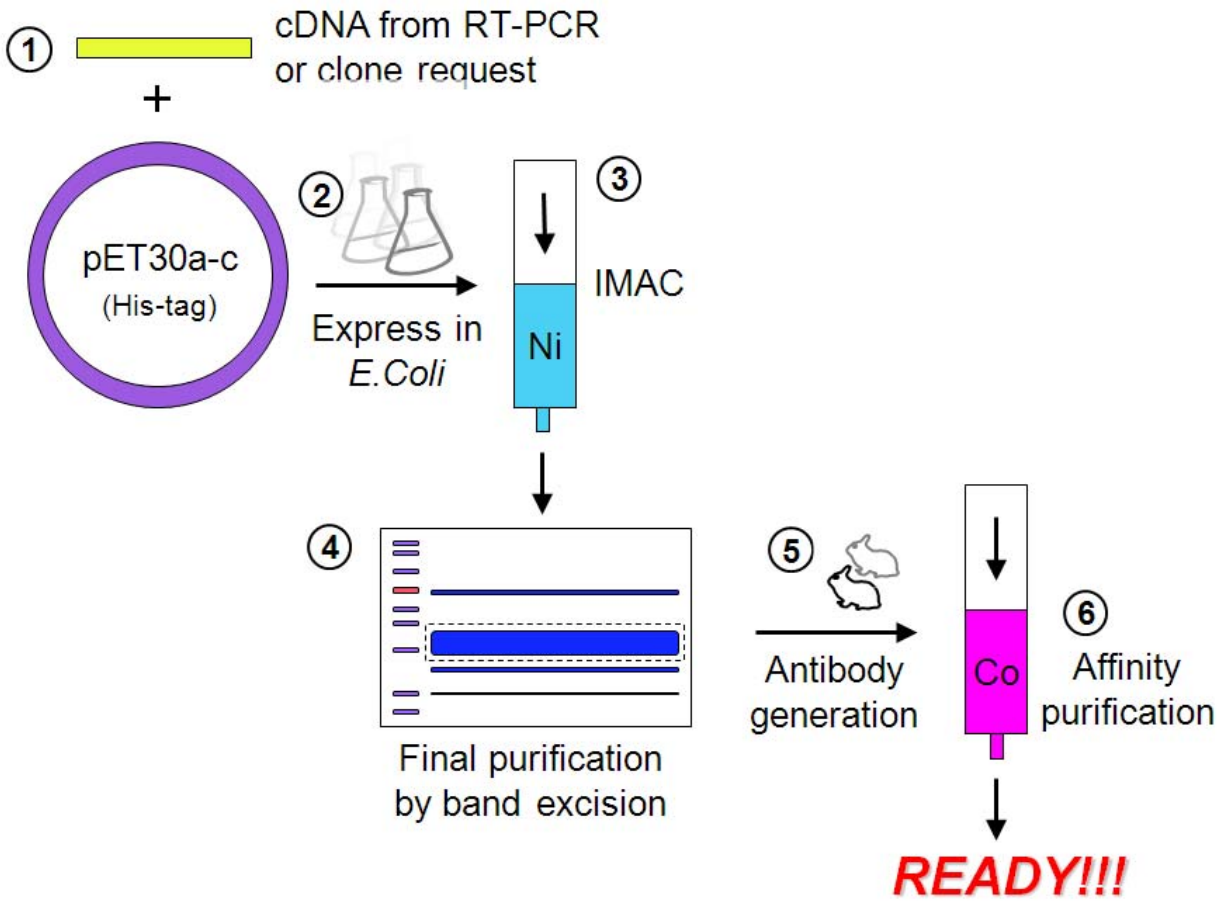


Figure 6. Specificity of RBPs for prolamine zipcode and control bait sequences

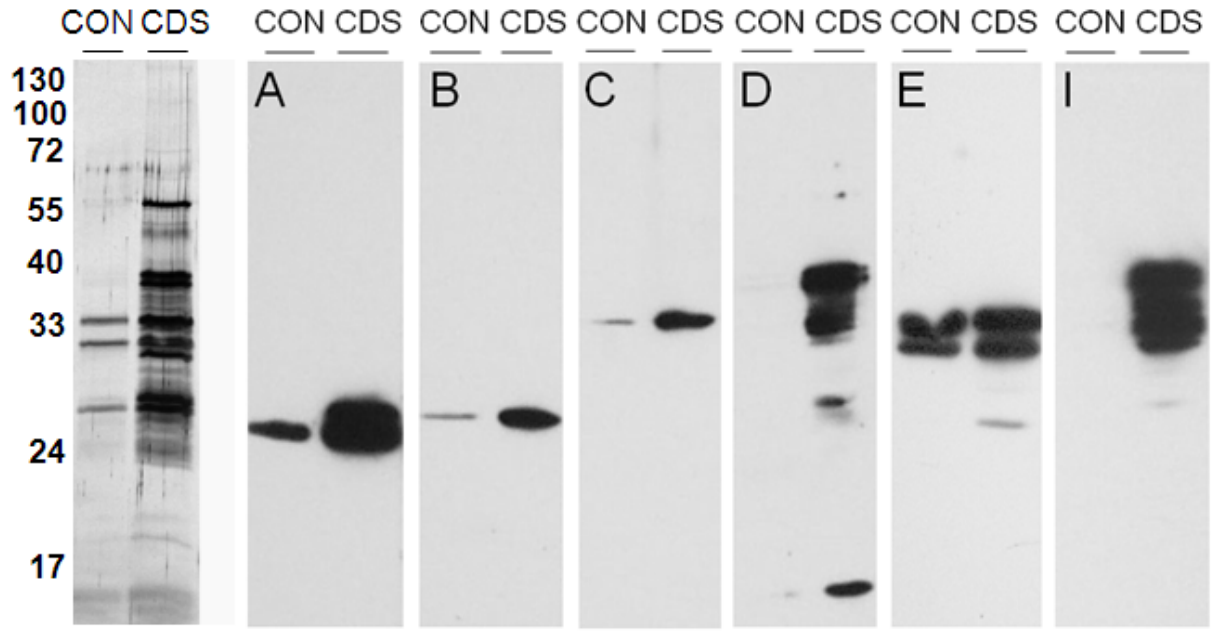


Figure 7. Protein expression analysis of candidate RBPs in developing seeds, mature seeds, leaf and root material

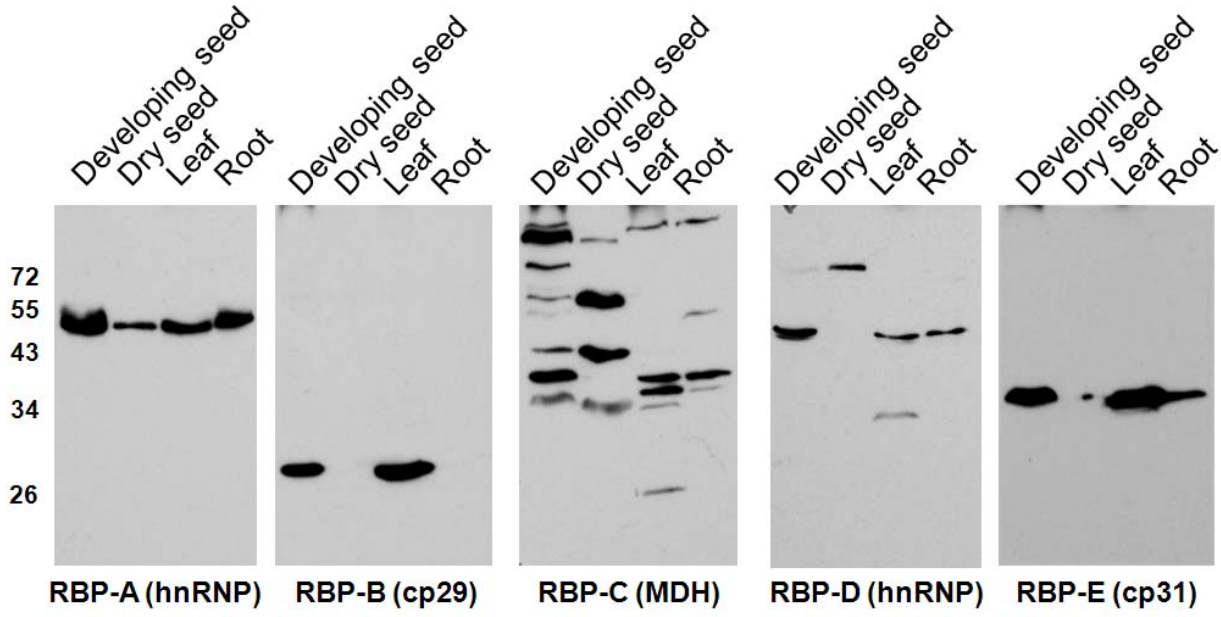


Figure 8. Domain structure of RBP-A, RBP-D and RBP-I

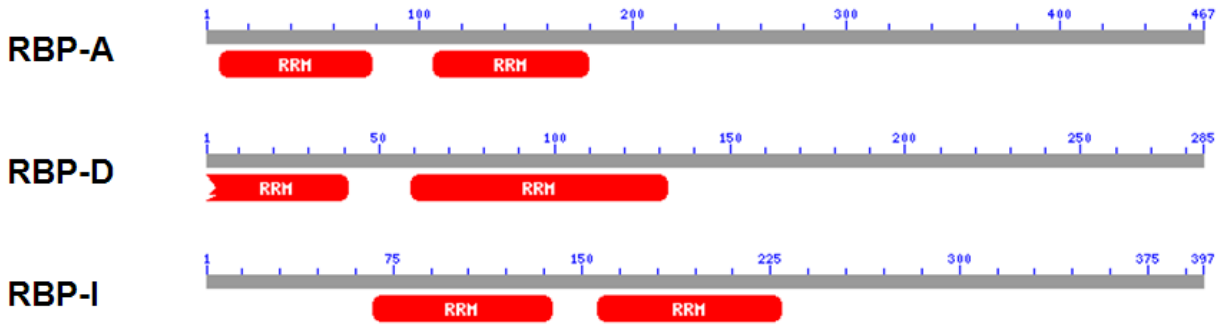


Figure 9. Genomic DNA analysis for RBP-D

(A)


```
RBP-D           MAGYGEENQNGMNGYEEEEEEEEVEEVEEEVEEEEEEEEEEGADATAAAADAAEEVAEE 60
AK059225        ------------------------------------------------------------


RBP-D           RRGGGGEVEGVGNGEEAGRTAGGGEGGDSSGKIFVGGVAWETTEESFTKHFEKYGAISDS 120
AK059225        ------------------------------------------------------------


RBP-D           VIMKDKHTKMPRGFGFVTFSDPSVIDKVLQDEHTIDGRTVEVKRTVPREEMSSKDGPKTR 180
AK059225        --MKDKHTKMPRGFGFVTFSDPSVIDKVLQDEHTIDGRTVEVKRTVPREEMSSKDGPKTR 58
                  ************************************************************


RBP-D           KIFVGGIPPSLTEDKLKEHFSSYGKVVEHQIMLDHGTGRSRGFGFVTFENEDAVERVMSE 240
AK059225        KIFVGGIPPSLTEDKLKEHFSSYGKVVEHQIMLDHGTGRSRGFGFVTFENEDAVERVMSE 118
                ************************************************************


RBP-D           GRMHDLAGKQVEIKKAEPKKPGGGDSSSNGRHSHGSGGGHRSSYRGSGGGNSGSSSSGGY 300
AK059225        GRMHDLAGKQVEIKKAEPKKPGGGDSSSNGRHSHGSGGGHRSSYRGSGGGNSGSSSSGGY 178
                ************************************************************


RBP-D           GGYGGGYRSAAAAYYGSTGYAGYGRGYGYGGNPAFGSGFGSGYGGSMYGGPYGAYGAYGG 360
AK059225        GGYGGGYRSAAAAYYGSTGYAGYGRGYGYGGNPAFGSGFGSGYGGSMYGGPYGAYGAYGG 238
                ************************************************************


RBP-D           AYGGGGAYGAPGGYGAGGYGAYGGAGGMGGGGSTSGRGSSRYHPYGK 407
AK059225        AYGGGGAYGAPGGYGAGGYGAYGGAGGMGGGGSTSGRGSSRYHPYGK 285
                ************************************************
```

RBP-D: translated full length cDNA
AK059225: translated cDNA clone

(B)

Genomic DNA Sequence from AP005456 (chr 6)

```
27841 tagtggtggt ggtggtgggg agggatggcg gggtacgggg aggagaacca gaacggcatg
27901 aacgggtacg aggaggagga ggaggaagag gaggtggagg aggtcgagga agaggtggag
27961 gaggaggagg aggaagagga ggaggaggag ggggctgatg ccaccgccgc ggctgccgat
28021 gcggcggagg aggttgcaga ggagaggcga ggcggaggtg gagaagtgga gggcgtcggc
28081 aatggcgagg aagcggggag gactgcgggt ggtggagagg gcggtgattc gtcggggtga
28141 gtcccctttcc tttcgatgtt cggtctgttt tggttggatg gatgaagttg gctgaatttt
28201 gttggttttg tttgatctgt gtgcctctgc aggaaaattt ttgttggagg tgtagcctgg
28261 gagacaactg aaggtagagc ggatggattc tttttggggg atttttgtta aacatgagaa
28321 ttctagttgt gcaaagcaac gtgtttaaga gatgattttt agaggttaat actagtagca
28381 tatagtagat gtgaatgaag tagcgtcctt cagtagctag caacctaaag aagaatttag
28441 ggaagagtgt ttctctgaaa ggataggttt cttgttgttg tgcttaataa cttaggatag
28501 caggactaac tgaactaaag tttgtttttg cttaggttga gatgtgaaaa ggaatctagc
28561 aatctgcaag ccatagaact taaatgctac tttgtaatac ttttaagaag caatcttgta
28621 cctttgtttt gtgtaggtgt attttcacat tgaaatgtag aagcaggttg aacagcttgg
28681 gctaacttag ttaaccaaat ttagaaactt gtccattggt aacttcatta ggaattttaa
28741 cttttaagca tagatgctat gcgaatgctt caatcatgat taagctgcat tgcttttaat
28801 catgttgttc catctagtcg ttgagacatg ttgtgcatgt ctgttgtcta atgcatacag
28861 gggtatattt atctttttgt ttggttgatt atgtttgcac tgtttttttt tctgatttgg
28921 attttccaat agtaaatact aaaataacat gtatctgctg tattgatggt catgtgtgtt
28981 gaattctttt gtttggaaga taaaatgctt atattgcatt tggtttgaat tgggaaatgg
29041 gagttacttt tgattaaata gcaattgagc aggagctgac tcagtatgac ctggcacatt
29101 tgctgttgtt atttgggtcc aaatgcttcc tgctattttc ttttctcttg agttttcagt
29161 aattgagcag gagctgactc agtatgacct ggcacatttg ctgttgttat ttgggtccaa
29221 atgcttcctg ctatttttctt ttctcttgag ttttcagtga atgcactatt ctgtttattt
29281 acattgtgca tgatcttatt gttagtgcgt ttggtgctgg gttctagttt gtactattta
29341 atttcattgc tgtcggcaga taggctaaac atgttcatat gattttaatc tggaccctaa
29401 tttcacctgg aaaaatgttc atattattgt aatcccatca atattgttag ataaggcttt
29461 gatttacata tgttgtttat gctttatatg ttttttttta ttcagaatcg ttcaccaagc
29521 attttgagaa atatggagct atatctgatt ctgtgatcat gaaagacaag catacaaaga
29581 tgcctcgtgg atttggtttt gttacatttt cggatccatc tgttattgac aaggttctgc
29641 aagatgaaca cactatagat ggaagaacag taagccttct atgtatcact tcaccaaagt
29701 atgcaaagct tgctgtagag tctgttctca tttgtctgtc atacattttg atctatcaag
29761 gttgaagtta aaaggactgt cccgagggag gaaatgtcgt cgaaagatgg tcctaaaacg
29821 agaaaaatct ttgttggtgg tattccacca tctcttactg aaggtagcca gcagaactaa
29881 gttgtttcct ttttacaagc aaaaatatat ttgtattcta gaaaacaatg gcttgaatac
29941 tctaaataat gttatgtctt gataaatatg attgtcatga aacttccaga taaattgaag
30001 gagcatttct cttcatatgg gaaggtagtt gagcatcaga taatgcttga ccatggcact
30061 gggcggtcac gaggctttgg ctttgtcaca tttgaaaatg aggacgctgt tgaaagggtt
30121 atgtcagaag ggagaatgca tgatcttgca gggaaacagg tcactagctt tccttgcttc
30181 caagttttcc ttctatttaa tattgttaac ttttcagtac ttatgagctt aagtatcaat
30241 tggtataaat ggatatgctt ctaatgaatt gattttacta ctattcactg aatttatttg
30301 ttagcacgag gatcttgttt tctgctttac tacttgaagt ttgaatattg ttccttgttg
30361 attatgtcta ccgtgacaag aagtcagttt tgctgttgta ttagaatgga cctttcttta
30421 ttcattccta tttgaaggaa ctgaaatgaa tcagtatcat atccatttga atcaatacct
30481 actagaatca gtaggtagtc ttcttggtgc acatcactac ctattctgta tgtgatttga
30541 gatcaggaga tctttaattt ctttaccttg tagtttttt tatattggtt tgttgcgaag
30601 ttctgttctc cctcgcaaag ggggtgaact tttgcattct atcatatgag acaagcgatc
30661 ctgatgctta tatgtttgcc atctgctaaa atgatgtgcc tagccctcca tgttacttcc
30721 tagtgaactg atgaactgct gaaatgcagt agagtttcct gttcaacttc ccttttaggt
30781 ctctcgtctt tttcatttgc aactatcttt gttaatttag tggtaggagg taagatgcat
30841 gattttatat cttaattgta ctgtgcgtat gaagtattac actactatgg aatgggacta
30901 tggtattgaa gattagtttg ataacttcac ctgttctgta actctgcagg ttgaaataaa
30961 gaaggctgag cctaagaaac ctggcggggg tgattctagc tctaatggga gacatagtca
31021 cgggagtggt ggtggccacc gtagttctta ccgtggtagt ggtggtggca attctggaag
31081 cagcagcagc ggtggctatg gtggatatgg tggtggctat cggtcagctg ctgcagctta
31141 ttacggtagc acaggttatg ctggctatgg cagaggttat ggatatggag gtaaccccgc
31201 atttgggtca ggctttggct ctggctatgg tggttcaatg tatggaggtc catatggtgc
31261 ctatggtgca tatggtggtg cctatggagg tggtggtgca tatggtgcac cgggtggcta
31321 cggtgcagga ggatatggtg cttatggcgg agctggaggt atgggcggtg gtgggagcac
31381 aagtggtaga ggctcgagca gataccaccc gtacggaaaa tgagcataag atcattgtgg
```

(C)

Putative cDNA sequence of N-terminus RBP-D

AGGGATGGCGGGGTACGGGGAGGAGAACCAGAACGGCATGAGGAGGAGGAGGAGGAAGAGGAGG
TGGAGGAGGTCGAGGAAGAGGTGGAGGAGGAGGAGGAGGAAGAGGAGGAGGAGGAGGGGGCTGA
TGCCACCGCCGCGGCTGCCGATGCGGCGGAGGAGGTTGCAGAGGAGAGGCGAGGCGGAGGTGGA
GAAGTGGAGGGCGTCGGCAATGGCGAGGAAG**CGGGGAGG**ACTGCGGGTGGTGGAGAGGGCGGTG
ATTCGTCGGGGAAAATTTTTGTTGGAGGTGTAGCCTGGGAGACAACTGAAGAATCGTTCACCAA
GCATTTTGAGAAATATGGAGCTATATCTGATTCTGTGATCATGAAAGACAAGCATACAAAGATG
CCTCGTGGATTTGGTTTTGTTACATTTTCGGATCCATCTGTTATTGACAAGGTTCTGCAAGATG
AACACACTATAGATGGAAGAACAGTTGAAGTTAAAAGGACTGTCCCGAGGGAGGAAATGTCGTC
GAAAGATGGTCCTAAAACGAGAAAAATCTTTGTTGGTGGTATTCCACCATCTCTTACTGAAGAT
AAATTGAAGGAGCATTTCTCTTCATATGGGAAGGTAGTTGAGCATCAGATAATGCTTGACCATG
GCACTGGGCGGTCACGAGGCTTTGGCTTTGTCACATTTGAAAATGAGGACGCTGTTGAAAGGGT
TATGTCAGAAGGGAGAATGCATGATCTTGCAGGGAAACAGGTTGAAATAAAGAAGGCTGAGCCT
AAGAAACCTGGCGGGGGTGATTCTAGCTCTAATGGGAGACATAGTCACGGGAGTGGTGGTGGCC
ACCGTAGTTCTTACCGTGGTAGTGGTGGTGGCAATTCTGGAAGCAGCAGCAGCGGTGGCTATGG
TGGATATGGTGGTGGCTATCGGTCAGCTGCTGCAGCTTATTACGGTAGCACAGGTTATGCTGGC
TATGGCAGAGGTTATGGATATGGAGGTAACCCCGCATTTGGGTCAGGCTTTGGCTCTGGCTATG
GTGGTTCAATGTATGGAGGTCCATATGGTGCCTATGGTGCATATGGTGGTGCCTATGGAGGTGG
TGGTGCATATGGTGCACCGGGTGGCTACGGTGCAGGAGGATATGGTGCTTATGGCGGAGCTGGA
GGTATGGGCGGTGGTGGGAGCACAAGTGGTAGAGGCTCGAGCAGATACCACCCGTACGGAAAAT
GA

Glay portion: sequence unconfirmed region
Blue portion: newly sequenced region
Red region: present in AK059225 cDNA from KOME

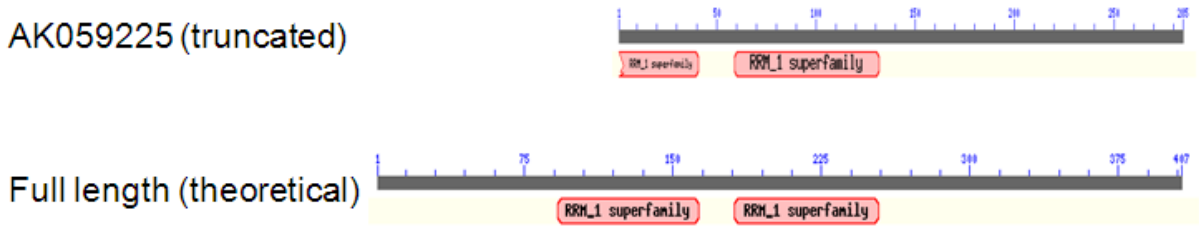
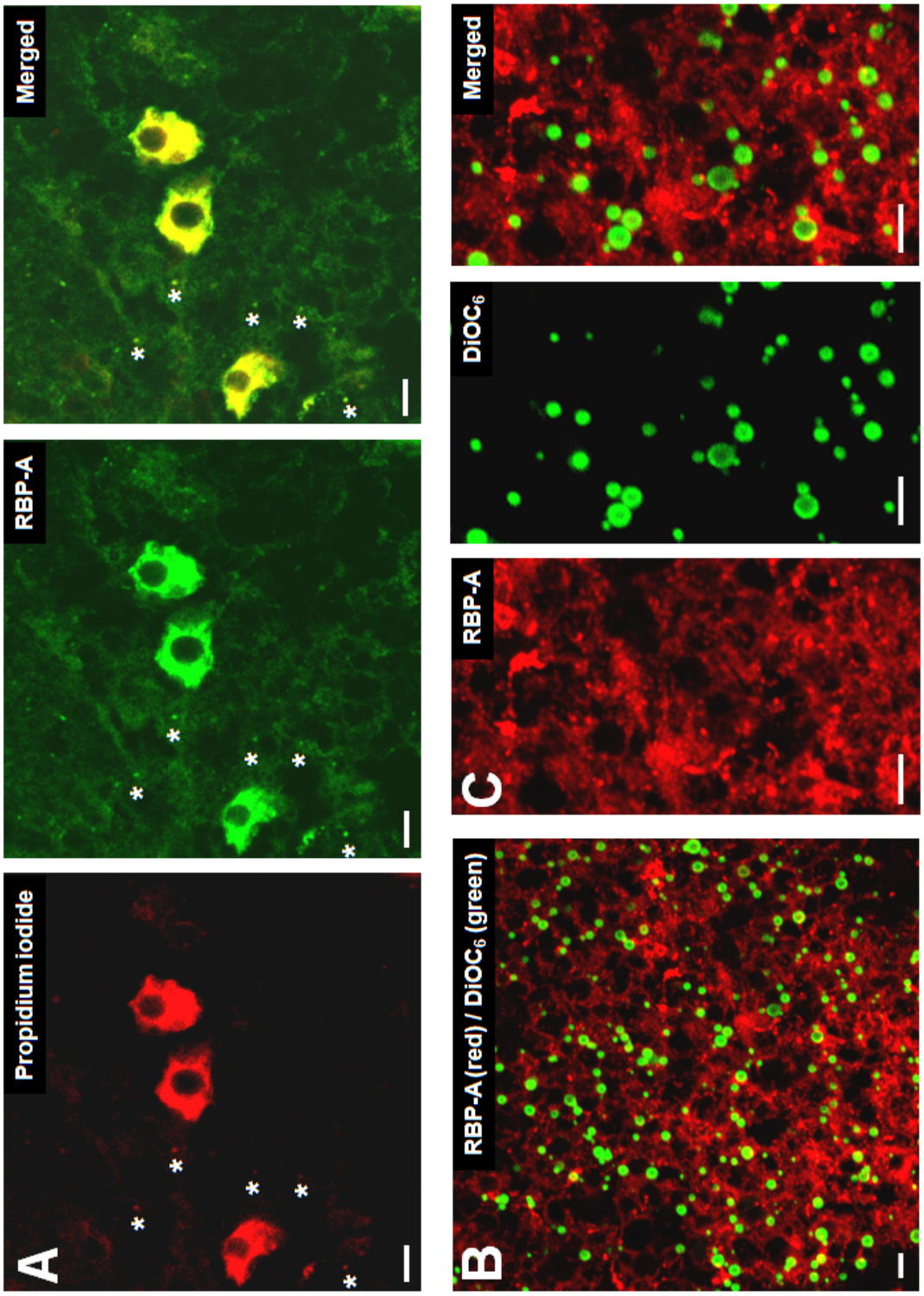


Figure 10. RBP-A is present on cisternal ER membranes and co-localizes with microtubules



D
RBP-A
N
Tubulin
Merged
N
N

Figure 11. Co-localization of RBP-D with actin filaments and protein bodies

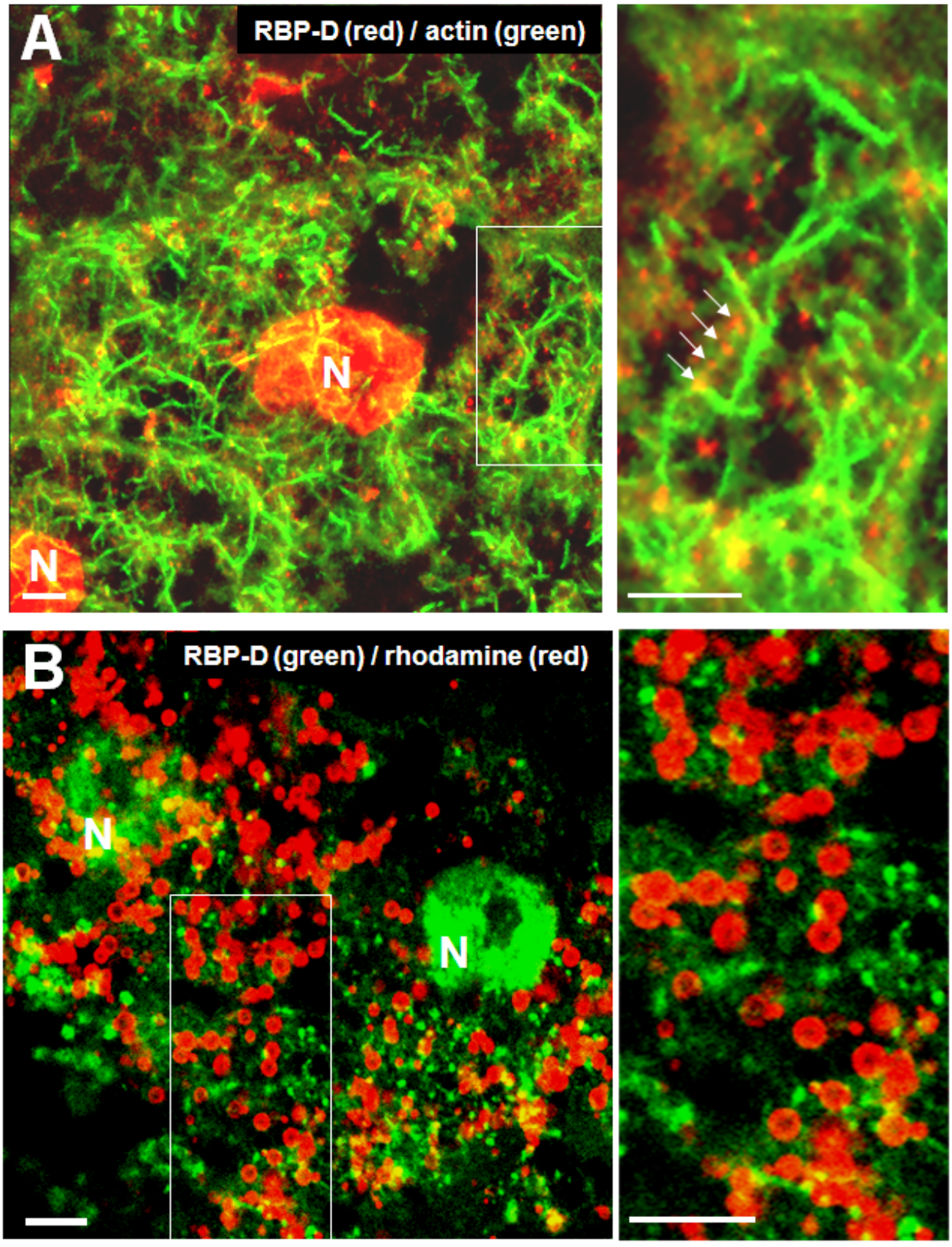


Figure 12. RBP-I localizes in nuclei

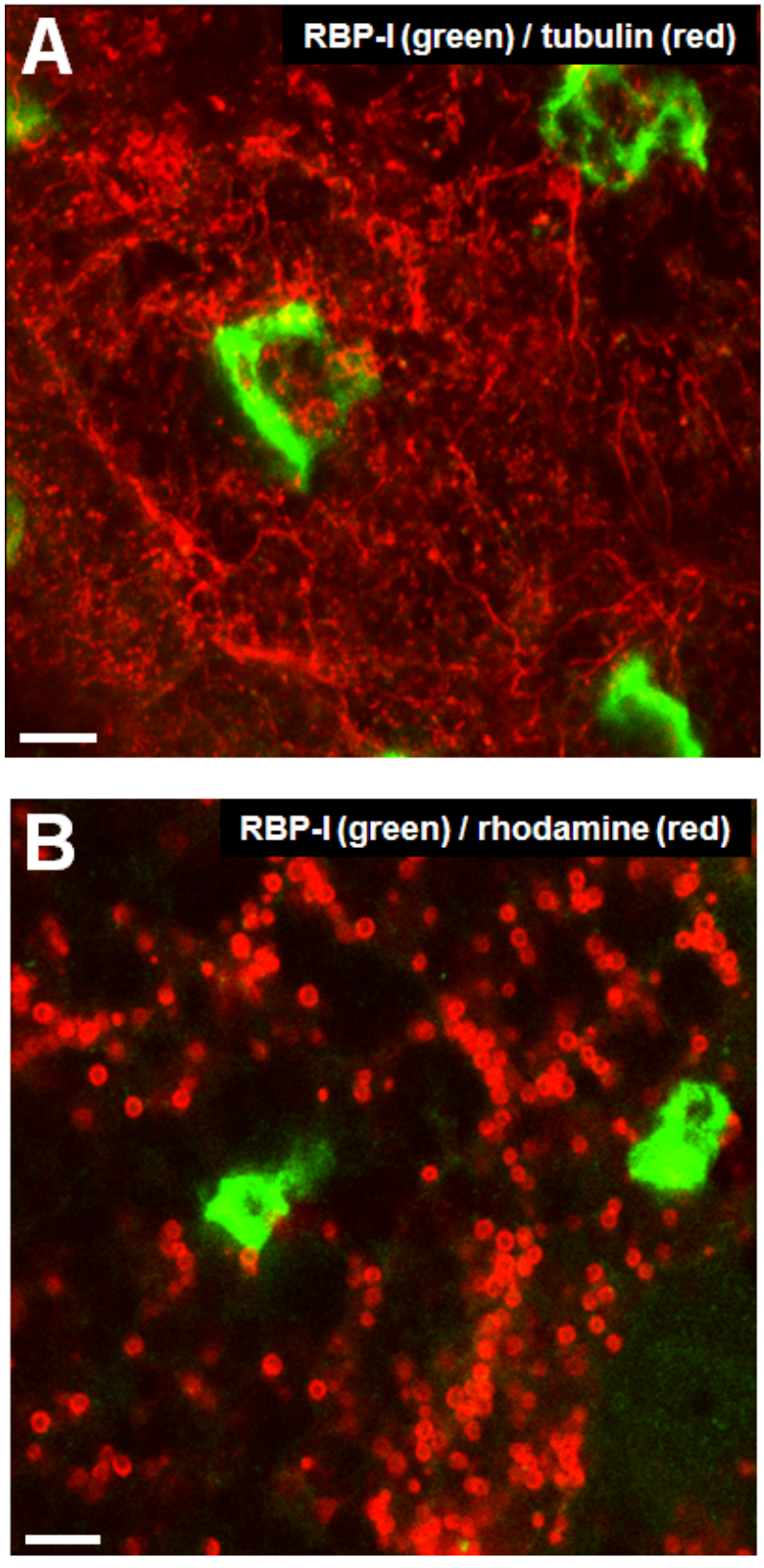


Figure 13. Localization of RBP-B, RBP-E, RBP-F, RBP-N and RBP-Q in developing rice endosperm cells

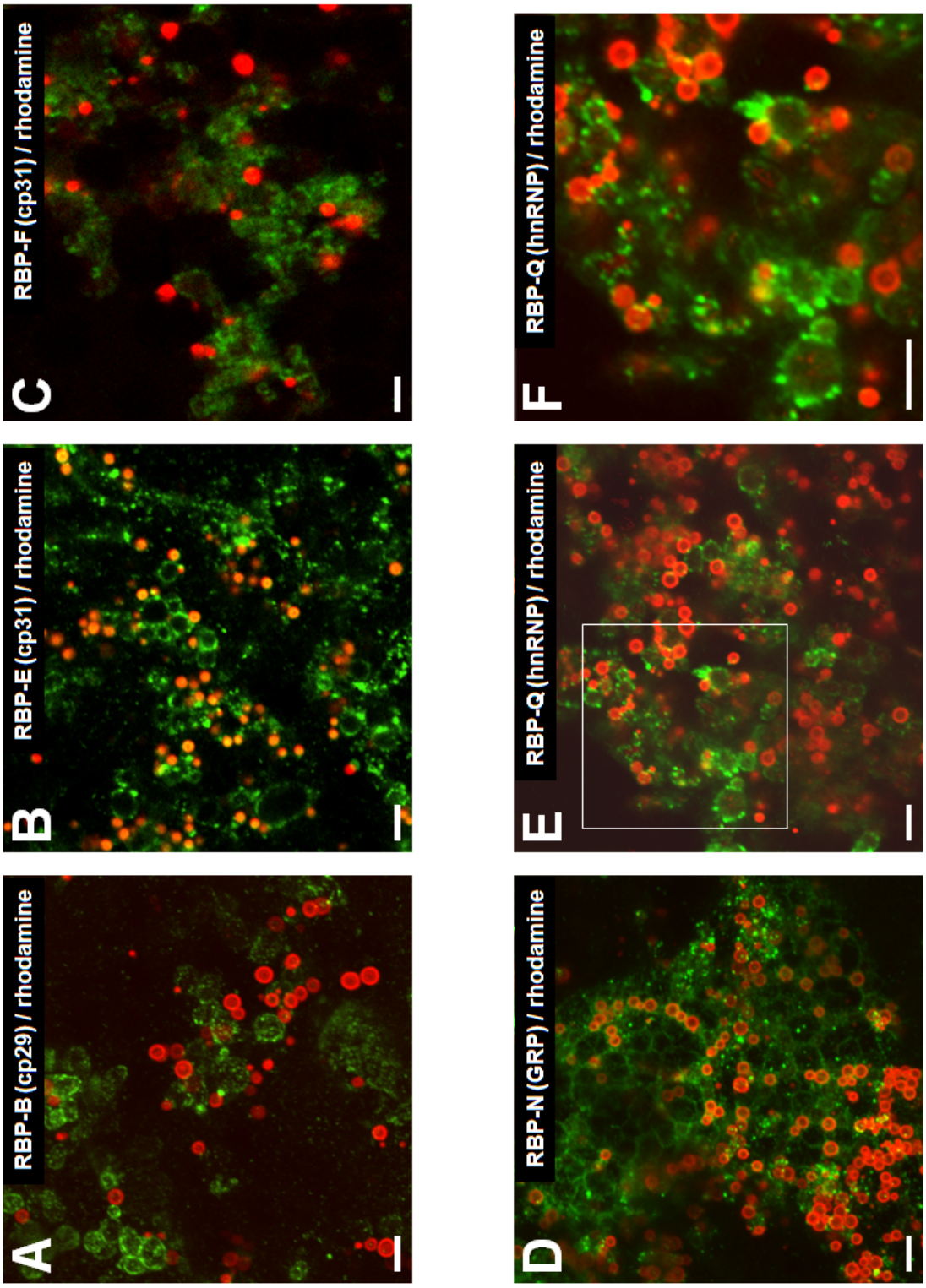


Figure 14. RBP-A expression correlates with storage protein accumulation

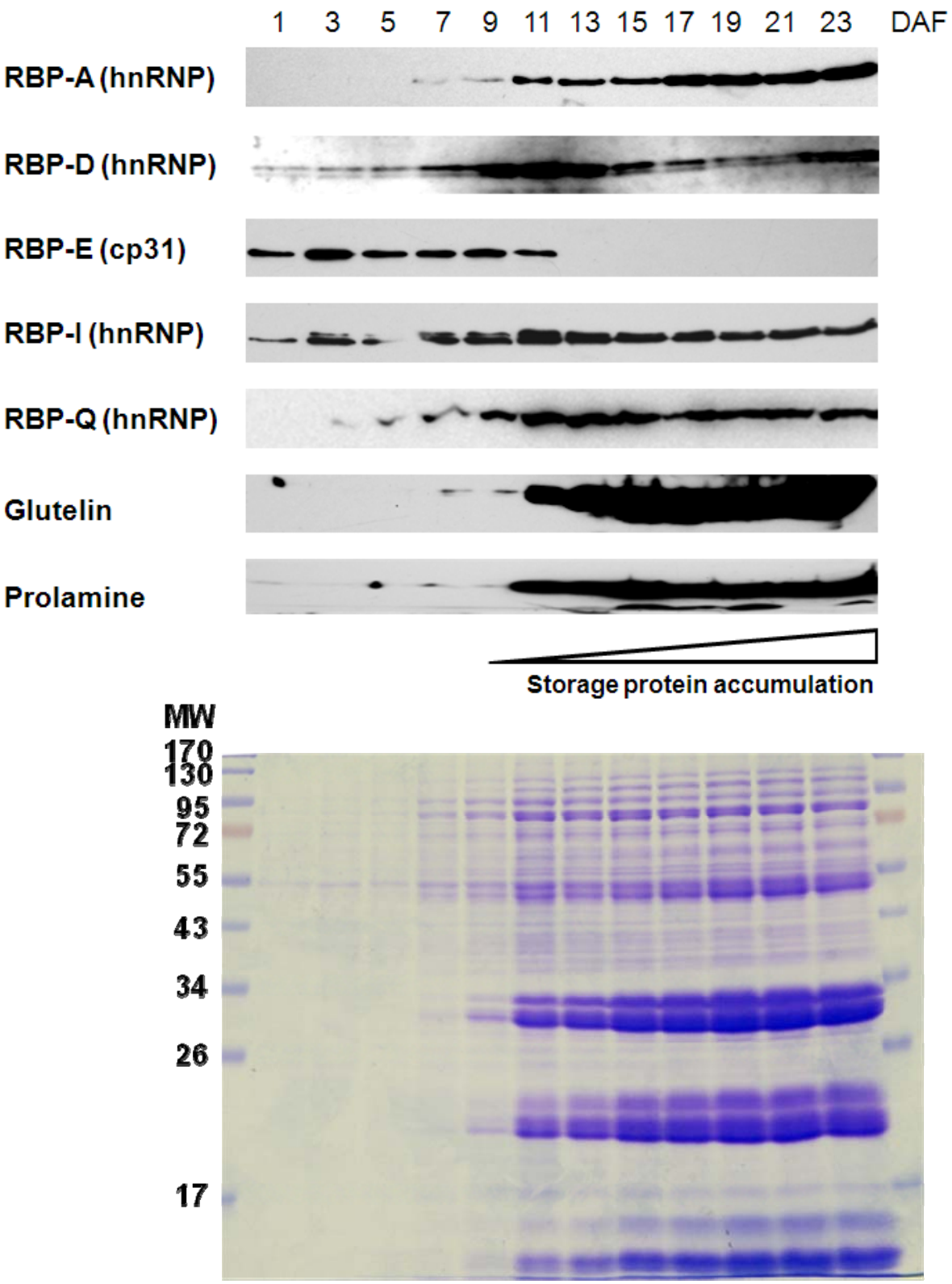


Figure 15. RNA expression levels of RBPs in developing rice seeds

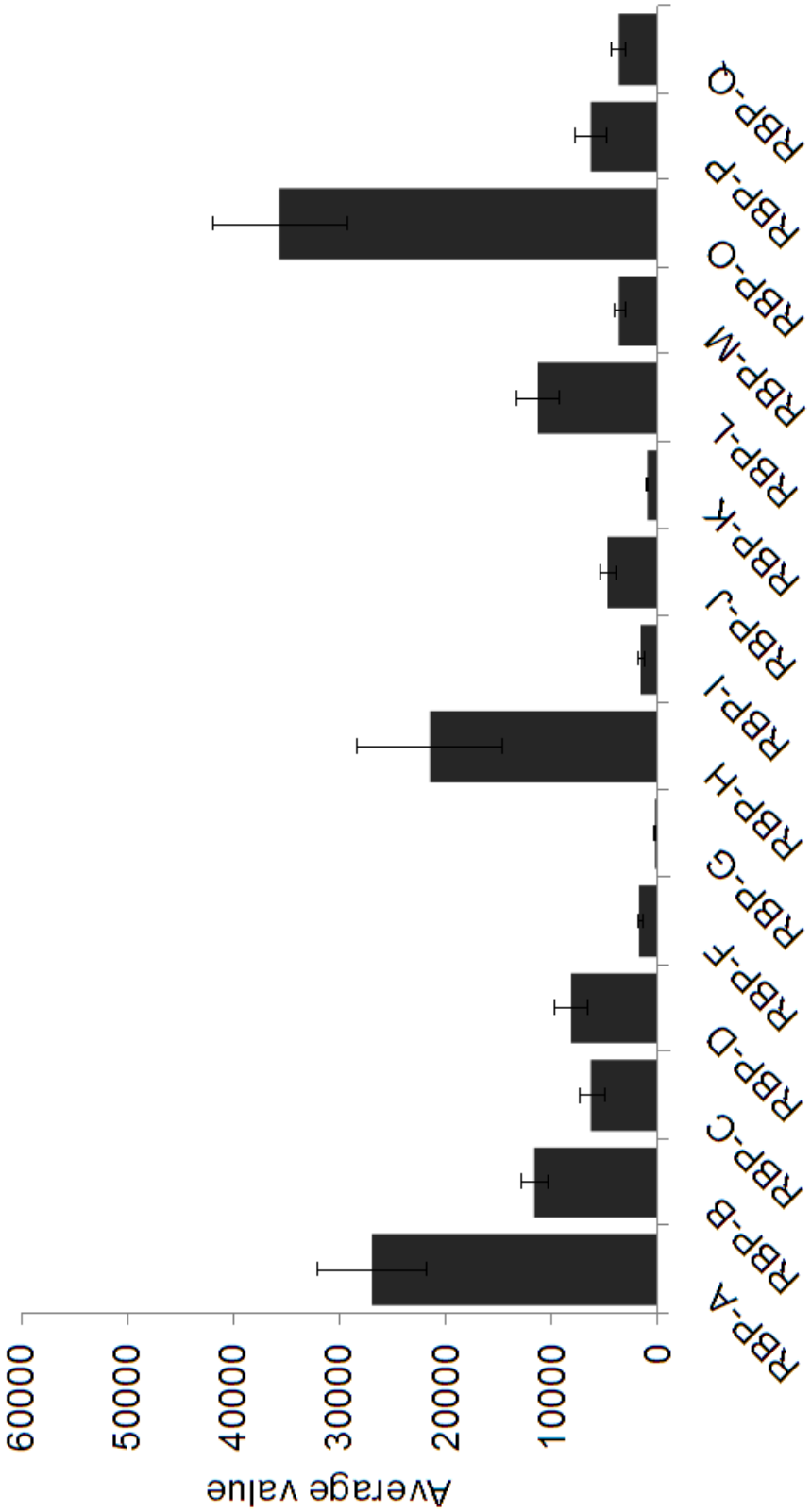


Figure 16. RBP-A is highly expressed in rapidly dividing cells

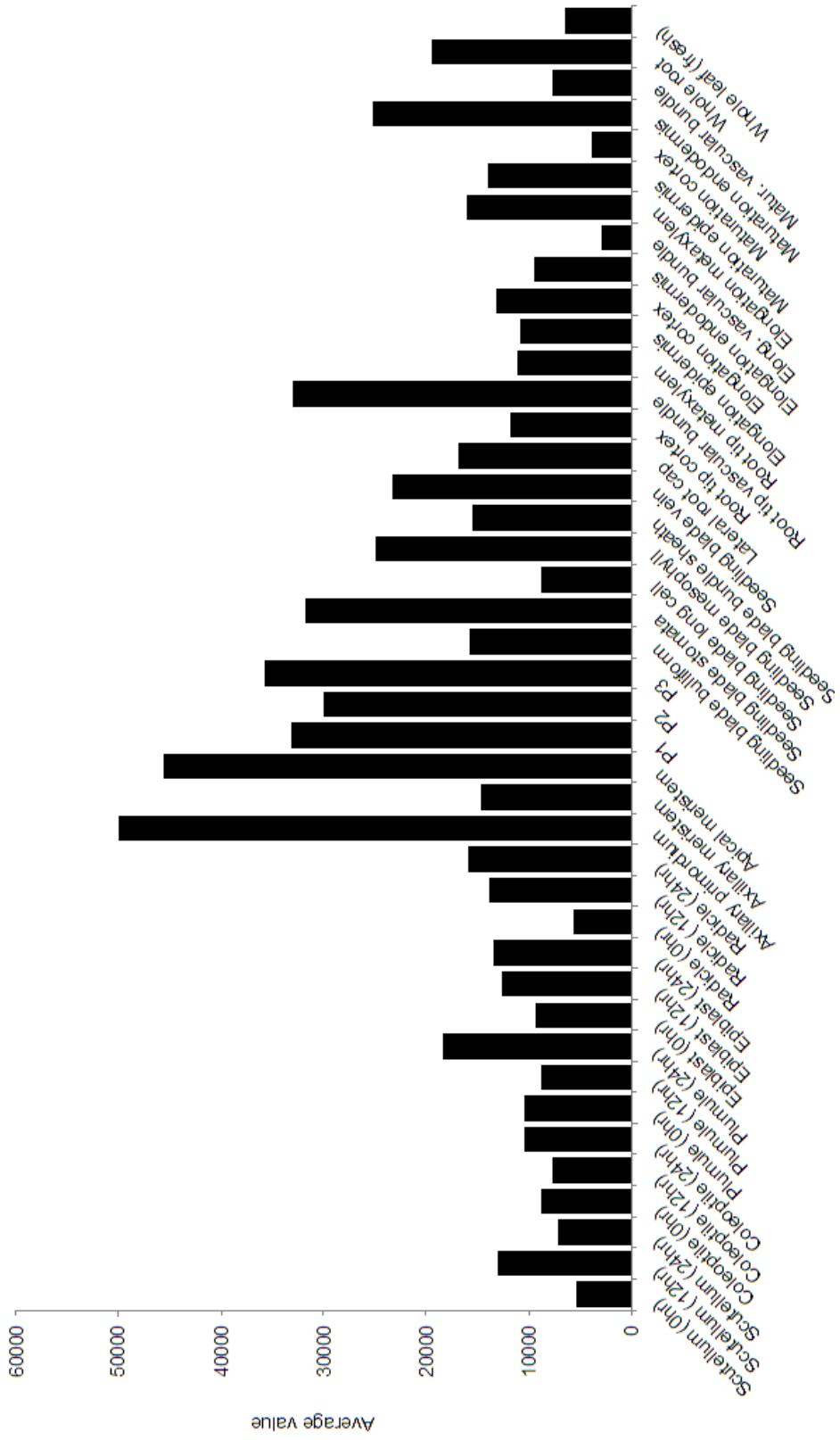


Table 1. Identification of candidate RBPs by tandem mass spectrometry

Name	gi	AK	Homology	MW / pI	Domain(s)
RBP-A	32990960	AK105751	hnRNP	49.0 / 8.44	2X RRM
RBP-B	32971090	AK061072	cp29	28.1 / 7.8	2X RRM
RBP-C	32975973	AK065955	MDH	37.4 / 8.1	MDH
RBP-D	32969243	AK059225	hnRNP	29.1 / 9.4	2X RRM
RBP-E	32979145	AK069121	cp31	35.4 / 4.4	2X RRM
RBP-F	32980766	AK070743	cp31	33.3 / 4.4	2X RRM
RBP-G	32977743	AK067725	hnRNP	50.5 / 5.9	2X RRM
RBP-H	32980735	AK070711	cold-shock	22.7 / 6.6	1X CSD 1XAIR1
RBP-I	32986113	AK100904	hnRNP	40.2 / 5.6	2X RRM
RBP-J	37990585	AK120962	hnRNP	45.6 / 6.4	2X RRM
RBP-K	37991425	AK121802	hnRNP	36.7 / 9.2	1X RRM
RBP-L	37990457	AK120834	RBP45	46.2 / 6.0	3X RRM
RBP-M	32977123	AK067105	glyoxysomal MDH	37.0 / 8.1	MDH
RBP-N	115488680	AK064923	glycine rich	19.5 / 6.6	1X RRM
RBP-O	303852	AK060925	ribosomal L3	44.5 /10.1	ribosomal L3
RBP-P	6815067	AK099896	UBP1 interacting	49.3 / 5.1	2X RRM
RBP-Q	115438550	AK120636	hnRNP	44.2 / 8.4	2X RRM
RBP-R	27311237	N/A	hypothetical	14.4 / 5.9	N/A

Table 2. List of peptides identified by tandem mass spectrometry

RBPs	AK	Coverage (%)	Peptides	Position
RBP-A	AK105751	15	LFIGGISWDTNEDR	8
			YGEVVEAVIMR	29
			GFGFIVFADPAVAER	48
			IFVGGLASTVTEADFR	108
			GFGFITYDSEDAVDK	148
RBP-B	AK061072	31	SVDSAQLAG	90
			SVDSAQLAGLFEQAGSVEMVEVVYDR	90
			GFGFVTMSTAEEAGAAIEQFNGYTFQGR	122
			RGFGFVTYGSAEEVNNAISNLDGVDLDGR	221
RBP-C	AK065955	14	VAILGASGGIGQPLALLMK	47
			DDLFNINAGIVR	131
			LLGVTTLDVVR	186
			SLGPLNEFER	324
RBP-D	AK059225	16	GFGFVTFSDPSVIDK	11
			KIFVGGIPPSLTEDKL	59
			GFGFVTFENEDAVER	100
RBP-E	AK069121	25	VYVGNLPYDIDSERLAQLFEQAGIVEVSEVIYNR	145
			GFGFVTMSTVEEAEK	185
			QFGPSFRIYVGNLPWQVDDSRLVQLFSEHGK	232
RBP-F	AK070743	30	VYVGNLPYDVDSEGL	129
			GFGFVTMSTIEEADK	169
			YVGNLPWQVDDSRLLQLFSEHGEVVNASVVYDR	224
			GFGFVSMASKEELDDAISALDGQELDGR	263
RBP-G	AK067725	3	GFGFVVFTDAGVAER	48
RBP-H	AK070711	5	SVDVTGPDGSFVK	67
RBP-I	AK100904	20	DDGRAPGGDSSGKIFVGGVAWETTE	55
			KIFVGGVAWETTEESFSKH	67
			KYGAITDSVIMKD	88
			KIFVGGLPSSLTEDELRE	156
			RGFGFVTFESEDSVERV	196
RBP-J	AK120962	4	KIFVGGLPSNLTEDEFRQ	
RBP-K	AK121802	5	RGFGFITFADPAVVDRV	56
RBP-L	AK120834	24	YYAAPPPQAMPAPAAADEVKT	71
			RDKQSGQLQGYGFVEFTSRA	123
			KFGDPTEQARA	228
			RVPNAQGAQSENDPNNTTIFVGGLDPNVTEDVLKQ	268
			KQVFAPYGEVVHVKI	301
RBP-M	AK067105	15	KVAVLGAAGGIGQPLSLLMKL	44
			NLISNPVNSTVPIAAEVFKK	156
			RLLGVTTLDVARA	183

RBPs	AK	Coverage (%)	Peptides	Position
RBP-N	AK064923	11	GGGGGGGGYQGGGGGYGGNNGGYGNRGG	151
RBP-O	AK060925	4	KDEMIDIIGVTKGKGY	214
RBP-P	AK099896	6	DADSIQALLNSFPK	145
			GGGLGGLYGAGTSGGR	385
RBP-Q	AK120636	7	KIFVGNVPADMPSERL	178
			KTPEGAQASLVDSVKV	226
RBP-R	AJ302060	38	RSLEAAFSTYGEILESKI	22
			RGFGFVTFSSEQAMRD	49
			RGFGFVTFSSEQAMRDAIEGMNGKE	49
			RDAIEGMNGKELDGRN	63
			RGGGGGYGGGGYGGGGGGGGYGQRR	100
			RGGGGGYGGGGYGGGGGGGGYGQRRE	100

CHAPTER THREE

DEVELOPMENT AND USE OF A METHOD FOR THE ANALYSIS OF RNA SPECIES INTERACTING WITH A RICE CYTOSKELETON-ASSOCIATED PROLAMINE ZIPCODE RNA BINDING PROTEIN

Abstract

Rice is an excellent model system for RNA localization since mRNAs encoding the two major seed storage proteins, prolamine and glutelin, are targeted to distinct subdomains of the cortical endoplasmic reticulum. This initial RNA localization event is crucial since it determines the final site of protein accumulation. We have previously identified a number of cytoskeleton-associated RNA binding proteins (RBPs) by affinity capture using a prolamine zipcode sequence as bait. One of the candidate proteins, termed RBP-A, a member of the heterogeneous ribonucleoprotein (hnRNP) family, shows a similar expression pattern to those of the seed storage proteins and localizes as particles present on both microtubules and ER membranes. In order to confirm that RBP-A, and the other candidate RBPs identified, are able to bind RNA *in vivo*, an RNA-immunoprecipitation (RNA-IP) method was established using developing rice endosperm cells as the starting material. The development of this method involved the optimization and use of reversible formaldehyde cross-linking to stabilize pre-existing protein-RNA complexes. RT-PCR experiments using material captured using this RNA-IP method show that RBP-A binds significant amounts of prolamine and glutelin mRNA compared to controls. In order to determine which other RNA species are bound by RBP-A, material captured by RNA-IP was subjected to reverse transcription, followed amplification, labeling and hybridization to whole genome microarrays. These experiments once again confirmed binding to storage proteins

mRNAs, but showed that RBP-A also binds to RNAs encoding glycine rich RBPs and several proteins involved in stress responses, in addition to RNAs encoding proteins involved in regulating protein synthesis and turnover. Computational sequence analysis of the captured mRNAs will allow prediction of consensus binding sequences which should prove valuable in understanding the mechanism of mRNA localization. The newly developed methods described here will also be directly applicable to the identification of the RNA species bound by other candidate prolamine zipcode binding proteins.

Introduction

The localization of storage protein mRNAs in rice seeds is perhaps the best studied model system for RNA sorting in the plant kingdom. Rice endosperm cells are unique in that they accumulate high levels of two different classes of storage proteins, prolamines and glutelins, and store them in separate intracellular compartments (Crofts et al., 2004; Crofts et al., 2005). This protein sorting is dependent on the asymmetric mRNA localization of prolamine and glutelin mRNAs to distinct subdomains of the endoplasmic reticulum (ER). Prolamine mRNAs are targeted to the protein body-ER (PB-ER) membranes which surrounds prolamine-containing protein bodies and are then translated, forming intracisternal granules called protein body I (PB-I). In contrast, glutelin mRNAs are targeted to adjacent cisternal ER (cis-ER) subdomains, and following translation, encoded precursors are exported and stored in protein storage vacuoles (PSV or protein body II, PB-II) (Li et al., 1993). *Cis*-acting localization sequences have been identified in both prolamine and glutelin mRNAs and are required for both PB and cis-ER targeting pathways, respectively. Whilst prolamine contains two zipcodes, present in the 5' coding sequence (5'CDS) and 3' untranslated region (3'UTR) (Hamada et al., 2003a), glutelin mRNA contains at least three cis-acting localization signals, at least one of which is present in the 3'UTR (Washida et al., submitted). The replacement of the prolamine 3'UTR with the glutelin 3' end sequences (which contains two cis-ER localization signals) causes the hybrid mRNA to be mis-targeted to the cis-ER (Choi et al., 2000). Additionally, experiments using heterologous expression of maize and sunflower seed storage proteins have shown that mis-localizing mRNAs to alternative ER subdomains directly influences the final protein localization. For example, the additional of zein zipcodes, which target RNAs to PB-ER, can at least partially re-localize the otherwise cis-ER targeted sunflower seed albumin, with a portion of the encoded

protein now being retained within PB-I (Washida et al., 2009). Thus, the site of initial RNA localization defines the final site of storage protein accumulation. The next question is therefore, which trans-acting RNA binding proteins are recognizing these *cis*-acting zipcode signals.

Previously, (see Chapter 2) the use of a prolamine RNA zipcode affinity column allowed the isolation of a large number of cytoskeleton-associated RNA binding proteins which bound under high stringency conditions. Following their identification by tandem mass spectrometry, these proteins, named RBP-A through RBP-R, were expressed from cDNAs in *E. coli* and used to generate antibodies which were subsequently affinity purified. Interestingly, of the eighteen individual RNA binding proteins identified, a total of seven were members of the heterogeneous ribonucleoprotein (hnRNPs) family. The hnRNPs are known to perform a wide range of cellular functions based on their association with RNAs to form RNP complexes. The different compositions of such complexes have been proposed to serve as a form of molecular memory, documenting an RNA's history as it moves towards its final destination (reviewed by Dreyfuss et al., 2002). As an example, some of the roles that have been reported for hnRNPs include, associate with newly transcribed pre-mRNA as part of splicing complexes (reviewed by Martinez-Contreras et al., 2007), transport of mRNAs from the nucleus to the cytoplasm that are associated with cytoskeletal elements (Kim et al., 2005; Percipalle et al., 2009). hnRNPs are also involved in mRNA localization (Norvell et al., 2005; Lewis et al., 2008; Kroll et al., 2009), in translational silencing (Cáceres and Nilson, 2008) and in anchoring RNAs once the final destination has been reached (Delanoue et al., 2007). Finally, hnRNPs have also been shown to regulate the stability of mRNAs (Bolognani et al., 2008).

RBP-A, a member of the hnRNP family, contains two RNA recognition motifs at its N-terminus and its homologs in *Drosophila* and *Xenopus* are known to be involved in mRNA

localization. As described in Chapter 2, our affinity capture experiments show that RBP-A exhibits good binding specificity for prolamine zipcode RNA compared to a non-zipcode control RNA of the same length. The expression pattern of RBP-A during seed development closely matches the accumulation pattern of prolamines and glutelins and this protein is localized to the nucleus as well as being present in small particles that are associated with microtubules in the cortical region of endosperm cells. The actin cytoskeleton is known to be required for the active transport of prolamine mRNA to PB-ER membranes (Hamada et al., 2003b), but the possible role of microtubule association of RBP-A is currently not clear. However, previous work has shown that prolamine protein bodies are closely associated with both the microtubule and actin cytoskeleton and are more prevalent in the cortical region of the cell (Muench et al., 2000). This suggests that the microtubules may be acting as a region for the mRNA storage (see Discussion).

Although RBP-A is clearly able to bind prolamine zipcode sequences *in vitro*, it is very important to confirm that this protein is also able to bind to prolamine mRNAs *in vivo*. Once this *in vivo* binding to prolamine mRNA is established, then the next question to ask will be what other RNA species are present or co-transported in RBP-A-containing RNP complexes. Unfortunately, the choice of methodology to determine the RNA targets of RBPs *in vivo* is currently limited. One direct way of RNA target identification is the use of RNA-immunoprecipitation (RNA-IP) to pull down RNP complex an antibody against RBP of interest (Schmitz-Linneweber et al., 2005; Peritz et al., 2006; Baroni et al., 2008; Keene et al., 2006). Use of extracts prepared after treatment of the target tissue with formaldehyde, a reversible nucleotide-protein cross-linker, enables the maintenance of *in vivo* assembled RNA-protein complexes. Importantly, it also prevents RBPs from binding to non-target RNAs immediately following extraction and during the course of the experiment (Yang et al., 2005). Successful

RNA-IP experiments have been reported in yeast (Gilbert and Svejstrup, 2006), mammalian cells (Niranjanakumari et al., 2002), *Arabidopsis* leaf tissue (Terzi and Simpson, 2009) and isolated chloroplasts from maize (Schmitz-Linneweber et al., 2005). These small or thin samples allow formaldehyde to penetrate evenly and are easy to handle during the fixation procedures. Mid-stage developing rice seeds, however, are much larger in volume, and therefore obviously required the development of a new fixation protocol. In addition, the abundance of storage protein mRNAs in developing seeds means that the number of washing steps needed to be increased in order to reduce background binding.

The aims of this study were therefore three fold; to establish a fixation and RNA-immunoprecipitation (RNA-IP) procedure using rice developing seed; to confirm the *in vivo* interaction of RBP-A and prolamine mRNA, and finally; to identify the other RNA species present in complexes with RBP-A using microarrays. In the future, once glutelin zipcode binding proteins have been identified, this method will allow us to identify the common and distinct RNA binding proteins involved in both prolamine and glutelin RNA localization pathways. A combination of RNA-IP and either microarray analysis, or high-throughput sequencing technology, should provide valuable insight into the mechanism of RNA localization in rice developing seeds.

Materials and Methods

Rice growth conditions

Oryza sativa L. cultivar Kitaake was grown as described in Chapter 2 with cycles of light period at 26 ^{o}C for 12 hours and dark periods at 22 ^{o}C for 12 hours with daily watering. Developing seeds at 10-14 days after flowering (DAF) were used for experiments.

Fixation of developing seeds

Freshly harvested 10-14 DAF developing rice seeds were used and their husks were removed prior to sectioning. Developing seeds were sectioned to approximately 0.5 mm thickness using a razor blade, generally yielding around 12 sections per rice seed. Seed sections were immediately incubated for 1 hour in 10 volumes of 10 mM DMA (dimethyl adipimidate•2 HCl) in modified cytoskeleton stabilization buffer (mCSB; 20 mM HEPES-KOH, pH 8.0, 0.2 M sucrose, 10 mM $Mg(OAc)_2$, 2 mM EGTA, 1 % Triton X-100, 1 mM PMFS, 1 mM DTT, 1 unit/mL RNase inhibitor) (Muench et al., 1998). Sections were washed once with 10 volumes of mCSB and cross-linked by incubating with 10 volumes of 0.1 % formaldehyde, in mCSB for 10 min at room temperature. Following fixation, rice sections were quenched by three washes with excess Tris-containing immunoprecipitation buffer (IP; 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5 % NP40, 1 mM PMFS, 1 mM DTT, 1 unit/mL RNase inhibitor) and stored at -80 ^{o}C.

Immunoprecipitation:

Protein A conjugated agarose resin (Invitrogen) was washed with IP buffer and a 10 % (v/v) slurry was made. Approximately 2 μg of affinity-purified anti-RBP-A or anti-GFP polyclonal rabbit antibodies were incubated with 50 μL of 10 % resin slurry in 500 μL of IP buffer for at least 24 hours at 4 ^{o}C with rotation. Antibody-resin was then washed three times with 500 μL of IP buffer. Fixed seeds were ground to powder in liquid nitrogen using a mortar and pestle. Powder was extracted with IP buffer (0.1g powder/mL IP buffer) supplemented with RNaseout (20 U), centrifuged at 12,000 rpm for 10 min. 2 mL each of supernatant was added to washed antibody-Protein A resin and incubated for 4 hours at 4 ^{o}C with rotation. Samples were

washed five times with 1 mL IP buffer at room temperature by repeating mix, spin and removal of supernatant.

RNA purification from immunoprecipitated material:

Resin was washed twice with 1X DNase buffer containing 10 mM Tris-HCl, pH 7.5, 2.5 mM $MgCl_2$, 0.1 mM $CaCl_2$, RNaseout (4 U) leaving a final volume of 30 μL which was then supplemented with RNase-free DNase I (1 U) and RNaseout (20 U) and incubated at 37 ^oC for 30 min in order to remove genomic DNA. Samples were then supplemented with final concentrations of 0.5 % SDS, 2.5 mM EDTA and proteinase K (50 μg/mL) and incubated at 60 ^oC for 1 hour in order to reverse formaldehyde cross-linking and to remove proteins. 200 μL of TRIzol (Invitrogen) were added to resin and stored at -80 ^oC overnight. Samples were defrosted, supplemented with 1 μg of glycogen and centrifuged to remove resin. Supernatants were then extracted with 50 μL of chloroform:isoamyl alcohol (24:1) and centrifuged. 150 μL of the resulting water-phase was then mixed with 50 μL of nuclease-free water, 700 μL of RLT buffer and 500 μL ethanol and applied onto RNeasy column (QIAGEN). Columns were washed with 500 μL RPE buffer followed by 500 μL 80 % ethanol and RNAs were eluted in 14 μL of nuclease-free water.

Microarray sample preparation

First strand cDNA synthesis:

An Amino Allyl MessageAmp^{TM} aRNA amplification kit (Ambion) was used for probe preparation and the protocol was adapted slightly from that supplied by the manufacturer. RNAs were dried down to allow the use of a 10 μL reaction scale. Total RNA were first primed with T7

oligo(dT) primer and cDNAs were synthesized using ArrayScript in 20 μL volumes at 42 ^{o}C for 2 hours.

Second strand cDNA synthesis and cDNA purification:

Second strand cDNA synthesis was performed to generate double stranded transcription template by adding 80 μL of second strand synthesis reaction mixture, containing DNA polymerase, and incubating at 16 ^{o}C for 2 hours. Double stranded DNA was then purified by mixing with 250 μL of cDNA binding buffer, applied onto a cDNA filter cartridge, washed, dried and eluted in 2X 9 μL of nuclease-free water. Purified dsDNA were concentrated to 14 μL by speed vacuum.

In vitro transcription and aRNA purification:

26 μL of *in vitro* transcription reaction mixture was mixed with dsDNA template and antisense RNA (aRNA) was amplified at 37 ^{o}C for 16 hours before terminating the reaction by the addition of 80 μL nuclease-free water. aRNA was then mixed with 350 μL of aRNA binding buffer, 250 μL of ethanol and purified using aRNA filter cartridges. After washing and drying, aRNA was eluted in 100 μL nuclease-free water.

First strand synthesis for second round amplification-:

Amplified RNA was dried down to 5 μL in a vacuum centrifuge and primed with second round oligo(dA) primer. 4 μL of first strand reaction mixture was added and incubated at 42 ^{o}C for 2 hours for reverse transcription. The RNA strands were then digested with RNase H at 37 ^{o}C for 30 min.

Second strand synthesis for second round amplification and cDNA purification:

cDNA was primed with T7 oligo(dT) primer and second strand synthesis reaction mixture was added at a 50 μL scale and incubated at 16 ^{o}C for 2 hours. Double stranded DNAs were then purified by mixing with 125 μL of cDNA binding buffer and applied onto cDNA filter cartridges,

before washing, drying and elution twice using 2X 9 μL of nuclease-free water. Purified dsDNA was concentrated to 14 μL by vacuum centrifuge.

In vitro transcription for amino allyl aRNA amplification and purification:

26 μL of *in vitro* transcription reaction mixture containing aaUTP was added and incubated at 37 ^{o}C for 16 hours to amplify aRNA while incorporating aaUTP. The reaction was terminated by the addition of 60 μL nuclease-free water. aRNA was then mixed with 350 μL of aRNA binding buffer, 250 μL of ethanol and purified using an aRNA filter cartridge. After washing, aRNA was eluted in 100 μL of nuclease-free water.

Dye coupling reaction:

Purified aRNA was dried to completion in a vacuum centrifuge and resuspended in 3.3 μL of coupling buffer. 40 nmol *N*-hydroxysuccinimide (NHS) ester Cy3 or Cy5 dye (GE Healthcare) was dissolved in 11μL of DMSO. 3.6 μL each of Cy3 dye (for control sample) or Cy5 dye (for experimental sample) were mixed with aRNA and incubated for coupling in the dark for 1 hour. The coupling reaction was stopped by the addition of 2.25 μL 4M hydroxylamine and incubation for 15 min prior to dilution with 30 μL of nuclease-free water.

Labeled aRNA purification and fragmentation:

aRNA was mixed with 105 μL of aRNA binding buffer, 75μL of ethanol and applied onto a labeled aRNA filter cartridge. After washing and drying, labeled aRNAs were eluted twice with 2X 14 μL nuclease-free water and dried down to 18 μL final volume. 2 μL of X10 fragmentation buffer (Ambion) was added and samples were incubated at 70 ^{o}C for 15 min prior to the addition of 2 μL Stop solution. Cy3 and Cy5 labeled probes were mixed and dried to completion in speed vacuum.

Hybridization:

Rice NSF45K arrays were purchased through the NSF Rice Oligonucleotide Array Project (http://www.ricearray.org/index.shtml) described by Jung et al., (2008). These paired array slides contain 43,311 50 to 70-mer oligos based on 45,116 gene models, each of which is supported by either an EST and/or full-length cDNA in TIGR rice annotation release 3. Array slides were pre-hybridized with 5 X SSC, 0.1 % SDS, 1 % BSA at 42 ^{o}C for at least 2 hours with rotation and then rinsed twice, each with 10 dipping cycles, in water prior to drying under a stream of nitrogen. Probes were resuspended in 58 μL of hybridization buffer (50 % formamide, 5X SSC, 0.1 % SDS, 10 mM DTT) and supplemented with 20 μg of fragmented salmon sperm DNA. Probes were heat denatured for 3 min, snap cooled on ice and then centrifuged to remove any precipitates. 28 μL was then applied onto each of the pre-hybridized paired slides. Coverslips were placed on the slides which were then sealed in hybridization chambers and incubated at 48 ^{o}C for 16 hours in the dark. After hybridization, slides were placed in a washing dish with low stringency buffer (2 X SSC, 0.1 % SDS) at 48 ^{o}C and washed for 10 min twice with stirring. Washing buffer was exchanged with high stringency buffer (0.1 X SSC, 0.1 % SDS) and slides were washed twice for 5 min with stirring at room temperature. Final washes were performed twice with 0.1X SSC for 10 min with stirring at room temperature. Slides were then dried under a stream of nitrogen gas.

Scanning and analysis:

Slides were scanned with ScanArray 4000XL (Packard BioScience, now PerkinElmer) and data were collected by GenePix Pro 6.1 (Molecular Devices, Sunyvale, CA).

Reverse transcription-polymerase chain reaction (RT-PCR)

RNAs were extracted from immunoprecipitates using 200 μL TRIzol (Invitrogen) and centrifuged to allow removal of resin. Supernatants was extracted using 40 μL chloroform:isoamyl alcohol (24:1), the water-phase was supplemented with 1 μg glycoblue and precipitated with 100 μL isopropanol. Precipitates were rinsed with 70 % ethanol and resuspended in 5 μL of nuclease-free water. Oigo(dT) primed cDNA was produced using Moloney murine leukemia virus (M-MLV) reverse transcriptase in 5 μL scale reactions containing 1X MMLV buffer, 0.5 mM dNTPs, SUPERaseIn 1 U (Ambion), 25 ng oligo(dT), M-MLV-RT enzyme (Promega) 10 U and 2.5 μL of purified RNA by incubation at 37 ^{o}C for 1 hour. PCR amplifications were performed in 20 μL reactions containing 50 mM Tris-HCl, pH 9.2, 16 mM $(NH_4)_2SO_4$, 0.35 mM $MgCl_2$, 0.1 % Tween 20, 6 U of KlenTaq1/DeepVent DNA polymerase (NEB) using 80 μM forward and reverse primers. Prolamine primers (AK240910-107f 5'-GTCAAAGTTATAGGCAATATCAGCTGC-3' and AK240910-428r 5'-CCTAGGGT-AGATACCATATCTAGATGGC-3') or glutelin primers (Gt2-exon2-SacIf 5'-ATAGAGCTC-AGAGGTATAACAGGGCCAAC-3' and Gt2-exon3-SacIr 5' ATAGAGCTCGCACTCATC-TGTACAAGATTGAG-3') with following conditions: 1 cycle at 94 ^{o}C for 3 min, 28 cycles (unless indicated) of 94 ^{o}C for 30 sec, 53 ^{o}C for 30 sec, 72 ^{o}C for 30 sec, and 1 cycle at 72 ^{o}C 5 min. Samples were analyzed by agarose gel electrophoresis and visualized using ethidium bromide.

Results

Establishment of fixation conditions for cross-linking of protein-RNA interactions in developing rice endosperm cells.

Mid-developing seeds were harvested 10-14 days after flowering since this is when both the RNA expression levels of seed storage proteins (Kim and Okita, 1988) and RBP-A are maximal. Seeds within this developmental range have significant starch content, but remain relatively soft and can be sliced easily.

In order to identify fixation conditions that allow uniform fixation at an appropriate level, either whole seeds or sliced, hand-sectioned, seeds were incubated with either the protein-protein cross-linker dimethyl adipimidate (DMA) or the reversible nucleotide-protein cross-linker formaldehyde. Varying incubations times were used as indicated in Figure 1. Both DMA and formaldehyde solutions were made with modified cytoskeleton stabilization buffer (mCSB), which has a higher pH to enhance cross-linking while helping to maintain the integrity of the cytoskeleton. The degree of cross-linking was determined by analyzing the SDS-urea denatured protein present in extract supernatants after centrifugation to remove starch and insoluble materials. Figure 1A shows that incubation of whole seeds in 10 mM DMA does not alter the amount of extractable protein, while the seed sections showed slightly less total extractable protein available only after incubation for 120 min. In addition, western blot analysis of RBP-A did not show any decrease in signal at this time point indicating that RBP-A is stable in the sectioned seeds and is readily extracted after DMA cross-linking. When the seeds were incubated with 1 % formaldehyde as shown in Figure 1B, the whole seeds did not show any change in the amount of appearance of total protein, whilst the sectioned seeds showed a drastic decrease in protein recovery after less than 15 min of fixation. The coomassie-stained gel also shows

streaking of high molecular weight proteins due to excessive protein-protein cross-linking. The amount of RBP-A recovered, as assessed by western blotting followed the same pattern, with the recovery efficiency dropping rapidly in less than 15 mins.

In order to confirm that RBP-A can be recovered from cross-linked seed sections using IP buffer, shorter formaldehyde fixation time course experiments were performed as shown in Figure 1C. Sectioned seeds were incubated with 0.1 % formaldehyde for between 5 and 20 min and then washed with Tris-containing IP buffer to quench residual formaldehyde prior to extraction in the same IP buffer and centrifugation. Since the majority of RBP-A localizes either to the nucleus or the cortical region of rice endosperm cells, it was not surprising that RBP-A signals were found in both pellet and supernatant samples. Both total protein and western analysis of the supernatant fraction failed to show a significant reduction of total protein or RBP-A after 5 min, whilst the pellet signal showed only a slight decrease in RBP-A signal. Therefore, the procedure we adopted included initial protein-protein cross-linking with DMA (to stabilize the cytoskeleton), followed by cross-linking with 0.1 % formaldehyde for 10 min. The sectioning of mid-developing rice seeds prior to fixation allowed formaldehyde to efficiently to penetrate the tissue, yielding a shorter fixation time still allowing easy handling during solution exchanges. This level of formaldehyde cross-linking was selected in order to prevent excess cross-linking which might lead to high background levels for both protein and RNA.

Immunoprecipitation of RBP-A from formaldehyde cross-linked developing rice seed extracts.

In order to test whether RBP-A can be captured from formaldehyde-treated material, immunoprecipitation was performed using affinity purified anti-RBP-A antibody. Five washing

steps were used to minimize background. Figure 2 shows that RBP-A was efficiently captured by immunoprecipitation, with only a very slight degradation product being detected. No RBP-A was detected in the antibody or extract control immunoprecipitations. The fact that RBP-A protein is not seen in the silver stain is most likely due to masking by IgG heavy chains which have a similar molecular weight. The silver stained gel does, however, show two additional bands of around 30 kDa which are only observed in the presence of both antibody and extract. These bands are likely to represent interacting proteins since they are not recognized by RBP-A antibody in western blots.

RBP-A associates with both prolamine and glutelin mRNA *in vivo.*

Having demonstrated that immunoprecipitation with RBP-A antibody can capture RBP-A-containing complexes from formaldehyde cross-linked material, we can now determine whether this same method can be used to co-precipitate interacting mRNAs. Since RBP-A is expected to interact with prolamine mRNA *in vivo*, a similar IP experiment to that described above was performed. Since developing rice seeds contain large amounts of storage protein mRNAs, we needed to confirm that sufficient washing had been performed. Therefore, samples from the final (5th) wash were taken in addition to samples from both the unbound (PCR positive control) and captured material. Following DNase and proteinase treatment, which was performed at the same time as cross-link reversal, RNA was isolated from these samples and used for oligo(dT)-primed RT-PCR with gene-specific prolamine primers. Prolamine genes lack introns (Kim and Okita, 1988), so intron spanning primers could not be used. DNase treatment was therefore required in order to eliminate the possibility of genomic DNA contamination.

The gel image in Figure 3A shows that five washes were sufficient to essentially eliminate any background prolamine mRNA. Although a weak signal was seen in the no antibody control capture, much more prolamine mRNA was captured in the presence of RBP-A antibody. This indicates that RBP-A and prolamine mRNA are associated *in vivo*. In order to more clearly define the difference in the amount of prolamine mRNA between the control and experimental samples, PCR reactions were performed for a different number of cycles (22, 24, 26, 28 or 30). Figure 3B shows that prolamine PCR product is detectable after 22 amplification cycles for the experimental sample, but 28 amplification cycles are required before the same product is detected from the control sample.

In order to ensure that prolamine mRNA was not being captured non-specifically by RBP-A antibody, we used a similarly affinity purified GFP antibody at the same concentration as an additional control. Figure 4A shows that prolamine mRNA cannot be detected by RT-PCR from the GFP control immunoprecipitation and is only visible in the experimental sample. Indeed, it was observed that the addition of GFP antibody appears to block the background binding of RNA to the Protein A resin (data not shown). Analysis of the same captured materials by RT-PCR with intron spanning glutelin-specific primers gave essentially the same results (Figure 4B), with glutelin PCR products only being detected from the experimental sample. Collectively, these experiments show that RBP-A is associated with both prolamine and glutelin mRNA *in vivo*. It is not currently clear, however, RBP-A is present in complexes containing both prolamine and glutelin mRNAs or whether there are distinct populations of RBP-A interacting with either prolamine or glutelin mRNA separately. It should also be emphasized that it is not yet clear if the interactions of RBP-A with prolamine and glutelin RNAs are direct, or involve additional partner proteins.

Identification of RNA targets of RBP-A using rice 45K cDNA array

In order to determine which, if any, other RNA species are bound by RBP-A, RNAs isolated by RNA-IP using RBP-A (experimental) and GFP (control) antibodies were used to generate Cy5 and Cy3-labeled probes, respectively. These were then hybridized to whole genome rice microarrays source from the NSF Array project. RNA-IP using anti-GFP was chosen as the microarray control sample to allow the assessment of false positive signals resulting from RNAs binding to either to antibodies or protein A resins. Particularly of concern are the mRNA encoding storage proteins which are highly expressed in mid-developing rice seeds. Since the amounts of RNAs were predicted to be low for both experimental and control samples, probes were generated using two rounds of cDNA amplification. Array slides were scanned at a gain setting that yielded maximum signal with minimum background and spots whose intensities varied between control and experimental samples were detected and quantified using GenePix Pro 6.1.

As expected, the experimental sample yielded far more spots of higher intensity than were seen with the GFP control material. In fact, essentially no spots were detected on the control slides at an intensity significantly greater than background. Table 1 shows that several of the most abundant mRNAs associating with RBP-A were saturating under the gain settings used for slide scanning. This means than an absolute enrichment of signal intensity in the control (GFP) vs experimental (RBP-A) cannot be calculated. In future experiments it would be wise to hybridize a smaller amount of probe, or to reduce the hybridization time. Irrespective of this, as expected based on our previous RT-PCR data, we were able to show that both prolamine and glutelin mRNAs are binding to RBP-A *in vivo*. Altogether, a total of 115 distinct RNA species were identified as ligands for RBP-A. Based on previous microarray experiments, this represents

around 0.5 % of the total number of genes in seed material of this age. Amongst the RNA species interacting with RBP-A are seed storage protein transcripts, RNAs encoding glycine-rich RNA binding proteins and a number of ribosomal mRNAs.

The significance of RBP-A binding to non-storage protein RNAs is currently unclear, however, this data represents a very useful starting point for the analysis and identification of putative binding site consensus sequences. Preliminary analysis indicates that several consensus sites can be identified within the 3'UTR regions of many RBP-A ligands. Most importantly, these consensus sequences can be mapped to known *cis*-acting targeting sequences with prolamine mRNA (Morris and Wyrick unpublished data). Although this may not be surprising, since RBP-A was initially identified as a prolamine zipcode binding protein, it supports the validity of our method.

Discussion

In this study, we have developed a novel cross-linking procedure which allows reliable RNA-IP experiments to be performed using developing rice seed. This method was successfully used to meet our first aim of demonstrating that the hnRNP RBP-A is able to bind, or is present in complexes with, prolamine mRNA *in vivo*. RBP-A was previously identified as a prolamine zipcode binding protein which although primarily nuclear in localization, is also present in particles which associate with cortical microtubules and ER membranes. The fact that we were able to demonstrate that this protein is associated with both prolamine and glutelin mRNA indicates that it may have a general role in moving RNAs from the nucleus to the cortical region of the cell. Whether RBP-A will again be identified in glutelin zipcode binding assays is currently unclear, but the evidence we have to date suggests that this is quite possible. Similarly

further work will be required to identify the role of the microtubule-associated RBP-A. Perhaps microtubules act as a storage site for mRNAs prior to their anchoring on specific subdomains of the ER.

Figure 5 outlines the basic procedure used for material preparation prior to RNA-IP. Developing rice seeds were sectioned to 0.5 mm thickness and then transferred to DMA-containing buffer. Sections were rinsed with modified cytoskeleton stabilizing buffer and fixed with 0.1 % formaldehyde buffer for 10 min to maintain *in vivo* assembled nucleotide-protein complexes. In order to quench formaldehyde, sections were then washed with Tris-containing IP buffer prior to extraction for use in immunoprecipitation assay with affinity purified antibodies. RNAs from immunoprecipitated materials were treated with DNase I to remove genomic DNA contamination and RNA-protein cross linking were reversed by incubation at 60 ^{o}C for one hour during protein digestion with proteinase A. RNAs were then purified using TRIzol reagent and used for either RT-PCR or probe generation for microarray hybridization. By performing immunoprecipitations with affinity purified anti-RBP-A, and similarly affinity purified anti-GFP as a negative control, we were able to demonstrate that RNA-IP is a valid strategy either to establish whether specific RNAs are present in a RNA-proteins complex by RT-PCR or to determine which other RNA species are bound using microarray analysis. The duration of formaldehyde fixation and buffer components should be adjusted empirically when analyzing other RBPs using the same approach. Since this protocol results in the cross-linking of RNA-protein complexes, extracts can be prepared using higher concentrations of detergent or denaturant. Extracts can then be diluted as needed prior to immunoprecipitation. An alternative means of identifying the target RNA sequences of RBPs is to use the CLIP (cross-linking immunoprecipitation) method. CLIP uses UV cross-linking to stabilize *in vivo* assembled RNA-

protein complexes. These complexes are then treated with RNase and purified by immunoprecipitation. RNAs isolated in this way are then be analyzed using linker-mediated PCR followed by ligation and sequencing (Jensen and Darnell, 2008; Wang et al., 2009). This technique is powerful since it can identify the specific sequence of the RNA region that is bound by an RBP and can therefore identify binding consensus sequences. Of course, since the protein of interest may associate with other RNA binding proteins, it may be difficult to identify which protein is binding which RNA. In addition, UV radiation can only penetrate roughly two mammalian cell layers (Jensen and Darnell, 2008). Since plant cell walls contain phenolic and flavonoid compounds to protect them from ultraviolet radiation (Li et al., 1993), working with large rice seeds may mean it is difficult to give a uniform and appropriate level of cross-linking without overly dehydrating the seed material.

Identification of RBP-A ligands using microarrays revealed transcripts encoding storage proteins, cytoskeletal elements (actin and tubulin), ubiquitin-modifying enzymes, chaperones (PDI, PPI) and membrane transport components (SAR1A, ARF1). These results suggest that RBP-A is a general RNA binding protein, which binds to a subset of highly expressed genes and is involved in the regulation of protein synthesis, transport and turn-over. The association of RBP-A with tubulin mRNA is interesting, given that RBP-A and tubulin proteins co-localize in immunofluorescence micrographs.

The second major aim of this study was to identify all the RNAs associating with RBP-A *in vivo*. This was accomplished using our newly developed fixation and RNA-IP procedure followed by the reverse transcription, amplification, labeling and hybridization of captured RNAs to whole genome microarrays. Although this data should currently be regarded as preliminary it serves as a useful starting point for further analysis on both biochemical and

bioinformatics levels. In addition to its association with prolamine and glutelin (expected from our previous RT-PCR experiments), RBP-A also was seen to interact with globulin RNAs and a number of other classes of mRNAs.

Microarray analysis showed that RBP-A is present with not only mRNAs encoding storage protein (prolamine, globulin and glutelin), but also a number of other class of mRNAs encoding RNA binding proteins, defense-related proteins in addition to proteins involved in regulating protein synthesis and turnover. Computational analysis of these mRNA sequences to identify potential consensus binding sites, led to the identification of a number of motifs, many of which are present within known prolamine zipcode sequence (Morris and Wyrick, unpublished data). This is interesting, but perhaps not surprising, given that RBP-A was initially identified based on its ability to bind to a prolamine zipcode sequence under high stringency conditions.

In summary, our newly developed cross-linking method, in combination with immunoprecipitation, will allow us to analyze other candidate RBPs that are potentially involved in RNA localization in rice endosperm cells. With the ever-increasing efficiency of high-throughput of sequencing techniques, identification of target RNA sequences for many more RBPs is just a matter of time.

Acknowledgements

John Wyrick and Derek Pouchnik are thanked for the valuable advice while establishing the use of microarray experiments in our laboratory. This work was supported by grants from the National Science Foundation.

References

Baroni, T. E., Chittur, S. V., George, A. D., and Tenenbaum, S. A. (2008). Advances in RIP-chip analysis : RNA-binding protein immunoprecipitation-microarray profiling. *Methods Mol Biol.* **419**, 93-108.

Bolognani, F., and Perrone-Bizzozero, N. I. (2008). RNA-protein interactions and control of mRNA stability in neurons. *J Neurosci Res.* **86**, 481-489.

Cáceres, L., and Nilson, L. A. (2008). Translational repression of gurken mRNA in the Drosophila oocyte requires the hnRNP Squid in the nurse cells. *Dev Biol.* **326**, 327-334.

Choi, S. B., Wang, C., Muench, D. G., Ozawa, K., Franceschi, V. R., Wu, Y., and Okita, T. W. (2000). Messenger RNA targeting of rice seed storage proteins to specific ER subdomains. *Nature.* **407**, 765-767.

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

Crofts, A. J., Washida, H., Okita, T. W., Satoh, M., Ogawa, M., Kumamaru, T., and Satoh, H. (2005) The role of mRNA and protein sorting in seed storage protein synthesis, transport, and deposition. *Biochem. Cell Biol.* **83**, 728-737.

Delanoue, R., Herpers, B., Soetaert, J., Davis, I., and Rabouille, C. (2007). Drosophila Squid/hnRNP helps Dynein switch from a gurken mRNA transport motor to an ultrastructural static anchor in sponge bodies. *Dev Cell.* **13**, 523-538.

Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol*. **3**, 195-205.

Gilbert, C. and Svejstrup, J. Q. (2006). RNA immunoprecipitation for determining RNA-protein associations *in vivo*. *Curr Protoc Mol Biol.* Chapter 27: Unit 27.4.

Hamada, S., Ishiyama, K., Choi, S. B., Wang, C., Singh, S., Kawai, N., Franceschi, V. R., and Okita, T.W. (2003b). The transport of prolamine RNAs to prolamine protein bodies in living rice endosperm cells. *Plant Cell*. **10**, 2253-2264.

Hamada, S., Ishiyama, K., Sakulsingharoj, C., Choi, S. B., Wu, Y., Wang, C., Singh, S., Kawai, N., Messing, J., and Okita, T. W. (2003a). Dual Regulated RNA Transport Pathways to the Cortical Region in Developing Rice Endosperm. *Plant Cell.* **15**, 2265-2272.

Jensen, K. B., and Darnell, R. B. (2008). CLIP: crosslinking and immunoprecipitation of *in vivo* RNA targets of RNA-binding proteins *Methods Mol Biol.* **488**, 85-98.

Jung, K. H., Dardick, C., Bartley, L. E., Cao, P., Phetsom, J., Canlas, P., Seo, Y. S., Shultz, M., Ouyang, S., Yuan, Q., Frank, B. C., Ly, E., Zheng, L., Jia, Y., Hsia, A. P., An, K., Chou, H. H., Rocke, D., Lee, G. C., Schnable, P. S., An, G., Buell, C. R., and Ronald P. C. (2008) Refinement of light-responsive transcript lists using rice oligonucleotide arrays: evaluation of gene-redundancy. *PLoS One.* **3**, e3337.

Keene, J. D., Komisarow, J. M., and Friedersdorf, M. B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc.* **1**, 302-307.

Kim, D. H., Langlois, M. A., Lee, K. B., Riggs, A. D., Puymirat, J., and Rossi J. J. (2005). HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Res.* **33**, 3866-3874.

Kim, W. T., and Okita, T. W. (1988). Structure, Expression, and Heterogeneity of the Rice Seed Prolamines. *Plant Physiol.* **88**, 649-655.

Kroll, T. T., Swenson, L. B., Hartland, E. I., Snedden, D. D., Goodson, H. V., and Huber, P. W. (2009). Interactions of 40LoVe within the ribonucleoprotein complex that forms on the localization element of Xenopus Vg1 mRNA. *Mech Dev.* **126**, 523-38.

Lewis, R. A., Gagnon, J. A., and Mowry KL. (2008). PTB/hnRNP I is required for RNP remodeling during RNA localization in Xenopus oocytes. *Mol Cell Biol*. **28**, 678-686.

Li, J., Ou-Lee, T. M., Raba, R. Amundson, R. G., and Last, R. L. (1993). Arabidopsis Flavonoid Mutants Are Hypersensitive to UV-B Irradiation. *Plant Cell.* **5**, 171-179.

Li, X., Franceschi, V. R. and Okita, T. W. (1993). Segregation of storage protein mRNAs on the rough endoplasmic reticulum membranes of rice endosperm cells. *Cell.* **72**, 869-879.

Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J. F., Revil, T., and Chabot, B. (2007). hnRNP proteins and splicing control. *Adv Exp Med Biol.* **623**, 123-147.

Muench, D. G., Chuong, S. D., Franceschi, V. R. and Okita TW. (2000). Developing prolamine protein bodies are associated with the cortical cytoskeleton in rice endosperm cells. *Planta.* **211**, 227-238.

Muench, D. G., Wu, Y., Coughlan, S. J., and Okita, T. W. (1998). Evidence for a Cytoskeleton-Associated Binding Site Involved in Prolamine mRNA Localization to the Protein Bodies in Rice Endosperm Tissue. *Plant Physiol.* **116**, 559-569.

Niranjanakumari, S., Lasda, E., Brazas, R., and Garcia-Blanco, M. A. (2002). Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions *in vivo*. *Methods.* **26**, 182-190.

Norvell, A., Debec, A., Finch, D., Gibson, L., and Thoma, B. (2005). Squid is required for efficient posterior localization of oskar mRNA during Drosophila oogenesis. *Dev Genes Evol.* **215**, 340-349.

Percipalle, P., Raju, C. S., and Fukuda, N. (2009). Actin-associated hnRNP proteins as transacting factors in the control of mRNA transport and localization. *RNA Biol.* [Epub ahead of print]

Peritz, T., Zeng, F., Kannanayakal, T. J., Kilk, K., Eiríksdóttir, E., Langel, U., and Eberwine, J. (2006). Immunoprecipitation of mRNA-protein complexes. *Nat Protoc.* **1**, 577-580.

Schmitz-Linneweber, C., Williams-Carrier, R., and Barkan, A. (2005). RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. *Plant Cell.* **17**, 2791-2804

Terzi, L. C., and Simpson, G. G. (2009). Arabidopsis RNA immunoprecipitation. *Plant J.* [Epub ahead of print]

Wahl, M. C., Will, C. L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell.* **136**, 701-718.

Washida, H., Sugino, A., Kaneko, S., Crofts, N., Sakulsingharoj, C., Kim, D., Choi, S. B., Hamada, S., Ogawa, M., Wang, C., Esen, A., Higgins, T. J., and Okita, T. W. (2009) Identification of cis-localization elements of the maize 10 kD delta-zein and their use in targeting RNAs to specific cortical ER subdomains. *Plant J.* (Accepted)

Yang, Z., Edenberg, H. J., and Davis, R.L. (2005). Isolation of mRNA from specific tissues of Drosophila by mRNA tagging. *Nucleic Acids Res*. **33**, e148

Figure Legends

Figure 1. Establishment of cross-linking conditions using developing rice seeds

(A) Either whole or sectioned rice developing seeds were incubated with or without of protein-protein cross-linker 10 mM DMA for 0, 15, 30, 60 or 120 min as indicated. Samples were extracted with urea-containing SDS sample buffer and resolved by SDS-PAGE. The upper panel is a CBB-stained gel and the lower panel is a western blot probed with anti-RBP-A. Note that levels of RBP-A extracted do not differ between the samples, whilst total protein extracted from sectioned seeds decreases slightly at 120 min. Pre-stained molecular weight markers (M) were loaded and molecular weights in kDa are given on the left.

(B) Either whole or sectioned rice developing seeds were incubated with or without the reversible, nucleic acid-protein cross-linker, formaldehyde at 1 % for 0, 15, 30, 60 or 120 min as indicated. Samples were extracted with urea-containing SDS sample buffer and resolved by SDS-PAGE. The upper panel is a CBB-stained gel and lower panel is western blot probed with anti-RBP-A. Note that levels of both total protein and RBP-A extracted from the sectioned seeds decrease drastically with increasing cross-linking time. Pre-stained molecular weight markers (M) were loaded and molecular weights in kDa are given on the left.

(C) Sectioned seeds were incubated with 0.1 % formaldehyde for 0, 5, 10 and 20 min. Following washing, samples were extracted in IP buffer (total) and centrifuged to obtain soluble fraction (supernatant). Pellets were extracted with urea-containing SDS sample buffer. Pre-stained molecular weight markers (M) were loaded and molecular weights in kDa are given on the left.

Figure 2. Immunoprecipitation of RBP-A from formaldehyde cross-linked developing rice seed extracts

(A) Samples immunoprecipitated with and without extract and antibody (as indicated) were resolved by SDS-PAGE, blotted onto membrane and probed with anti-RBP-A antibody. Starting material (SM) was loaded as a positive control. Note that RBP-A is readily captured in the presence of both antibody and extract. A low intensity degradation product was detected around 26 kDa. Molecular weights are indicated in kDa.

(B) The same IP samples as in A were resolved by SDS-PAGE and silver-stained. Arrows indicate potential RBP-A interacting proteins. Note that abundant IgG heavy chain masks the RBP-A protein.

Figure 3. RT-PCR analysis of RNA-IP material shows that RBP-A associates with prolamine mRNA *in vivo*

(A) Following RNA-immunoprecipitation (RNA-IP), prolamine RT-PCR was performed on both the final (5th) wash sample and on the immunoprecipitated material as indicated and samples were analyzed by agarose gel electrophoresis and ethidium bromide staining. Unbound material and a water control were used as positive and negative controls respectively. Note that prolamine PCR products are much more abundant in the experimental reaction than in the negative controls. No PCR products were detected from the wash material.

(B) Prolamine RT-PCR was performed using different numbers of PCR cycles (as indicated) to allow comparison of real and background binding levels. Unbound material and water were used as positive and negative PCR controls respectively. Note that PCR amplification was linear until

28 cycles and that anti-RBP-A IP captured significantly more prolamine RNA than the background (resin + extract) control.

Figure 4. RBP-A associates with both prolamine and glutelin mRNA *in vivo*

(A) Following RNA immunoprecipitation using either GFP (control) or RBP-A antibodies, prolamine RT-PCR was performed using oligo(dT) cDNA as indicated. Total RNA and water were used as positive and negative PCR controls respectively.

(B) Following RNA immunoprecipitation using either GFP (control) or RBP-A antibodies, glutelin RT-PCR was performed using oligo(dT) cDNA as indicated. Total RNA and water were used as positive and negative PCR controls respectively.

Figure 5. RNA-immunoprecipitation procedure

Husks were removed from freshly harvested developing rice seeds (1), and they were sliced into 0.5 mm thick sections (2). These were collected in protein-protein cross-linker solution prior to nucleic acid-protein cross-linking by formaldehyde to maintain *in vivo* assembled ribonucleoproteins (RNP) complexes (3). Excess formaldehyde was quenched by Tris (4) and sections were then extracted (5) for use in immunoprecipitation using affinity purified antibodies (6). Following extensive washing, RNAs were treated with DNase to remove possible contaminating genomic DNA. Samples were then protease-treated whilst formaldehyde cross-link were reversed. Finally RNAs were purified (7) and used for either RT-PCR or microarray analysis (8 and 9). For RT-PCR, oligo(dT) primed cDNA (8) was generated and PCR amplified using gene specific primers prior to agarose gel electrophoresis (9). For microarray analysis, dye

coupled, amplified cDNA probes were generated (8) and hybridized to rice 45K oligo microarray slides (9).

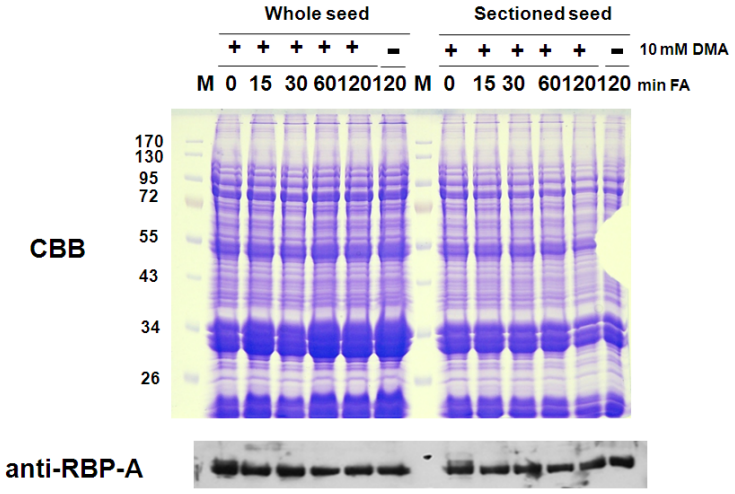
Table legends

Table 1. List of RNA species associated with RBP-A *in vivo*

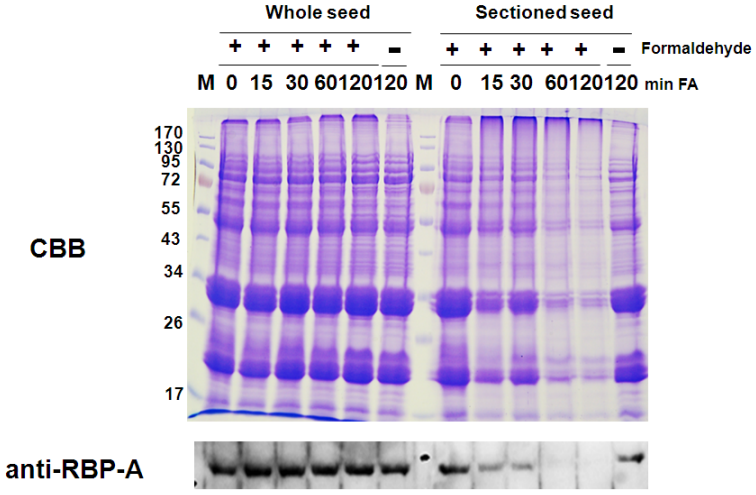
RNA species associated with RBP-A are listed in the order intensities (high to low). Mean background subtracted signal intensities from RNA-IP with anti-RBP-A and anti-GFP control are indicated, together with the fold ratio calculated as RBP-A signal / GFP signal. Note that signals from spots with intensities greater than 45000 were at least partially saturated and therefore fold ratios are expressed as minimums. The actual fold ratios may be significantly higher in reality. RNAs encoding seed storage proteins (green), glycine rich RNA binding proteins (yellow), cytoskeletons (purple), ubiquitination related proteins (orange) and ribosomal proteins (blue) are highlighted.

Figure 1. Establishment of cross-linking conditions using developing rice seeds

(A)



(B)



(C)

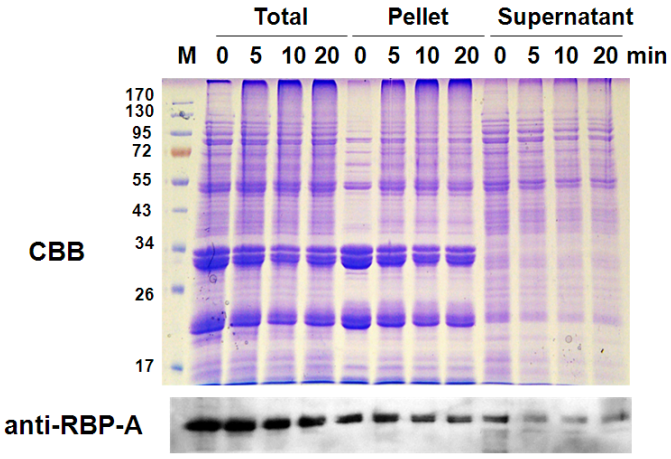
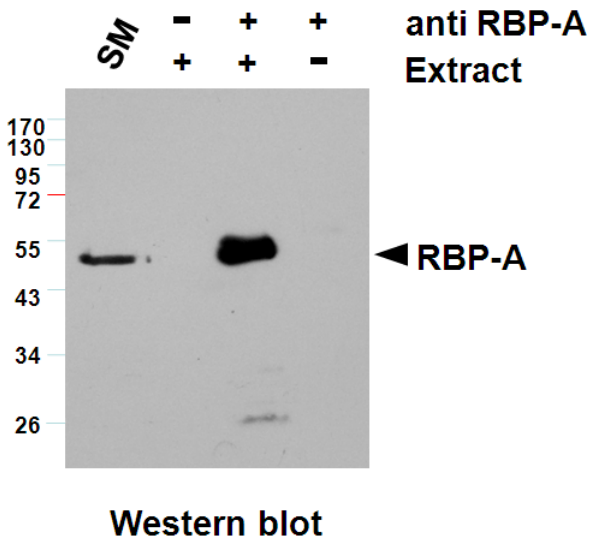


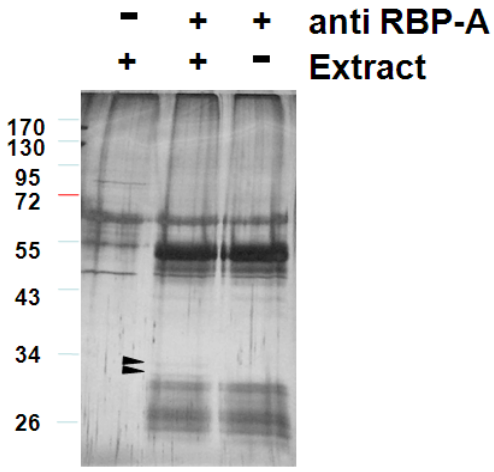
Figure 2. Immunoprecipitation of RBP-A from formaldehyde cross-linked developing rice seed extracts

(A)



Western blot

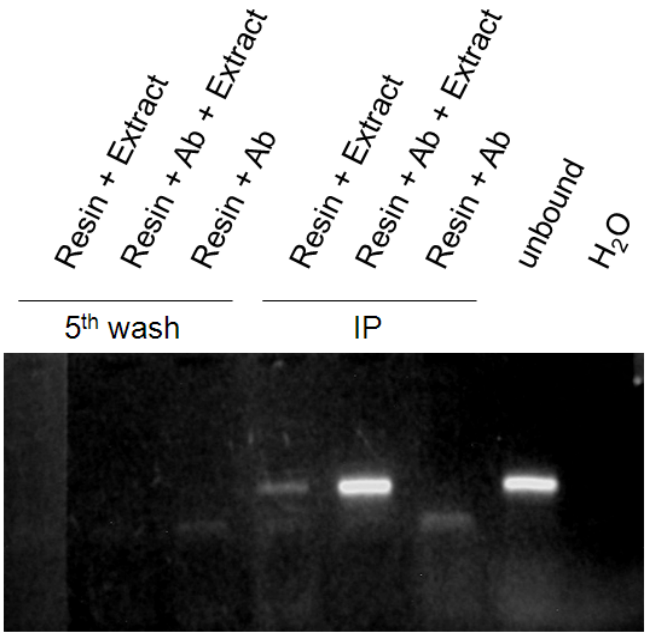
(B)



Silver stain

Figure 3. RT-PCR analysis of RNA-IP material shows that RBP-A associates with prolamine mRNA *in vivo*

(A)



(B)

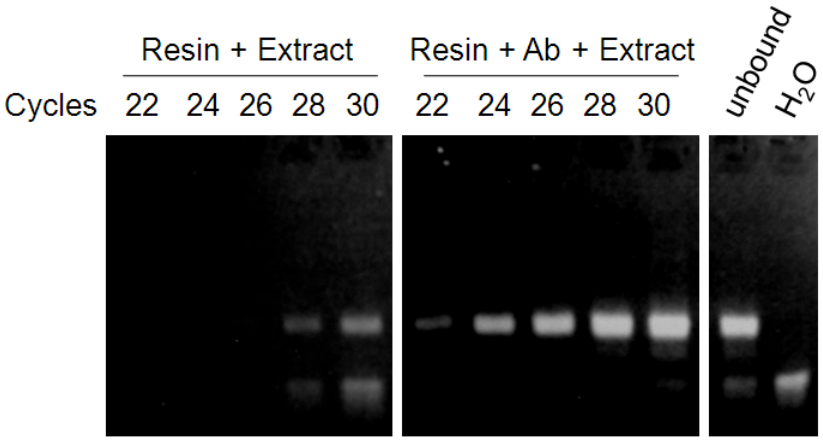
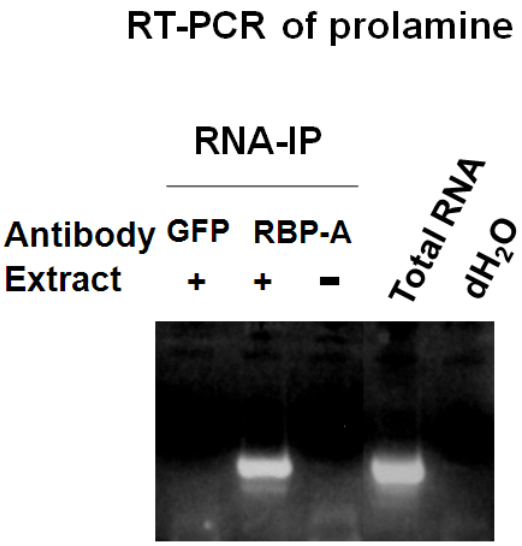


Figure 4. RBP-A associates with both prolamine and glutelin mRNA *in vivo*

(A)



(B)

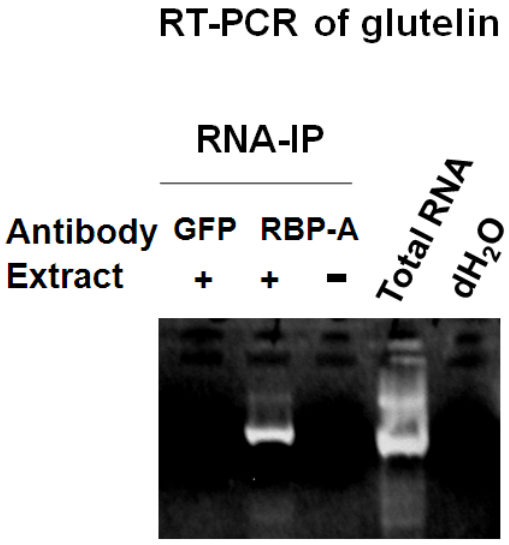


Figure 5. RNA immunoprecipitation (RNA-IP) procedure

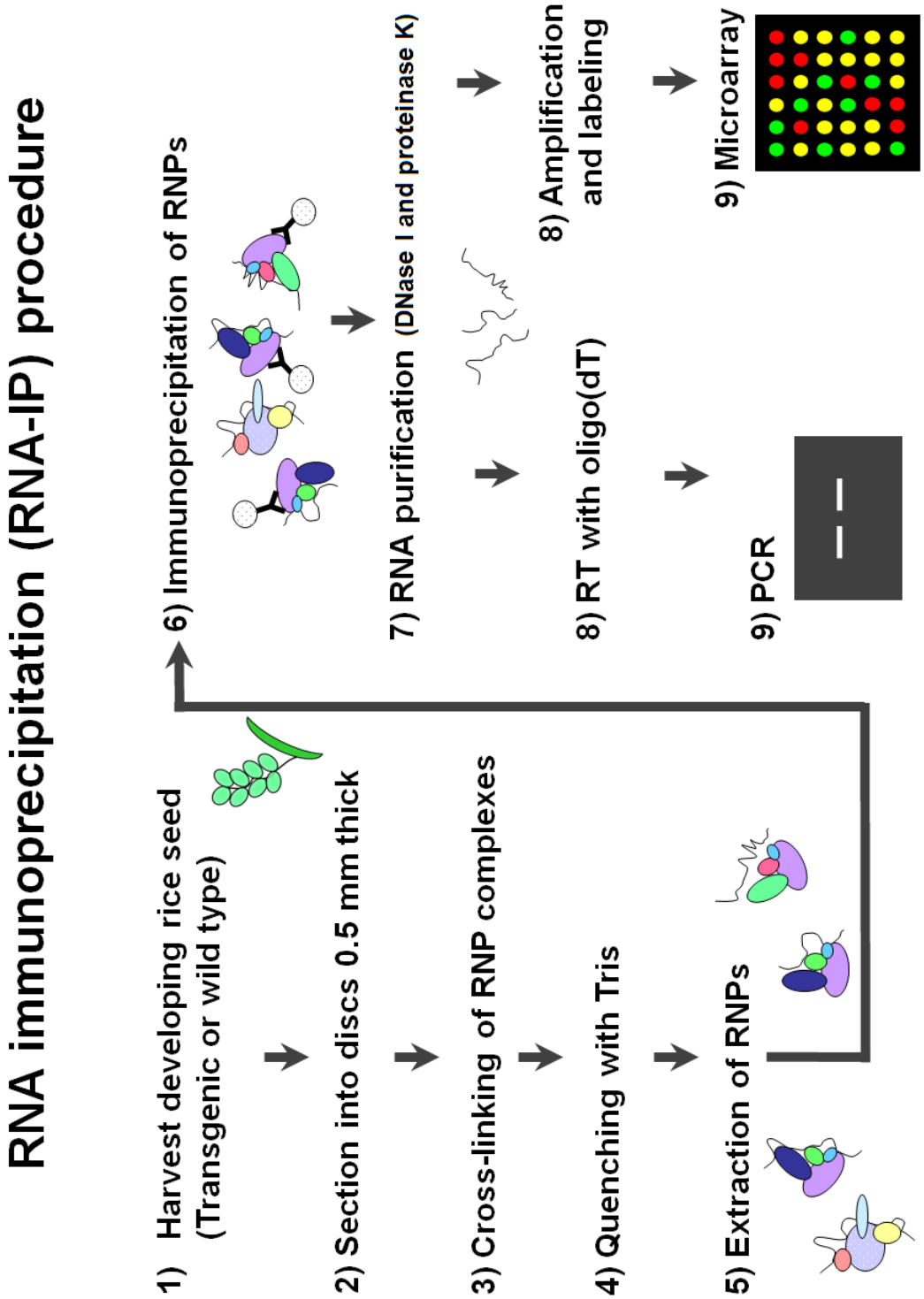


Table 1. List of RNA species associated with RBP-A *in vivo*

Annotation	ID	Ratio	RBP-A	GFP
Glycine-rich RNA-binding protein GRP1A	TR067076	315	65144	207
19 kDa globulin precursor	TR063837	324	62601	193
1,4-alpha-glucan branching enzyme IIB	TR037182	281	60048	214
NAC domain-containing protein 77	TR031876	267	58820	220
Seed allergenic protein RAG2 precursor	TR048267	285	55894	196
Prolamin precursor	TR048206	307	55320	180
Retrotransposon	TR071149	218	53712	246
Metallothionein family	TR061512	244	51926	213
Glutelin type-A 2 precursor	TR036729	214	48973	229
N/A	TR066353	297	45993	155
Oleosin 18 kDa	TR042579	240	44611	186
Flower-specific gamma-thionin precursor	TR039389	233	43088	185
Seed specific protein Bn15D17A	TR031972	151	34291	227
Polyubiquitin containing 7 ubiquitin monomers	TR058800	143	33885	237
Senescence-associated protein	TR056926	184	31625	172
Translationally-controlled tumor protein	TR069752	147	31428	214
Aspartic proteinase oryzasin-1 precursor	TR032949	170	30458	179
Plasma membrane associated protein	TR048887	158	29461	186
Embryonic abundant protein 1 related cluster	TR062914	173	28715	166
Endothelial differentiation-related factor 1	TR052216	154	27870	181
OsGrx_C2.2 - glutaredoxin subgroup I	TR046201	143	24782	173
Expressed protein	TR058832	163	23406	144
SKP1-like protein 1B	TR061159	141	23350	166
Prolamin precursor	TR048205	141	22543	160
Legumin precur	TR050824	110	22137	202
Tubulin alpha-3 chain	TR042786	144	21112	147
AWPM-19-like family protein	TR055132	109	19013	175
Tubulin alpha-3 chain	TR073189	78	18575	238
40S ribosomal protein S9	TR073240	76	18508	245
Transposon	TR056425	115	18215	158
Nonspecific lipid-transfer protein 2	TR069519	116	17452	150
Prolamin	TR069014	98	16815	171

Annotation	ID	Ratio	RBP-A	GFP
Tesmin/TSO1-like CXC domain	TR063902	91	16338	180
Expressed protein	TR063293	90	16202	181
B12D protein	TR041939	114	16138	141
EF-hand Ca2+-binding protein CCD1	TR058811	115	15361	133
AN1-like Zinc finger family protein	TR058385	84	15245	182
Mono- or diacylglycerol acyltransferase	TR072328	78	14102	180
60S acidic ribosomal protein P1	TR050747	83	13821	166
Glutelin type-A 2 precursor	TR054716	68	13793	203
Plasma membrane ATPase 1	TR048114	73	13664	188
Linker histone H1 and H5 family protein	TR056141	61	13340	219
Protein disulfide-isomerase precursor	TR073168	69	13338	194
Protease inhibitor/seed storage/LTP family	TR057833	69	12951	189
Peptidyl-prolyl cis-trans isomerase	TR035252	84	12774	152
Early nodulin 93	TR056228	69	11961	174
Cupin family protein	TR040854	75	11862	159
Expressed protein	TR046902	68	11802	174
Trypsin/factor XIIA inhibitor precursor	TR048296	68	11301	167
60S ribosomal protein L10-3	TR072653	63	11037	175
Dehydration stress-induced protein	TR056765	57	10799	191
Nonspecific lipid-transfer protein precursor	TR039242	55	10738	196
ADP-ribosylation factor	TR063760	63	10698	170
Stem-specific protein TSJT1	TR064796	64	10678	167
Ubiquitin-conjugating enzyme E2-17 kDa	TR071570	49	10507	213
60S ribosomal protein L34	TR060386	52	9853	190
Ethylene-responsive protein	TR042977	55	9840	179
Globulin-1 S allele precursor	TR043283	61	9532	156
Stem-specific protein TSJT1	TR067399	42	9474	228
Ubiquitin-conjugating enzyme E2-17 kDa	TR070426	51	9251	182
Actin-7	TR034311	49	9190	189
40S ribosomal protein SA	TR039761	63	8999	143
Chorismate mutase, chloroplast precursor	TR066727	41	8803	214
Seed allergenic protein RAG2 precursor	TR048268	53	8615	164
N/A	TR060833	47	8215	175

Annotation	ID	Ratio	RBP-A	GFP
60S ribosomal protein L13-2	TR041773	36	8148	227
Inositol-tetrakisphosphate 1-kinase 3	TR042787	39	8103	210
60S ribosomal protein L44	TR071724	37	8060	220
N/A	TR070909	39	7925	203
40S ribosomal protein S5	TR068745	44	7750	177
Xylanase inhibitor protein 2 precursor	TR069963	41	7626	186
Ubiquitin-fold modifier 1 precursor	TR034980	43	7280	169
Uncharacterized plant-specific domain	TR063671	33	7098	218
GTP-binding protein SAR1A	TR066613	43	7076	165
Translocase of chloroplast 34	TR040177	31	7072	226
14-3-3-like protein B	TR037480	47	7061	150
Disease resistance protein RPM1	TR048162	40	7038	174
Calmodulin	TR063775	41	6907	169
Expressed protein	TR055643	40	6853	172
14-3-3-like protein S94	TR042670	39	6850	177
Peptidase family M1 containing protein	TR072028	37	6763	184
Expressed protein	TR050072	40	6655	166
F-box domain containing protein	TR066422	27	6621	242
Late embryogenesis abundant protein	TR072834	39	6544	168
Expressed protein	TR061525	31	6499	213
60S ribosomal protein L10-1	TR067872	43	6483	150
60S acidic ribosomal protein P0	TR050845	37	6464	175
Expressed protein	TR060906	38	6445	169
MRNA capping enzyme	TR067850	38	6378	170
Eukaryotic translation initiation factor 5A	TR049935	35	6333	180
Expressed protein	TR072447	41	6317	153
Protease inhibitor/seed storage/LTP	TR063676	40	6288	158
Lysosomal Pro-X carboxypeptidase precursor	TR033652	36	6244	172
AP2 domain containing protein	TR063819	45	6141	138
Glycine-rich protein 2b	TR050834	31	6109	197
Phosphoglycerate kinase, cytosolic	TR070581	34	6094	177
Ethylene-responsive transcription factor 4	TR046894	40	6033	152
Hypothetical protein	TR055218	36	5881	163

Annotation	ID	Ratio	RBP-A	GFP
Actin-7	TR070460	32	5859	185
60S ribosomal protein L26-1	TR064824	29	5830	203
DNA-binding protein MNB1B	TR072452	35	5824	166
Expressed protein	TR047195	33	5787	173
Glutelin type-B 4 precursor	TR036094	30	5729	194
Elongation factor 1-alpha	TR039730	35	5654	162
60S ribosomal protein L39	TR038892	31	5553	180
Oleosin family protein	TR064395	29	5498	187
60S ribosomal protein L21	TR055140	32	5477	173
40S ribosomal protein S27a	TR031479	31	5393	172
60S ribosomal protein L2	TR066663	39	5387	137
40S ribosomal protein S20	TR053957	26	5365	206
Expressed protein	TR067018	37	5355	144
Ubiquitin-conjugating enzyme E2 M	TR072501	29	5281	183
60S ribosomal protein L23	TR055149	27	5214	191
Lipid binding protein	TR043286	28	5106	181
Polyubiquitin containing 7 ubiquitin monomers	TR071548	34	5048	147
40S ribosomal protein S14	TR070687	26	5030	195