RNA Binding Motif Protein 48 is required for U12 splicing and maize endosperm differentiation

Fang Bai[1,4], Jacob Corll[2,4], Donya Shodja[2], Ruth Davenport[3], Guanqiao Feng[3], Janaki Mudunkothge[1], Christian J. Brigolin[2], Federico Martin[1], Gertraud Spielbauer[1], Chi-Wah Tseung[1], Amy E. Siebert[2], W. Brad Barbazuk[3,*], Shailesh Lal[2,*], and A. Mark Settles[1,5,*]

Affiliations

[1]Horticultural Sciences Department and Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, Florida, USA

[2]Department of Biological Sciences, Oakland University, Rochester, Minnesota, USA

[3]Department of Biology and Genetics Institute, University of Florida, Gainesville, Florida, USA

[4]These authors contributed equally: Fang Bai and Jacob Corll

[5]Lead contact: A. Mark Settles

*Correspondence: bbarbazuk@ufl.edu; lal@oakland.edu; settles@ufl.edu

Abstract

The last eukaryotic common ancestor had two classes of introns that are still found in most eukaryotic lineages. Common U2-type and rare U12-type introns are spliced by the major and minor spliceosomes, respectively. Relatively few splicing factors have been shown to be specific to the minor spliceosome. We found that the maize RNA Binding Motif Protein48 (RBM48) is a U12 splicing factor that functions to promote cell differentiation and repress cell proliferation. RBM48 is coselected with the U12 splicing factor, ZRSR2/RGH3. Protein-protein interactions between RBM48, RGH3, and U2 Auxiliary Factor (U2AF) subunits suggest major and minor spliceosome factors may form complexes during intron recognition. Human RBM48 interacts with ARMC7. Maize RBM48 and ARMC7 have a conserved protein-protein interaction. These data predict that RBM48 is likely to function in U12 splicing throughout eukaryotes and that U12 splicing promotes endosperm cell differentiation in maize.

Keywords

U12 splicing, minor spliceosome, maize, endosperm, cell differentiation, cell proliferation, stem cell

Introduction

Eukaryotic protein-coding genes generally are interrupted by introns, and splicing of pre-mRNA is fundamental for protein expression. Eukaryotic introns originate from bacterial, self-splicing Group II introns (Burge et al., 1998; Roy and Irimia, 2009). However, eukaryotic introns are no longer self-splicing, and dynamic RNA-protein complexes, known as spliceosomes, direct accurate recognition of splice sites (Simpson and Filipowicz, 1996; Brown, 1998; Lorkovic et al., 2000; Ru et al., 2008; Rappsilber et al., 2002). The last eukaryotic common ancestor had two classes of introns that were spliced by different spliceosome complexes (Irimia and Roy, 2014). The vast majority of introns, termed U2-type introns, are spliced by the major spliceosome (Lee and Rio, 2015). There are also rare, U12-type introns, which are spliced by the minor spliceosome (Will and Luhrmann, 2005). U12-type introns have a higher degree of sequence conservation at the 5' splice site and branch point sequence (BPS) relative to U2-type introns (Turunen et al., 2013).

The minor spliceosome is at a low concentration in the cell, and splicing of U12-type introns is significantly slower than U2-type introns both *in vitro* and *in vivo* (Tarn and Steitz, 1995; Patel et al., 2002; Montzka and Steitz, 1988). Consequently, splicing of U12-type introns can limit the abundance of mature transcripts from minor intron-containing genes (MIGs) (Konig et al., 2007; Patel et al., 2002; Niemela et al., 2014; Niemela and Frilander, 2014; Younis et al., 2013). In human cells, U12 splicing efficiency is responsive to a stress-activated signal transduction pathway (Younis et al., 2013). These data suggest that U12 splicing efficiency could be a deeply conserved regulatory mechanism. However, evolutionary evidence indicates the minor spliceosome

is dispensable in many eukaryotes. The minor spliceosome has been lost in multiple independent lineages including model organisms such as *C. elegans, S. cerevisiae, D. discoideum,* and *C. reinhardtii* (Dávila et al., 2008). Among metazoan species with U12 splicing, the number of MIGs varies drastically, ranging from 758 in humans to only 15 in *D. melanogaster* (Alioto, 2006). Loss of U12-type introns is the most common mechanism for reduction of MIGs within a genome (Burge et al., 1998).

Despite species variation for the presence and number of MIGs, U12 splicing appears to have roles in cell proliferation, cell differentiation, and development across divergent eukaryotes. A striking example is revealed by mutations of ZRSR2 orthologs in human and maize. Somatic mutations in human *ZRSR2* primarily affect U12 splicing and cause myelodysplastic syndrome, a disease that reduces differentiation of mature blood cell types (Madan et al., 2015; Gault et al., 2017). The *rgh3* mutation of the maize ZRSR2 ortholog has a conserved U12 splicing function and inhibits differentiation of endosperm cells (Fouquet et al., 2011). Other U12 splicing mutants in humans and mouse models demonstrate roles in hematopoiesis, muscle strength, as well as bone and neural development (Edery et al., 2011; He et al., 2011; Horiuchi et al., 2018). A mutation in zebrafish indicates U12 splicing functions in intestinal epithelium and endodermal organ development (Markmiller et al., 2014). Arabidopsis knockdowns of minor spliceosome proteins show defects in leaf development and flowering time (Kim et al., 2010; Jung and Kang, 2014; Xu et al., 2016). The different pleiotropic phenotypes of U12 splicing factors have made it difficult to identify unifying developmental functions for MIGs.

Moreover, there are relatively few splicing factors that have been identified as specific to the minor spliceosome. Biochemical characterization of small nuclear

ribonucleoproteins (snRNPs) showed that most minor spliceosome proteins are shared with the major spliceosome with only seven novel proteins uniquely associated with the minor spliceosome (Schneider et al., 2002; Will et al., 2004). Genetic analysis of human ZRSR2, human FUS, mouse SMN1, and maize RGH3 indicates that these factors predominantly affect U12-type intron splicing even though there is substantial evidence for the proteins to participate in other RNA metabolic processes (Gault et al., 2017; Madanieh et al., 2015; Reber et al., 2016). Here we identify the maize RNA Binding Motif Protein (RBM48) as a minor spliceosome factor which functions to promote cell differentiation and repress cell proliferation. RBM48 is conserved in organisms that retain the minor spliceosome predicting that this protein will have a function in U12 splicing throughout eukaryotes. Protein-protein interactions and co-localization between RBM48, RGH3, and U2 Auxiliary Factor (U2AF) subunits suggests major and minor spliceosome factors may form complexes as part of recognizing introns.

Results

Loss of RBM48 causes a rough endosperm (rgh) defective kernel phenotype

The *rbm48-umu1* reference allele was identified from a collection of 144 UniformMu *rgh* mutants (Fouquet et al., 2011; McCarty et al., 2005). The *rgh* locus was mapped to the long arm of chromosome 4 with bulked segregant analysis (Liu et al., 2010) and compared to *Robertson's Mutator* transposon flanking sequence tags from the *rgh* mutant line (Fig. S1). The *rbm48-umu1* insertion was the only novel insertion in this line that co-segregated with the *rgh* phenotype (Fig. S1). The *rbm48-umu2* allele was obtained from the Maize Genetics Co-operative Stock Center (McCarty et al., 2013).

Self-pollination of *rbm48-umu2* heterozygotes also segregate for an *rgh* kernel phenotype (Fig. S1). Both alleles show similar defective kernel phenotypes and segregate at ratios consistent with a recessive mutant (Fig. 1; Table S1). The *rbm48-umu1* allele transmits fully through both male and female gametes but does not give a seed phenotype when crossed to normal inbred lines (Table S2). Crosses of the two alleles failed to complement the *rgh* mutant phenotype indicating that *rbm48* mutations disrupt maize seed development (Fig. S2).

Mature *rbm48* kernels have reduced endosperm size and embryos typically fail to develop (Fig. 1C-L). Consistent with these morphological defects, mutant kernel composition is affected with increased protein and reduced oil, starch, and seed density (Fig. S2). A small fraction of *rbm48* kernels have a less severe phenotype and develop a viable embryo that germinates. Mutant seedlings developed only 1-2 narrow leaves, stunted roots, and died around 20 days after sowing (Fig. S2).

RT-PCR of normal *Rbm48* cDNA from etiolated roots and shoots of W22, B73, and Mo17 inbred seedlings identified a common transcript isoform coding for a 219 aa protein (Fig. 1M; Fig. S3). Alternatively spliced isoforms had premature termination codons that are predicted to be targets of nonsense mediated decay (Shaul, 2015). The predicted RBM48 protein has an N-terminal RNA recognition motif (RRM) specific to the RBM48 protein family and a C-terminal RS-rich domain (Fig. 1M). This domain structure is common for SR proteins involved in pre-mRNA splicing (Graveley, 2000). We were not able to detect *rbm48* transcripts in mutant seedlings for either allele, and we infer that the two alleles are likely null mutations (Fig. 1N).

Based on the NCBI Conserved Domains Database (Marchler-Bauer et al., 2017), the RRM domain of RBM48 is found in 320 eukaryotic species (Fig. 1O). The RRM domains of RBM48 and ZRSR2/RGH3 appear to be coselected with 96% of species having an RBM48 domain also containing a RGH3 RRM domain. By contrast, only 72% of species with an RBM48 RRM domain also have the RRM domain for the core U2 splicing factor, U2AF1. Like RGH3, the RBM48 RRM domain is not found in model organisms that have lost MIGs and the U12 splicing machinery. Phylogenetic analysis of 16 representative species shows the broad distribution of RBM48 orthologs across multiple eukaryotic kingdoms (Fig. 2). Interestingly, RBM48 was lost in *D. melanogaster*, a species with a highly reduced number of MIGs. These phylogenetic data suggest a potential role for RBM48 in U12 splicing.

Aberrant splicing of U12-type introns in rbm48

RNA splicing was assessed with mRNA-seq comparing mutant and normal sibling endosperm RNA from both *rbm48* alleles. To quantify the extent of intron splicing defects, we calculated the percent spliced out (PSO) of individual introns using exon-exon junction and intron reads (Table S3). PSO compares the density of exon-exon junction reads and intron reads for each intron (Fig. S4). A PSO of 100 indicates complete splicing of an intron, and a value of 0 indicates all expressed transcripts retain the intron. The difference in PSO values for individual introns between normal siblings and *rbm48* mutants (ΔPSO) identifies introns that are differentially impacted. The major U2-type introns of the B73_v2 filtered gene set are largely unaffected with only 3-5% of introns in expressed genes have a ΔPSO >20% (Fig. 3A-B and S4). By contrast, 65% and

53% of U12-type introns in *rbm48-umu1* and *rbm48-umu2*, respectively, have a ΔPSO >20% suggesting more than half of U12-type introns are retained in *rbm48* mutants.

For example, the U12-type intron in GRMZM2G131321 has a ΔPSO of 65% for *rbm48-umu1*, which corresponds to a high density of intron reads in mutant RNA-seq data (Fig. 3C). A semi-quantitative RT-PCR assay spanning this U12-type intron shows splicing defects with increased intron retention in both endosperm and shoot tissues (Fig. 3D). RT-PCR assays confirmed eleven predicted U12-type splicing defects (Fig. 3D, 4B, S5, and S6). Hierarchical clustering of U12-type introns based on PSO values for normal and *rbm48* mutant alleles show highly overlapping effects on intron retention, except that the *rbm48-umu1* allele has slightly stronger defects (Fig. 3E).

Fisher's exact tests of intron and exon-exon junction read counts were used to identify statistically significant splicing differences for individual introns (Table S3). Approximately 60% of expressed MIGs had significant U12-type intron retention in either allele. U2-type introns were largely unaffected with only 7% and 10% of U2-type introns in *rbm48-umu1* and *rbm48-umu2* having significant differences relative to normal. The median ΔPSO in these significant U2-type introns is 4%, while mis-spliced U12-type introns show a median ΔPSO of 44% and 29% in *rbm48-umu1* and *rbm48-umu2* alleles, respectively (Fig. S4). These data further indicate *rbm48* mutants confer large magnitude effects on the majority of U12-type introns with rare effects on U2 splicing.

We also compared the two *rbm48* alleles with prior mRNA-seq analysis of *rgh3* (Gault et al., 2017). Of the 242 U12-type introns tested in all three mutant alleles, 196 (77%) were significantly affected in all experiments (Fig. 4A). Direct RT-PCR

comparisons of mutant seedling tissues confirmed similar U12 splicing defects in *rgh3* and *rbm48-umu1* (Fig. 4B and S6). These analyses indicate that *rgh3* and *rbm48* have extensive overlap in U12 splicing defects.

rbm48 disrupts endosperm cellular development

Differential gene expression analysis of the *rbm48* RNA-seq data identified large scale changes. Of 16,745 transcripts expressed at least one transcript per million (TPM) in one genotype, 5,771 transcripts were differentially expressed (FDR <0.05, >2-fold change) in either *rbm48-umu1* or *rbm48-umu2* (Fig. S7; Table S4). There were 1,811 up-regulated and 766 down-regulated transcripts in both alleles. A subset of these gene expression differences were confirmed with semi-quantitative RT-PCR (Fig. 4C). GO term enrichment analysis identified nutrient reservoir activity as the most significantly enriched term among differentially expressed genes (Table S5).

Nearly 80% the nutrient reservoir activity genes affected are seed storage proteins, such as zeins. At this stage of endosperm development, zein expression accounts for 45% of the total transcriptome in the normal sibling endosperm and only 20-30% of the total transcriptome in *rbm48* mutants (Fig. 4D). We confirmed that α-zein expression was reduced in *rbm48-umu1* using qRT-PCR in a developmental series of whole kernel RNA (Fig. 4E). The *rgh3* mutant also reduces and delays zein accumulation in the endosperm as part of an overall disruption of endosperm cell differentiation (Fouquet et al., 2011).

A focused analysis of endosperm cell type marker genes revealed that cell-type specific genes had aberrant expression from 10-13 days after pollination (DAP). Cell-

type markers for the aleurone (*Al9*), embryo surrounding region (*Esr1*), and basal endosperm transfer cell layer (*Betl2*, *Tcrr1*) had increased expression late in development (Fig. 4F, 4G, and S7).

Many of these whole kernel expression patterns were strikingly different from those observed from the lower half of kernels analyzed in the *rgh3* mutant (Fouquet et al., 2011). For example, *Esr1* is expressed at a basal level in the lower half of *rgh3* mutants, while *Esr1* is up regulated in whole *rbm48* kernels. Extractions of the upper and lower half of 12 DAP kernels showed that this discrepancy is due to reduced expression of *Esr1* in the lower half of the kernel, where *Esr1* is normally expressed, and ectopic expression of *Esr1* in the upper half of the kernel (Fig. 4H). Quantitative expression analysis of the lower half of 12 DAP kernels also showed that, while *Al9* is induced, *Betl2* and *Esr1* are reduced consistent with similar observations of *rgh3* (Fig. S7). These results indicate that *rbm48* endosperm cells fail to express cell type specific markers in the correct domains of the kernel.

Cellular phenotypes of *rbm48* endosperm are also consistent with cell differentiation defects. Transverse sections of 12 DAP kernels show that *rbm48* starchy endosperm cells are smaller than normal siblings (Fig. 5A and 5B). At this stage, the basal endosperm transfer layer (BETL) is differentiated in normal kernels, visible as elongated cells with intense Schiff's staining due to secondary cell wall ingrowths (Fig. 5C). The BETL in *rbm48* is not developed into elongated cells (Fig. 5D). Cell-type specific reporter lines indicate that the *rbm48* BETL region differentiates into aleurone. A GUS reporter gene driven by the *Vivparous1* promoter (pVP1:GUS) marks ABA responsive cells including the aleurone and embryo, while the *Betl1* promoter

(pBetl1:GUS) marks BETL cells. These reporter transgenes have aberrant expression in *rbm48-umu1* mutants with pVp1:GUS expressing in the basal epidermal endosperm and pBetl1:GUS having low expression in the same region (Fig. 5E-H and S8). These data further support the model that epidermal endosperm cells express aleurone markers regardless of position in the endosperm.

Mutant *rgh3* endosperm retains the ability to proliferate in cell culture at late time-points in seed development (Fouquet et al., 2011). Endosperm tissue culture responses for *rbm48-umu1* mutants and normal siblings were tested. At 6 DAP, normal and mutant endosperm grow equally well in the culture assay, but at 9-18 DAP, only *rbm48* endosperms are able to grow as a callus in the culture assay (Fig. 5I-J and S8). Thus, *rbm48* mutants have prolonged endosperm cell proliferation and delayed or aberrant differentiation, similar to *rgh3* mutants.

RBM48 interacts with U12 and U2 splicing factors

The U12 splicing and endosperm cell differentiation defects in *rbm48* and *rgh3* mutants raises the possibility that the proteins interact. We investigated the subcellular localization of RBM48 with GFP fusion proteins that were transiently expressed in *Nicotiana benthamiana* leaves. Full length RBM48 is localized to the nucleus and is associated with the nuclear speckles (Fig. S9). Deletion of the RRM domain still localized the protein in nuclear speckles, albeit at a lower expression level, while deletion of the RS region resulted in diffuse cytoplasmic signal (Fig. S9). These results are consistent with the localization of SR proteins in both plants and animals to sub-nuclear

speckles via the RS domain (Mori et al., 2012; Saitoh et al., 2012; Tillemans et al., 2005; Tillemans et al., 2006; Cazalla et al., 2002).

Co-expression of GFP-RBM48 with RFP-tagged RGH3 or the U2AF1 and U2AF2 subunits of U2AF resulted in extensive overlap of fluorescent signals (Fig. 6A). Human U2AF2 interacts with U2AF1 and the RGH3 ortholog, ZRSR2 (Tronchere et al., 1997; Kielkopf et al., 2001). The co-localization of RBM48 with these factors suggests that RBM48 may interact in a larger complex. Human U2AF is involved in the selection of 3' splice sites of U2-type introns, while ZRSR2 binds the 3' splice sites of U12-type introns (Tronchere et al., 1997; Guth et al., 2001; Shen et al., 2010).

Bimolecular fluorescence complementation (BiFC) assays of RBM48 with U2AF1, U2AF2, and RGH3 support the model of a larger protein complex of these factors. BiFC tests for fluorescence complementation of split, non-fluorescent fragments of YFP that are brought in close vicinity by interacting fusion proteins (Citovsky et al., 2006; Kerppola, 2008). The N-terminal or C-terminal fragments of YFP were fused with RBM48 and transiently co-expressed in different pairwise combinations with N- or C-terminal YFP fusion constructs of U2AF1, U2AF2, or RGH3 in *N. benthamiana* leaves. All combinations of RBM48 with full length U2AF1, U2AF2, and RGH3 resulted in YFP signal indicating these proteins are in close proximity in plant cells (Fig. 6B and S9). An in-frame deletion of the RGH3 RRM domain (RGH3ΔUHM), also known as the U2AF homology motif (UHM), fails to complement YFP indicating that the UHM domain is needed for RBM48 and RGH3 association (Fig. 6B). When expressed with U2AF2 fusions, RGH3ΔUHM can give BiFC signal but GFP-RGH3ΔUHM does not co-localize with U2AF2-RFP in the nucleus (Gault et al., 2017).

We then tested direct protein-protein interactions with *in vitro* pull-down assays. Recombinant GST-His-RBM48 fusion protein was immobilized to glutathione magnetic beads and incubated with *E. coli* lysate containing recombinant His-tagged U2AF1, U2AF2, RGH3, or Armadillo Repeat Containing7 (ARMC7, GRMZM2G106137). Human RBM48 interacts with ARMC7 (Hart et al., 2015). In all cases, RBM48 can purify the His-tagged proteins from bacterial lysates (Fig. 7A). Moreover, GST-His-RGH3 pulls down His-tagged U2AF1 and U2AF2, but not Cbes_2310, a protein originating from *Caldicellulosiruptor bescii* (Fig. 7B). These results support conserved interactions in maize for RBM48 with ARMC7 and RGH3 with U2AF2. In addition, we found that RGH3 and RBM48 interact with each other as well as U2AF1.

Discussion

RBM48 is a U12 splicing factor.

Genetic and molecular analysis of maize *rbm48* revealed extensive defects in U12-type intron splicing with approximately 60% of MIGs significantly mis-spliced. Defects in U2 splicing were observed in a fraction of introns, but these are predominantly of small magnitude with 85% of significantly affected U2-type introns splicing at ±10% of normal levels. Defects in MIG splicing have very similar patterns in both *rbm48* and *rgh3* with adjacent U2-type introns being mis-spliced in a subset of MIGs. Mis-splicing of neighboring U2-type introns in U12 splicing mutants has been observed in multiple species (Madan et al., 2015; Horiuchi et al., 2018). These data indicate that maize RBM48 is a U12 splicing factor. No RNA processing role has been assigned yet to RBM48, and the human ortholog was not identified in biochemical purification of U12

spliceosomes (Schneider et al., 2002; Will et al., 2004). However, RBM48 is conserved among organisms that have a U12 spliceosome and is coselected with ZRSR2/RGH3. Based on sequence and functional conservation, we predict that animal RBM48 proteins will also function in U12 splicing.

RBM48 interacts with U2AF, RGH3 and ARMC7.

The protein-protein interactions we found between RBM48 with U2AF, RGH3, and ARMC7 may help explain the impact on U2-type introns in a subset of U12 splicing mutants (Verma et al., 2017). A low frequency of U2 splicing defects have been observed for mutations in ZRSR2 homologs and RNPC3 (Madan et al., 2015; Gault et al., 2017; Horiuchi et al., 2018; Argente et al., 2014). RGH3/ZRSR2 and RNPC3 are required for U12-type intron recognition (Shen et al., 2010; Frilander and Steitz, 1999). Human ZRSR2 was known to interact with U2AF2 (Tronchere et al., 1997), but we found more extensive connection between U12 splicing factors and U2AF. Both U2AF and ZRSR2 are responsible for 3' splice site recognition of U2-type and U12-type introns, respectively (Shen et al., 2010). By contrast, U2 splicing defects have not been reported in human mutant alleles of U4atac, which are expected to recognize U12-type introns but not complete splicing reactions (Verma et al., 2017). The interaction data and the presence of a low frequency of U2-type splicing defects suggest that RBM48 and ARMC7 may also participate in U12-type intron recognition.

Our study also illustrates the importance of genetic analysis of RNA splicing factors to determine RNA processing functions. Recently, mutant models for several splicing factors including ZRSR2, ZRSR1, SMN, and FUS showed a predominant *in vivo*

function in minor intron splicing even though the proteins are thought to have broader roles based on *in vitro* splicing assays or protein-protein interactions (Madan et al., 2015; Gault et al., 2017; Horiuchi et al., 2018; Reber et al., 2016; Doktor et al., 2017; Jangi et al., 2017). RNA-seq analysis of mutant transcriptomes were able to resolve the predominant defects as impacting U12-type introns.

U12 splicing mutants in plants and animals have divergent developmental effects.

In plants, genetic analysis of U12 splicing factors has defined two phenotypic syndromes. Transgenic knockdown alleles for multiple U12 splicing factors in *Arabidopsis thaliana* develop serrated leaves, delay senescence, and dwarf inflorescence branches, which can be complemented by exogenous application of gibberellins (Kim et al., 2010; Jung and Kang, 2014; Xu et al., 2016). Strong alleles of U12 splicing factors in either Arabidopsis or maize disrupt seed development (Fig. 1) (Fouquet et al., 2011; Kim et al., 2010). In maize, aberrant U12 splicing disrupts both endosperm and embryo development. However, reduced U12 splicing efficiency does not cause cellular lethality, because *rgh3* and *rbm48* both show increased cell proliferation in endosperm tissue culture (Fig. 5) (Fouquet et al., 2011). These data argue that efficient U12 splicing is required for developmental functions rather than cell viability. Neither *rbm48* nor *rgh3* completely block U12 splicing, and it is likely that a complete block would cause cellular lethality.

Mutations disrupting U12 splicing in human cell lines reduce proliferation and are classified as core fitness genes (Hart et al., 2015; Blomen et al., 2015). Interestingly, ARMC7, an RBM48 interacting protein, was also found to be a core fitness gene in these

studies. It is somewhat surprising that homologous mutations in maize U12 splicing have contrasting cell proliferation defects in endosperm tissue culture (Fig. 5) (Gault et al., 2017; Fouquet et al., 2011). Gault et al. (2017) suggested that divergence in MIGs or in the location of U12-type introns within conserved MIGs are likely causes for this divergence in the cell arrest versus proliferation phenotypes in animals and plants.

Divergence of U12 splicing phenotypes likely result from fractionation of MIGs.

Although MIGs represent a small fraction of eukaryotic genes, plants and animals have MIGs in conserved cellular and biochemical pathways such as autophagy, secreted protein glycosylation, DNA damage repair, cell cycle, and histone methyltransferases (Gault et al., 2017). U12-type introns have longer and more conserved 5' splice site and branch point consensus sequences that are recognized through cooperative binding of the U11 and U12 snRNPs (Frilander and Steitz, 1999). Consequently, divergence in MIGs predominantly is due to U12-type intron loss or mutation to U2-type introns (Burge et al., 1998; Lin et al., 2010). Maintenance of MIGs in a conserved genetic architecture between plant and animal species suggests that U12 splicing efficiency could have similar effects on the cell.

A growing number of U12 splicing mutants in animals have been shown to cause defects in cell differentiation. A subset of human myelodysplasia patients have mutations in *ZRSR2* that result in aberrant U12 splicing and reduced terminal differentiation of myeloid blood cell types (Madan et al., 2015). Mutation of the mouse *ZRSR1* paralog of *ZRSR2* leads to blood and sperm differentiation defects (Horiuchi et al., 2018). Patients with Roifman syndrome have mutations in the human U4atac snRNA along with defects

in both myeloid and lymphoid blood cell differentiation (Heremans et al., 2018). These examples suggest a common requirement for U12 splicing to promote the differentiation of a subset of eukaryotic cell types. However, minor splicing mutations cause a range of human diseases and diverse phenotypes in animal models (Horiuchi et al., 2018; Verma et al., 2017; Pessa et al., 2010; Markmiller et al., 2014). The similarity of endosperm defects in *rbm48* and *rgh3* suggest a mutant ideotype for the developmental pathways sensitive to U12 splicing efficiency. Maize genetic screens for delayed endosperm cell differentiation and prolonged cell proliferation could discover both U12 splicing mutants and MIGs needed to promote cell differentiation.

Materials and Methods

Genetic stocks

The *rbm48-umu1* allele was isolated as a visual *rough endosperm* (*rgh*) mutant from the UniformMu transposon-tagging population (McCarty et al., 2005). The *rbm48-umu2* allele was identified from the UniformMu reverse genetics resource as *Mu* insertion mu1056623 in stock UFMu-07152 (McCarty et al., 2013; Settles et al., 2007). Both *rbm48-umu1* and *rbm48-umu2* were backcrossed five times into W22, B73, Mo17, and A636 for phenotypic analysis. An *rbm48-umu1* B73 x W22 F_2 mapping population was generated by self-pollinating the F_1 hybrid for the B73 introgression. The pBetl1:β-glucuronidase (GUS) and pVp1:GUS reporter transgenics have been described (Hueros et al., 1999; Cao et al., 2007). The *rgh3* mutant stock was maintained as a *Pr1 rgh3/pr1 Rgh3* stock in the W22 inbred background (Fouquet et al., 2011). All plant materials were

grown at the University of Florida Plant Science Research and Education Unit in Citra, FL or greenhouses located at the Horticultural Sciences Department in Gainesville, FL.

Molecular cloning of *rbm48*

The *rbm48-umu1* allele was mapped with bulked segregant analysis (BSA) using a pool of 93 homozygous *rgh* kernels from the B73 x W22 F_2 mapping population. DNA was extracted from the pooled kernels and genotyped with 144 single nucleotide polymorphism (SNP) markers using the Sequenom MassARRAY platform at the Iowa State University Genomic Technologies Facility as described (Liu et al., 2010). Relative enrichment for the W22 genotype was calculated from non-segregating pooled samples.

Mu flanking sequence tags (MuFSTs) were generated from a 12 x 12 grid of 144 UniformMu *rgh* mutants including *rbm48-umu1*. Transposon-flanking sequences were amplified from pooled DNA using MuTAIL-PCR (Settles et al., 2004). The PCR products were sequenced with a Roche Genome Sequencer FLX System using custom A primers to add pool-specific barcodes and sequence from the *Mu* terminal inverted repeat (TIR). Resulting sequences were filtered for intact TIR sequences and mapped to the B73_v2 genome with blastn. Identical insertion sites sequenced from a single row and a single column pool were assigned to individual *rgh* mutant isolates. The ten unique MuFSTs in *rbm48-umu1* were co-aligned to the BSA map position to identify *rbm48-umu1* as the likely causative mutation. Co-segregation of *rbm48-umu1* with the *rgh* kernel phenotype was tested by extracting DNA from segregating kernels and amplifying the wild-type or *Mu*-insertion alleles using gene specific primers and the TIR5 primer as described (Settles et al., 2007).

Kernel phenotypes

Segregating ears and mature kernels were imaged on a flatbed scanner. Both *rbm48* alleles were analyzed for kernel composition phenotypes by a single-kernel grain analyzer with a microbalance and near infrared spectrometer (Spielbauer et al., 2009; Gustin et al., 2013). At least 30 normal and 30 *rgh* kernels were sampled from three segregating ears of each allele. Four technical replicate kernel weight and NIR spectra were collected from each kernel. Average single-kernel composition predictions for each biological replicate were used as the kernel composition phenotypes shown in Fig. S2.

RT-PCR

Tissues were dissected, frozen in liquid nitrogen, and ground to a powder. RNA was extracted from 100 mg of tissue and mixed with 200 μL of RNA extraction buffer (50 mM Tris-HCl, pH 8, 150 mM LiCl, 5 mM EDTA, 1% SDS in DEPC treated water). The slurry was extracted using phenol: chloroform and Trizol (Invitrogen). RNA was precipitated from the aqueous fraction using isopropanol and washed with 70% ethanol. RNA pellets were re-suspended in nuclease free water (Sigma) treated with Purelink^{Tm} DNase (Invitrogen). RNA was further purified using an RNeasy MinElute Cleanup Kit (Qiagen). M-MLV reverse transcriptase (Promega) used 1 μg total RNA as template to synthesize cDNA.

Semi-quantitative RT-PCR was carried out by modifying the methods from Bai and DeMason (2006). The PCR conditions were 3 min at 94°C, followed by 24-28 cycles of 30 sec at 94°C, 30 sec at the appropriate annealing temperature for the primers and

1 min at 72°C, followed by a final extension of 5 min at 72°C. Quantitative RT-PCR used a StepOnePlus real-time PCR machine (Applied Biosystems) with 1X SYBR® Green PCR Master Mix (Applied Biosystems) as described (Fouquet et al., 2011; Bai et al., 2016). The normalized expression level of each gene represents the average of three replicates of three biologically distinct kernel pools relative to *actin1* using the comparative cycle threshold ($\Delta\Delta C_t$) method (Livak and Schmittgen, 2001). Primer sequences are Table S6.

Phylogenetic analysis

Coselection was detected by parsing the species from the specific protein lists for the RRM_RBM48, RRM_U2AFBPL, and RRM_U2AF35 domains in the NCBI Conserved Domains Database (Marchler-Bauer et al., 2017) A representative set of RBM48 proteins were aligned with MUSCLE (version 3.8.31) using default parameters (Edgar, 2004). The multiple sequence alignment of the RBM48_RRM was input to RAxML (version 8.2.3) (Stamatakis, 2014) to construct a maximum likelihood phylogeny with the LG4X amino acid replacement matrix model and 1,000 bootstrap replicates (Le et al., 2012). Eight tree searches were completed to further test for tree stability. The phylogeny was rooted by the oomycetes RBM48 protein and displayed with Dendroscope (version 3.2.10) (Huson and Scornavacca, 2012).

RNA-Seq analysis

Endosperm tissue from 16-18 DAP normal and *rbm48* kernels in the W22 genetic background was dissected and frozen in liquid nitrogen. Total mRNA was extracted from

four biological replicates of paired *rbm48* mutant and normal sibling pools. Non-strand specific TruSeq^{TM} (Illumina) cDNA libraries were prepared from 1 µg total RNA input with a 200 bp median insert length. All libraries were quantified using a Qubit, pooled, and 100 bp paired-end reads were sequenced on two lanes of the HiSeq 2000 platform.

Raw RNA-seq data were screened to remove adapter sequences using Cutadapt v1.1 (Martin, 2011) with the following parameters: error-rate=0.1, times=1, overlap=5, and minimum-length=0. Adapter trimmed sequences were quality trimmed with Trimmomatic v0.22(Bolger et al., 2014) using parameters (HEADCROP:0, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, and MINLEN:40) to truncate reads for base quality <15 within 4 base windows and kept only reads ≥40 bases after trimming.

Reads were uniquely aligned the B73 RefGen_v2 maize genome assembly with GSNAP (Version 2013-07-20) using the following parameters: --orientation=FR --batch=5 --suboptimal-levels=0 --novelsplicing=1 --localsplicedist=8000 --local-splice-penalty=0 --distant-splice-penalty=4 --quality-protocol=sanger --npaths=1 --quiet-if-excessive --max-mismatches=0.02 --nofails --format=sam --sam-multiple-primaries --pairmax-rna=8000 --pairexpect=200 --pairdev=150 --nthreads=4. The --use-splicing argument was implemented to guide GSNAP by the ZmB73_5b Filtered Gene Set annotations but was permitted to discover novel splicing events. The argument --use-snps was called with an IIT file containing W22/B73 (Gault et al., 2017) as used to allow GSNAP to perform SNP tolerant alignments ensuring proper mapping of the w22 RNA-Seq reads to the B73 reference genome.

Read counts/gene were determined with the HTSeq-Count utility in the HTSeq package (Anders et al., 2015) (Ver 0.8.0). Non-redundant introns and genomic

coordinates were identified from the ZmB73_5b annotation as described, except that U12-type introns were filtered to a non-redundant set of 372 introns (Gault et al., 2017) Percent spliced out (PSO) metrics for each intron were calculated from exon-exon junction spanning reads and intron reads for each unique 5' and 3' splice site (Gault et al., 2017; Katz et al., 2010). Fisher's exact test was calculated for each intron from the summed exon-exon junction reads and intron reads across the four normal and four mutant libraries for each *rbm48* allele. Test statistics were correct to a false discovery rate ≤0.05 using the Benjamini-Hochberg method.

Differentially expressed transcripts were detected with the DESeq2 (Love et al., 2014) Bioconductor package. Transcripts were considered expressed if TPM was ≥1 in at least one genotype. Differential expression criteria were an adjusted p-value ≤0.05 and ≥2- fold change. GO term enrichment analysis used agriGO v2.0 with default parameters and the expressed transcripts as a customized reference (Du et al., 2010).

Histology

Developing kernels from ears segregating for *rbm48* were harvested and fixed overnight at 4°C in FAA (3.7% formaldehyde, 5% glacial acetic acid, and 50% ethanol). Kernels were embedded in JB-4 plastic embedding media (Electron Microscopy Sciences) and sectioned at 4 μm thickness as described (Bai et al., 2016). Sections were stained in Schiff's reagent, counter-stained with 1% aniline blue-black in 7% acetic acid, washed, dried, and mounted. Imaging was completed with a Zeiss Axiophot light microscope and an Amscope digital camera.

For β-glucuronidase (GUS) reporter expression analysis, male plants bearing either pBetl1:GUS or pVp1:GUS were crossed by *rbm48*/+ heterozygous plants and resulting hybrids were self-pollinated (Hueros et al., 1999; Cao et al., 2007). Developing kernels were sampled from segregating ears, sectioned in the sagittal plane, and stained for GUS activity as previously described (Costa et al., 2003; Gutierrez-Marcos et al., 2006; Bai and Demason, 2008). At least 20 kernels for 3-4 biological replicates were observed, and 6-9 kernels were imaged to ensure that GUS reporter images were representative of the staining patterns. Images were captured with a Wild Heerbrugg dissecting microscope and an Amscope digital camera.

Endosperm tissue culture

Endosperm callus cultures were initiated as described (Fouquet et al., 2011; Shannon, 1994). Briefly, ears segregating for *rbm48* from the A636 introgression were harvested from 6-18 DAP. Ears were surface sterilized in 70% ethanol and 20% bleach for 10 minutes and washed with sterile water. Individual endosperm tissues were dissected and placed on Murashige and Skoog media (pH 5.7) and supplemented with 3% sucrose, 4 ppm thiamine, 0.2% asparagine, and 2% phytagel. Growth of 400 individual endosperm tissues per genotype and harvest date was assayed after incubating in the dark for 35 d at 29°C. Cultures were imaged with a Nikon digital camera.

Subcellular localization

RGH3, RGH3ΔUHM, U2AF1, and U2AF2 fusion protein constructs have been described previously (Gault et al., 2017; Fouquet et al., 2011). RBM48 fusion proteins were constructed with a similar strategy. Briefly, the full-length RBM48 ORF was

amplified from B73 seedling cDNA. Primer pairs to enable Gateway cloning of N- or C-terminal fusions were used to amplify full-length (RBM48_FL), N-terminal deletion of the RRM domain (RBM48ΔRRM), or C-terminal deletion (RBM48ΔRS), see Table S6 for primer sequences. PCR products were cloned into pDONR221 and recombined into Gateway destination vectors containing GFP, RFP, or split EYFP following the manufacturer's protocol (Invitrogen) (Karimi et al., 2002; Karimi et al., 2007). For BiFC constructs, recombined expression cassettes, pSAT5-DEST or pSAT4-DEST, were digested with I-CeuI or I-SceI and ligated into pPZP-RCS2-bar binary vector. Binary vectors were transformed in *Agrobacterium tumefaciens* strains ABi and GV3101 (Wise et al., 2006).

N. benthamiana plants were grown for 5-8 weeks in a growth chamber 16/8 h and 26/22°C day/night for transient expression assays. *Agrobacterium* was infiltrated into stomata of the abaxial surface of leaves using a needleless syringe. Leaves were imaged 24-48 h after infiltration. For co-expression analysis, *Agrobacterium* strains with individual constructs were mixed in a 1:1 ratio prior to infiltration. Leaf sections were stained with DAPI prior to visualization of fusion protein expression (Kapila et al., 1997). Representative images of subcellular localization were obtained using a Carl Zeiss Axio Imager .Z2 confocal microscope (Cevik and Kazan, 2013).

Protein pull-down assays

Coding sequences for maize U2AF1 (GRMZM2G177229) and U2AF2 (GRMZM2G093256) were amplified with primers containing Nco-I and Not-I restriction enzyme sites, cloned into a pCR4-TOPO vector (Invitrogen), then subcloned into pET-

28b with standard restriction-ligation protocols. *E. coli* codon-optimized coding sequences of maize RBM48, RGH3, and ARMC7 were synthesized and cloned into pET-28b using NcoI and XhoI restriction sites (Genscript, Inc.). *Caldicellulosiruptor bescii* Cbes_2310 (WP_015908637.1), a predicted periplasmic sugar binding protein was cloned into pET-28b using NheI and XhoI restriction sites (Dam et al., 2011). A stop codon was included in the inserted *Cbes_2310* sequence to prevent translation of the C-terminal His-tag in the pET-28b vector. RBM48 and RGH3 were also cloned in pET-42a using NcoI and XhoI restriction sites to produce N-terminal tandem GST-His-tag fusion proteins (GenScript, Inc.). A stop codon was included in the inserted *RBM48* and *RGH3* sequences to prevent translation of the C-terminal His-tag in the pET-42a vector.

Recombinant expression conditions for all tagged versions of RGH3, U2AF1, U2AF2, and ARMC7 was initiated with a 1:50 dilution of overnight cultures in fresh LB media supplemented with 50 µg/mL kanamycin. Cultures were grown at 37°C with shaking at 225 rpm until OD_{600} = 0.45. Expression was induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cultures were induced for 3 h at room temperature with shaking at 225 rpm. Bacterial cells were harvested by centrifugation at 2,500 x g for 15 min at 4°C. U2AF1, U2AF2, and ARMC7 cell pellets were stored at -80°C. RGH3 induction was improved by storing cell pellets overnight at 4°C with supernatant media. The supernatant was removed the following morning and the pellet stored at -80°C.

RBM48 recombinant proteins were expressed with a 1:20 dilution of an overnight grown culture into MagicMedia™ *E. coli* Expression Medium (Invitrogen) supplemented with 50 µg/mL kanamycin. Cells were grown at 37°C with shaking at 225 rpm until

OD_{600} = 0.6. The culture was then transferred to 18°C and grown for an additional 36 h at 225 rpm to allow for auto-induction of recombinant protein expression. After incubation at 18°C, the culture was centrifuged at 2,500 x g for 15 min at 4°C, and the pellet was stored at -80°C.

Cell pellets were resuspended in 2 mL of lysis buffer (50 mM Tris, 150 mM NaCl, 1X Halt® Protease Inhibitor Cocktail (ThermoScientific), 1 mg/mL lysozyme, 10 µg/mL DNase I, 1%Triton X-100) and mixed with a rotisserie tube-rotator at room temperature for 10 min. The lysate was sonicated at 20 kHz for 30 s on/off five times, and subsequently centrifuged at 10,000 x g for 10 min at 4°C.

GST protein pull-down assays were based on Pierce™ Glutathione Magnetic Agarose Beads instructions (ThermoScientific). Briefly, 400 µL bacterial lysate containing induced fusion protein was mixed with 100 µL equilibration buffer and incubated with 25 µL settled bead volume in a 1.5 mL microcentrifuge tube for 1 h at 4°C on a rotisserie tube-rotator. Beads were collected in a magnetic stand, the supernatant was discarded, and beads were washed twice with equilibration buffer. In the third wash, half of the slurry volume was transferred to a new microcentrifuge tube, and the GST fusion protein was eluted with 125 µL elution buffer to assay input protein for the pull-down.

The remaining protein-bound beads were separated on a magnetic stand, incubated with a mix of 200 µL of induced bacterial lysate containing His-tagged protein mixed with 50 µL of equilibration buffer. The lysate was incubated with the bound GST fusion protein for 1 h at 4°C on a rotisserie tube-rotator. The beads were collected and

supernatant discarded. Beads were washed three times using 250 µL of equilibration buffer. Bound proteins were eluted with 125 µL elution buffer.

GST-His-RBM48, GST-His-RGH3, RGH3-His, U2AF1-His, U2AF2-His, ARMC7-His, and His-Cbes_2310 were detected by western blot using a monoclonal His-tag antibody (Cell Signaling, cat# 2365S). Protein fragments of RGH3-His were excised from the gels, subjected to trypsin digestion, and sequenced according to the procedure described (Shevchenko et al., 1996) at the Biomolecular and Proteomics Mass Spectrometry Facility at the University of California, San Diego.

Data availability

All sequence data are available at the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA). Maize transposon flanking sequence tags for identification of *rbm48-umu1* are in accession: SRX006230. Endosperm RNA-seq data are in accession:_____. Genetic stocks for *rbm48* mutants and clones described are available upon request.

Acknowledgements

We thank John Baier, Jennifer Moses, J. Paige Gronevelt, Elizabeth Jankulovski, and Laurel Levine for technical assistance. This work was supported by National Science Foundation grants (MCB-1412218 to SL, WBB, AMS; IOS-1547787 to WBB; IOS-1623478 to AMS), the University of Florida HHMI Science for Life undergraduate research program, and the Vasil-Monsanto Endowment.

Author Contributions

A.M.S., S.L., W.B.B., F.B., and J.C. wrote the manuscript. A.M.S., F.B., G.S., and C.W.T. completed plant genetics and mutant mapping; F.B. characterized maize kernel, seedling, and cellular phenotypes; A.M.S., S.L., and G.F. completed protein sequence evolutionary analysis; W.B.B., R.D., F.B., A.M.S. designed and completed RNA-seq experiments; F.B., J.C., D.S., J.M., and A.E.S. completed RT-PCR experiments; A.M.S. and S.L. designed and supervised protein-protein interaction experiments; D.S. and F.M. completed co-localization and BiFC experiments; J.C., C.J.B., and F.M. developed recombinant expression constructs and methodology; J.C. completed protein pull-down experiments. The manuscript has been seen and approved by all authors.

References

Alioto, T.S. (2006). U12DB: a database of orthologous U12-type spliceosomal introns. Nucleic acids research *35*, D110-D115.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Argente, J., Flores, R., Gutierrez-Arumi, A., Verma, B., Martos-Moreno, G.A., Cusco, I., Oghabian, A., Chowen, J.A., Frilander, M.J., and Perez-Jurado, L.A. (2014). Defective minor spliceosome mRNA processing results in isolated familial growth hormone deficiency. EMBO Mol Med *6*, 299-306.

Bai, F., Daliberti, M., Bagadion, A., Xu, M., Li, Y., Baier, J., Tseung, C.W., Evans, M.M., and Settles, A.M. (2016). Parent-of-Origin-Effect rough endosperm Mutants in Maize. Genetics *204*, 221-231.

Bai, F., and DeMason, D.A. (2006). Hormone interactions and regulation of Unifoliata, PsPK2, PsPIN1 and LE gene expression in pea (Pisum sativum) shoot tips. Plant Cell Physiol *47*, 935-948.

Bai, F., and Demason, D.A. (2008). Hormone interactions and regulation of PsPK2::GUS compared with DR5::GUS and PID::GUS in Arabidopsis thaliana. Am J Bot *95*, 133-145.

Blomen, V.A., Majek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., *et al.* (2015). Gene essentiality and synthetic lethality in haploid human cells. Science *350*, 1092-1096.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Brown, J.W.S. (1998). Splice site selection in plant pre-mRNA splicing. Annu Rev Plant Physiol Plant Mol Biol *49*, 77-95.

Burge, C.B., Padgett, R.A., and Sharp, P.A. (1998). Evolutionary fates and origins of U12-type introns. Molecular cell *2*, 773-785.

Cao, X., Costa, L.M., Biderre-Petit, C., Kbhaya, B., Dey, N., Perez, P., McCarty, D.R., Gutierrez-Marcos, J.F., and Becraft, P.W. (2007). Abscisic acid and stress signals induce Viviparous1 expression in seed and vegetative tissues of maize. Plant physiology *143*, 720-731.

Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A.R., and Caceres, J.F. (2002). Nuclear export and retention signals in the RS domain of SR proteins. Mol Cell Biol *22*, 6871-6882.

Cevik, V., and Kazan, K. (2013). Agroinfiltration of Nicotiana benthamiana leaves for co-localization of regulatory proteins involved in jasmonate signaling. Methods Mol Biol *1011*, 199-208.

Citovsky, V., Lee, L.Y., Vyas, S., Glick, E., Chen, M.H., Vainstein, A., Gafni, Y., Gelvin, S.B., and Tzfira, T. (2006). Subcellular localization of interacting proteins by bimolecular fluorescence complementation in planta. Journal of molecular biology *362*, 1120-1131.

Costa, L.M., Gutierrez-Marcos, J.F., Brutnell, T.P., Greenland, A.J., and Dickinson, H.G. (2003). The globby1-1 (glo1-1) mutation disrupts nuclear and cell division in the developing maize seed causing alterations in endosperm cell fate and tissue differentiation. Development (Cambridge, England) *130*, 5009-5017.

Dam, P., Kataeva, I., Yang, S.J., Zhou, F., Yin, Y., Chou, W., Poole, F.L., 2nd, Westpheling, J., Hettich, R., Giannone, R., *et al.* (2011). Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium Caldicellulosiruptor bescii DSM 6725. Nucleic acids research *39*, 3240-3254.

Dávila, L.M., Rosenblad, M., and Samuelsson, T. (2008). Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. Nucleic acids research *36*, 3001.

Doktor, T.K., Hua, Y., Andersen, H.S., Broner, S., Liu, Y.H., Wieckowska, A., Dembic, M., Bruun, G.H., Krainer, A.R., and Andresen, B.S. (2017). RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns. Nucleic acids research *45*, 395-416.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. Nucleic acids research *38*, W64-70.

Edery, P., Marcaillou, C., Sahbatou, M., Labalme, A., Chastang, J., Touraine, R., Tubacher, E., Senni, F., Bober, M.B., Nampoothiri, S., *et al.* (2011). Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. Science *332*, 240-243.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Fouquet, R., Martin, F., Fajardo, D.S., Gault, C.M., Gomez, E., Tseung, C.W., Policht, T., Hueros, G., and Settles, A.M. (2011). Maize rough endosperm3 encodes an RNA splicing factor required for endosperm cell differentiation and has a nonautonomous effect on embryo development. Plant Cell *23*, 4280-4297.

Frilander, M.J., and Steitz, J.A. (1999). Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. Genes Dev *13*, 851-863.

Gault, C.M., Martin, F., Mei, W., Bai, F., Black, J.B., Barbazuk, W.B., and Settles, A.M. (2017). Aberrant splicing in maize rough endosperm3 reveals a conserved role for U12 splicing in eukaryotic multicellular development. Proceedings of the National Academy of Sciences of the United States of America *114*, E2195-E2204.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. Rna *6*, 1197-1211.

Gustin, J.L., Jackson, S., Williams, C., Patel, A., Armstrong, P., Peter, G.F., and Settles, A.M. (2013). Analysis of maize (Zea mays) kernel density and volume using microcomputed tomography and single-kernel near-infrared spectroscopy. J Agric Food Chem *61*, 10872-10880.

Guth, S., Tange, T.O., Kellenberger, E., and Valcarcel, J. (2001). Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. Mol Cell Biol *21*, 7673-7681.

Gutierrez-Marcos, J.F., Costa, L.M., and Evans, M.M. (2006). Maternal gametophytic baseless1 is required for development of the central cell and early endosperm patterning in maize (Zea mays). Genetics *174*, 317-329.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., *et al.* (2015). High-Resolution

CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515-1526.

He, H., Liyanarachchi, S., Akagi, K., Nagy, R., Li, J., Dietrich, R.C., Li, W., Sebastian, N., Wen, B., Xin, B., *et al.* (2011). Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. Science *332*, 238-240.

Heremans, J., Garcia-Perez, J.E., Turro, E., Schlenner, S.M., Casteels, I., Collin, R., de Zegher, F., Greene, D., Humblet-Baron, S., Lesage, S., *et al.* (2018). Abnormal differentiation of B cells and megakaryocytes in patients with Roifman syndrome. J Allergy Clin Immunol.

Horiuchi, K., Perez-Cerezales, S., Papasaikas, P., Ramos-Ibeas, P., Lopez-Cardona, A.P., Laguna-Barraza, R., Fonseca Balvis, N., Pericuesta, E., Fernandez-Gonzalez, R., Planells, B., *et al.* (2018). Impaired Spermatogenesis, Muscle, and Erythrocyte Function in U12 Intron Splicing-Defective Zrsr1 Mutant Mice. Cell Rep *23*, 143-155.

Hueros, G., Gomez, E., Cheikh, N., Edwards, J., Weldon, M., Salamini, F., and Thompson, R.D. (1999). Identification of a promoter sequence from the BETL1 gene cluster able to confer transfer-cell-specific expression in transgenic maize. Plant physiology *121*, 1143-1152.

Huson, D.H., and Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol *61*, 1061-1067.

Irimia, M., and Roy, S.W. (2014). Origin of spliceosomal introns and alternative splicing. Cold Spring Harb Perspect Biol *6*.

Jangi, M., Fleet, C., Cullen, P., Gupta, S.V., Mekhoubad, S., Chiao, E., Allaire, N., Bennett, C.F., Rigo, F., Krainer, A.R., *et al.* (2017). SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. Proceedings of the National Academy of Sciences of the United States of America *114*, E2347-E2356.

Jung, H.J., and Kang, H. (2014). The Arabidopsis U11/U12-65K is an indispensible component of minor spliceosome and plays a crucial role in U12 intron splicing and plant development. The Plant journal : for cell and molecular biology *78*, 799-810.

Kapila, J., Rycke, R.D., Van Montagu, M., and Angenon, G. (1997). An Agrobacterium-mediated transient gene expression system for intact leaves. Plant Sci *122*, 101-108.

Karimi, M., Depicker, A., and Hilson, P. (2007). Recombinational cloning with plant gateway vectors. Plant physiology *145*, 1144-1154.

Karimi, M., Inze, D., and Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. Trends in plant science *7*, 193-195.

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods *7*, 1009-1015.

Kerppola, T.K. (2008). Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. Annual review of biophysics *37*, 465-487.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell *106*, 595-605.

Kim, W.Y., Jung, H.J., Kwak, K.J., Kim, M.K., Oh, S.H., Han, Y.S., and Kang, H. (2010). The Arabidopsis U12-type spliceosomal protein U11/U12-31K is involved in U12 intron splicing via RNA chaperone activity and affects plant development. Plant Cell *22*, 3951-3962.

Konig, H., Matter, N., Bader, R., Thiele, W., and Muller, F. (2007). Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. Cell *131*, 718-729.

Le, S.Q., Dang, C.C., and Gascuel, O. (2012). Modeling protein evolution with several amino acid replacement matrices depending on site rates. Mol Biol Evol *29*, 2921-2936.

Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem *84*, 291-323.

Lin, C.F., Mount, S.M., Jarmolowski, A., and Makalowski, W. (2010). Evolutionary dynamics of U12-type spliceosomal introns. BMC evolutionary biology *10*, 47.

Liu, S., Chen, H.D., Makarevitch, I., Shirmer, R., Emrich, S.J., Dietrich, C.R., Barbazuk, W.B., Springer, N.M., and Schnable, P.S. (2010). High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. Genetics *184*, 19-26.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif) *25*, 402-408.

Lorkovic, Z.J., Wieczorek Kirk, D.A., Lambermon, M.H., and Filipowicz, W. (2000). Pre-mRNA splicing in higher plants. Trends in plant science *5*, 160-167.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Madan, V., Kanojia, D., Li, J., Okamoto, R., Sato-Otsubo, A., Kohlmann, A., Sanada, M., Grossmann, V., Sundaresan, J., Shiraishi, Y., *et al.* (2015). Aberrant splicing of U12-type

introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. Nat Commun *6*, 6042.

Madanieh, R., Mathew, S., Shah, P., Vatti, S.K., Madanieh, A., Kosmas, C.E., and Vittorio, T.J. (2015). Cardiac Magnetic Resonance Imaging Might Complement Two-Dimensional Echocardiography in the Detection of a Reversible Nonischemic Cardiomyopathy. Clinical medicine insights Case reports *8*, 109-114.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., *et al.* (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic acids research *45*, D200-D203.

Markmiller, S., Cloonan, N., Lardelli, R.M., Doggett, K., Keightley, M.C., Boglev, Y., Trotter, A.J., Ng, A.Y., Wilkins, S.J., Verkade, H., *et al.* (2014). Minor class splicing shapes the zebrafish transcriptome during development. Proceedings of the National Academy of Sciences of the United States of America *111*, 3062-3067.

Martin, M. (2011). Cutadapt removes adapter sequencing from high-throughput sequencing reads. EMBnet J *17*, 10-12.

McCarty, D.R., Settles, A.M., Suzuki, M., Tan, B.C., Latshaw, S., Porch, T., Robin, K., Baier, J., Avigne, W., Lai, J., *et al.* (2005). Steady-state transposon mutagenesis in inbred maize. Plant J *44*, 52-61.

McCarty, D.R., Suzuki, M., Hunter, C., Collins, J., Avigne, W.T., and Koch, K.E. (2013). Genetic and molecular analyses of UniformMu transposon insertion lines. Methods Mol Biol *1057*, 157-166.

Montzka, K.A., and Steitz, J.A. (1988). Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. Proceedings of the National Academy of Sciences *85*, 8885-8889.

Mori, T., Yoshimura, K., Nosaka, R., Sakuyama, H., Koike, Y., Tanabe, N., Maruta, T., Tamoi, M., and Shigeoka, S. (2012). Subcellular and subnuclear distribution of high-light responsive serine/arginine-rich proteins, atSR45a and atSR30, in Arabidopsis thaliana. Bioscience, biotechnology, and biochemistry *76*, 2075-2081.

Niemela, E.H., and Frilander, M.J. (2014). Regulation of gene expression through inefficient splicing of U12-type introns. Rna Biol *11*, 1325-1329.

Niemela, E.H., Oghabian, A., Staals, R.H., Greco, D., Pruijn, G.J., and Frilander, M.J. (2014). Global analysis of the nuclear processing of transcripts with unspliced U12-type introns by the exosome. Nucleic acids research *42*, 7358-7369.

Patel, A.A., McCarthy, M., and Steitz, J.A. (2002). The splicing of U12-type introns can be a rate-limiting step in gene expression. The EMBO journal *21*, 3804-3815.

Pessa, H.K., Greco, D., Kvist, J., Wahlstrom, G., Heino, T.I., Auvinen, P., and Frilander, M.J. (2010). Gene expression profiling of U12-type spliceosome mutant Drosophila reveals widespread changes in metabolic pathways. PLoS One *5*, e13215.

Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome research *12*, 1231-1245.

Reber, S., Stettler, J., Filosa, G., Colombo, M., Jutzi, D., Lenzken, S.C., Schweingruber, C., Bruggmann, R., Bachi, A., Barabino, S.M., *et al.* (2016). Minor intron splicing is regulated by FUS and affected by ALS-associated FUS mutants. The EMBO journal *35*, 1504-1521.

Roy, S.W., and Irimia, M. (2009). Splicing in the eukaryotic ancestor: form, function and dysfunction. Trends in ecology & evolution *24*, 447-455.

Ru, Y., Wang, B.B., and Brendel, V. (2008). Spliceosomal proteins in plants. Current topics in microbiology and immunology *326*, 1-15.

Saitoh, N., Sakamoto, C., Hagiwara, M., Agredano-Moreno, L.T., Jimenez-Garcia, L.F., and Nakao, M. (2012). The distribution of phosphorylated SR proteins and alternative splicing are regulated by RANBP2. Molecular biology of the cell *23*, 1115-1128.

Schneider, C., Will, C.L., Makarova, O.V., Makarov, E.M., and Luhrmann, R. (2002). Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. Mol Cell Biol *22*, 3219-3229.

Settles, A.M., Holding, D.R., Tan, B.C., Latshaw, S.P., Liu, J., Suzuki, M., Li, L., O'Brien, B.A., Fajardo, D.S., Wroclawska, E., *et al.* (2007). Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. BMC Genomics *8*, 116.
Settles, A.M., Latshaw, S., and McCarty, D.R. (2004). Molecular analysis of high-copy insertion sites in maize. Nucleic acids research *32*, e54.

Shannon, J.C. (1994). Establishment and culture of maize endosperm. In The Maize Handbook, M Freeling and V Walbot, eds (New York: Springer-Verlag), 719-722.
Shaul, O. (2015). Unique Aspects of Plant Nonsense-Mediated mRNA Decay. Trends in plant science 20, 767-779.

Shen, H., Zheng, X., Luecke, S., and Green, M.R. (2010). The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. Genes Dev 24, 2389-2394.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 68, 850-858.

Simpson, G.G., and Filipowicz, W. (1996). Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery. Plant molecular biology 32, 1-41.

Spielbauer, G., Armstrong, P., Baier, J.W., Allen, W.B., Richardson, K., Shen, B., and Settles, A.M. (2009). High-throughput near infrared reflectance spectroscopy measures quantitative and qualitative composition phenotypes of individual maize kernels. Cereal Chem 86, 556-564.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Tarn, W.Y., and Steitz, J.A. (1995). Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. Proceedings of the National Academy of Sciences of the United States of America 92, 2504-2508.

Tillemans, V., Dispa, L., Remacle, C., Collinge, M., and Motte, P. (2005). Functional distribution and dynamics of Arabidopsis SR splicing factors in living plant cells. Plant J 41, 567-582.

Tillemans, V., Leponce, I., Rausin, G., Dispa, L., and Motte, P. (2006). Insights into nuclear organization in plants as revealed by the dynamic distribution of Arabidopsis SR splicing factors. Plant Cell 18, 3218-3234.

Tronchere, H., Wang, J., and Fu, X.D. (1997). A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. Nature 388, 397-400.

Turunen, J.J., Niemela, E.H., Verma, B., and Frilander, M.J. (2013). The significant other: splicing by the minor spliceosome. Wiley interdisciplinary reviews RNA 4, 61-76.

Verma, B., Akinyi, M.V., Norppa, A.J., and Frilander, M.J. (2017). Minor spliceosome and disease. Semin Cell Dev Biol.

Will, C.L., and Luhrmann, R. (2005). Splicing of a rare class of introns by the U12-dependent spliceosome. Biol Chem 386, 713-724.

Will, C.L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T., and Luhrmann, R. (2004). The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. RNA 10, 929-941.

Wise, A.A., Liu, Z., and Binns, A.N. (2006). Three methods for the introduction of foreign DNA into Agrobacterium. Methods Mol Biol 343, 43-53.

Xu, T., Kim, B.M., Kwak, K.J., Jung, H.J., and Kang, H. (2016). The Arabidopsis homolog of human minor spliceosomal protein U11-48K plays a crucial role in U12 intron splicing and plant development. Journal of experimental botany.

Younis, I., Dittmar, K., Wang, W., Foley, S.W., Berg, M.G., Hu, K.Y., Wei, Z., Wan, L., and Dreyfuss, G. (2013). Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. Elife 2, e00780.

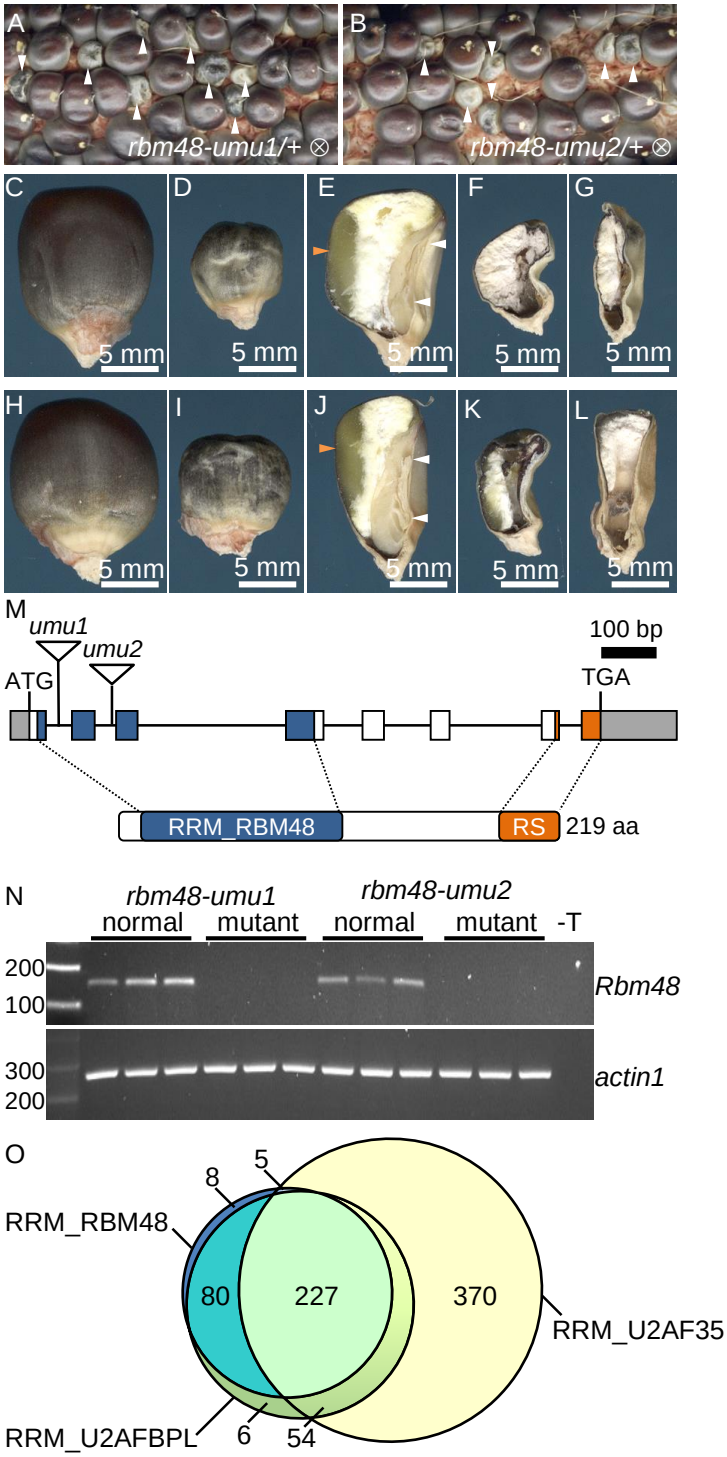

Figure 1. *Rbm48* is required for kernel development and is coselected with RGH3/ZRSR2. (A-B) Segregating self-pollinated ears for *rbm48-umu1* and *rbm48-umu2* alleles. Arrowheads indicate *rbm48* mutant kernels. (C-L) Mature kernel phenotypes for *rbm48-umu1* (C-G) and *rbm48-umu2* (H-L). Abgerminal view of normal siblings (C, H), *rbm48-umu1* (D), and *rbm48-umu2* (I). Sagittal sections of normal siblings (E, J), *rbm48-umu1* (D), and *rbm48-umu2* (I). Scale bars are 5 mm in all panels. White arrowheads indicate shoot and root of the embryo. Red arrowhead indicate vitreous endosperm. (M) Schematic of the *Rbm48* locus and protein domain structure. Triangles indicate transposon insertions causing *rbm48-umu1* and *rbm48-umu2*. (N) RT-PCR of *Rbm48* and *actin1* control in *rbm48-umu1* and *rbm48-umu2*, and their normal siblings. T is a no template DNA negative control. (O) Proportional Venn diagram showing the number of species in the NCBI Conserved Domain Database with RRM domains from RGH3 (RRM_U2AFBPL), RBM48 (RRM_RBM48), and U2AF2 (RRM_U2AF35).

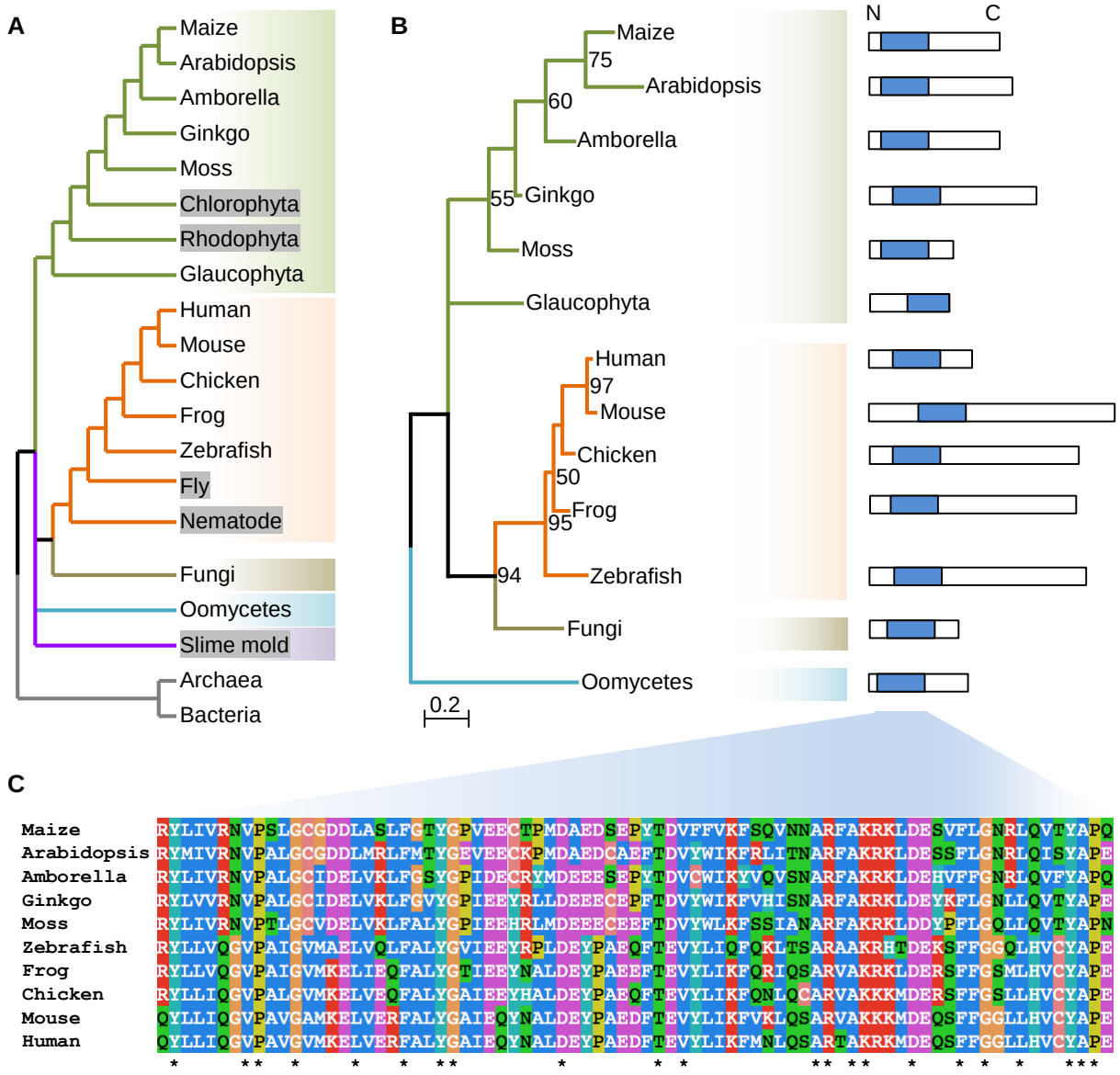

Figure 2. RBM48 is missing from eukaryotic clades that have lost U12 splicing. (A) Species tree including significant eukaryotic model organisms. Grey boxes indicate lineages that have lost RBM48. *Drosophila melanogaster* (fly) encodes a hypothetical gene with an atypical RBM48 RRM-like domain. (B) Maximum likelihood tree of the RBM48_RRM domain. Bootstrap values ≥50 are reported in the corresponding nodes. Protein schematics show the RBM48 domain in blue. The glaucophyta hypothetical RBM48 protein is from *Cyanophora paradoxa*; the fungi hypothetical RBM48 protein is from *Rhizophagus irregularis* DAOM 181602; the oomycetes hypothetical RBM48 protein is from *Phytophthora sojae*. (C) Multiple sequence alignment of representative RBM48 domains. Asterisks indicate completely conserved sites.

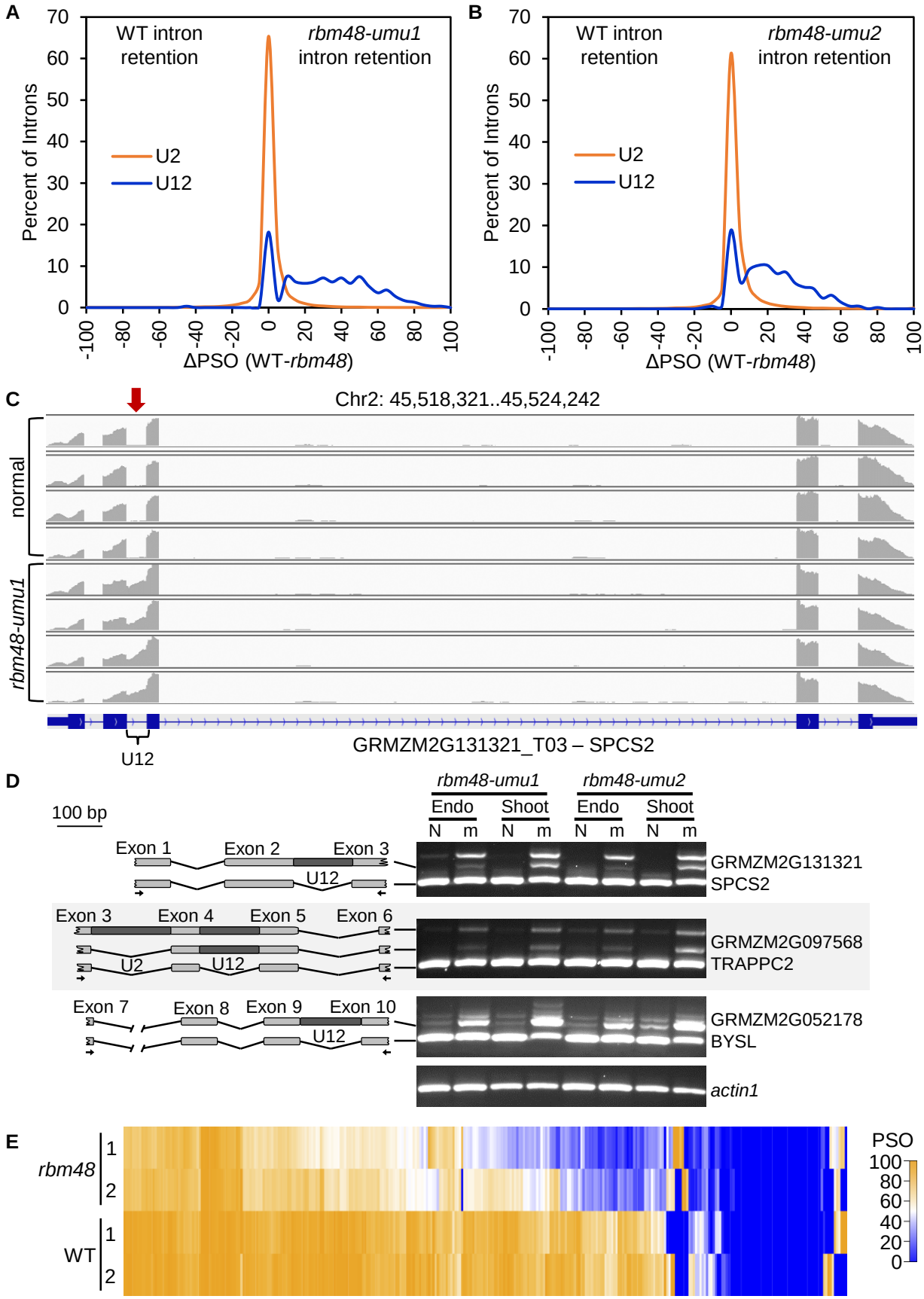

Figure 3. Legend on next page.

Figure 3. U12 splicing is defective in *rbm48* mutants. (A-B) Distribution of RNA splicing differences for U2-type and U12-type introns in *rbm48-umu1* (A) and *rbm48-umu2* (B) 16-18 DAP endosperm. Plots show the distribution of ΔPSO values for all introns with sufficient reads to calculate ΔPSO. Positive values have intron retention in mutant endosperm. (C) RNA-seq read depth for GRMZM2G131321, a homolog of human SPCS2. Red arrow and brace symbol indicate the U12-type intron with increased read depth in *rbm48-umu1*. (D) RT-PCR of normal (N) and *rbm48* mutant (m) RNA from endosperm and seedling shoot tissues. Schematics show amplified products with PCR primers indicated by arrows. (E) Heat map of PSO values for all U12-type introns measured by RNA-seq.

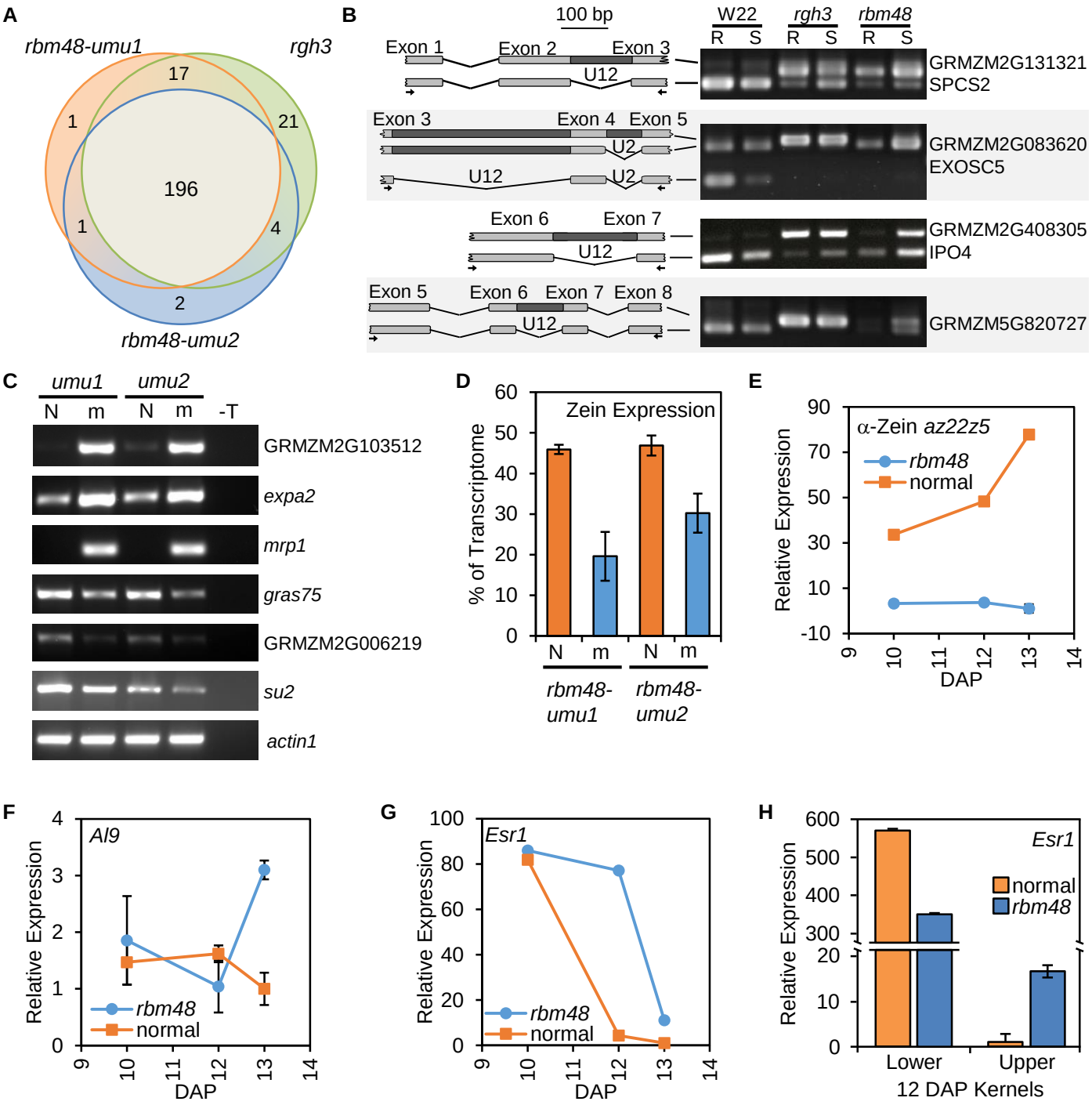

Figure 4. Expression defects in *rbm48*. (A) Venn diagram of U12-type introns that are significantly retained in *rbm48-umu1*, *rbm48-umu2*, and *rgh3* (Gault et al., 2017) RNA-seq experiments. The diagram shows introns that were tested in all three experiments. (B) RT-PCR of root (R) and shoot (S) tissues from W22 inbred, *rgh3*, and *rbm48-umu1* seedlings. Schematics show amplified products with PCR primers indicated by arrows. Gene names are from maize B73_v2 and human homologs. (C) RT-PCR of normal (N) and *rbm48* mutant (m) endosperm RNA used for RNA-seq. GRMZM2G103512, *expa2*, and *mrp1* were predicted to show increased levels in *rbm48* mutants, while *gras75*, GRMZM2G006219, and *su2* were predicted to show decreased expression in *rbm48*. The *actin1* locus was used as a loading control. T is a no cDNA template negative control. Gene names are MaizeGDB locus identifiers. (D) Proportion of total transcriptome represented by 39 zein transcripts based on endosperm RNA-seq. Error bars are standard deviation of four biological replicates. (E-G) Quantitative RT-PCR of total RNA extracted from whole kernels in *rbm48* and normal siblings for *α-Zein* (E), *Al9* (F), and *Esr1* (G). Error bars are standard deviation of three biological replicates. (H) Quantitative RT-PCR of *Esr1* expression in total RNA extracted from lower and upper half kernels. Error bars are standard deviation of three biological replicates.

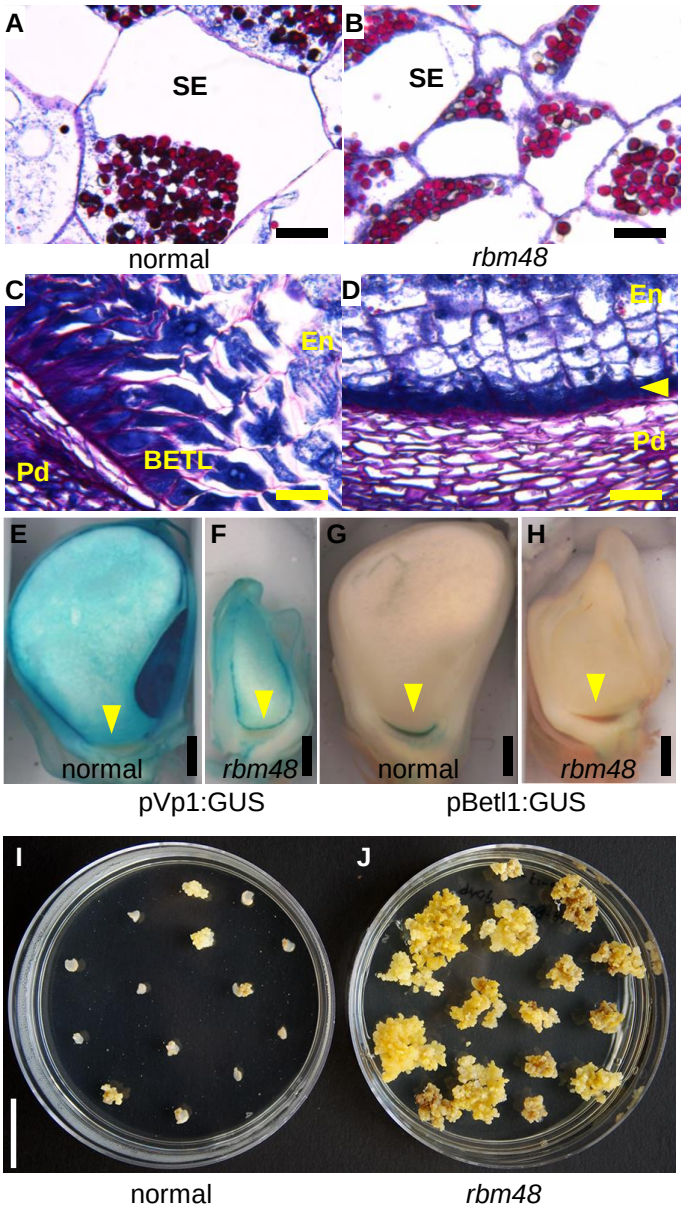

Figure 5. Endosperm cell differentiation requires *rbm48*. (A-D) Transverse sections of 12 DAP normal sibling (A, C) and *rbm48-umu1* (B, D) kernels stained with Schiff's reagent and aniline blue-black. Insoluble carbohydrates in cell walls and starch grains stain fuchsia; nucleoli, nuclei, and cytoplasm stain different intensities of blue. Scale bar is 50 µm. (A-B) Internal starchy endosperm cells (SE). (C-D) BETL region with inner endosperm (En) and maternal pedicel (Pd). Yellow arrowhead indicates small epidermal cells in the BETL region. Scale bar is 50 µm. (E-H) Maize transgene reporter lines for *Vp1* (pVp1:GUS) and *Betl2* (pBetl2:GUS) in normal siblings (E, G) and *rbm48-umu1* mutants (F, H) at 13 DAP. Blue stain indicates GUS expression. Yellow arrowhead indicates the BETL region. Scale bar is 1 mm. (I, J) Tissue culture response of 9 DAP endosperm from normal (I) sibling and *rbm48* (J) seeds. Plates show growth response after 35 days of culture. Scale bar is 2 cm.

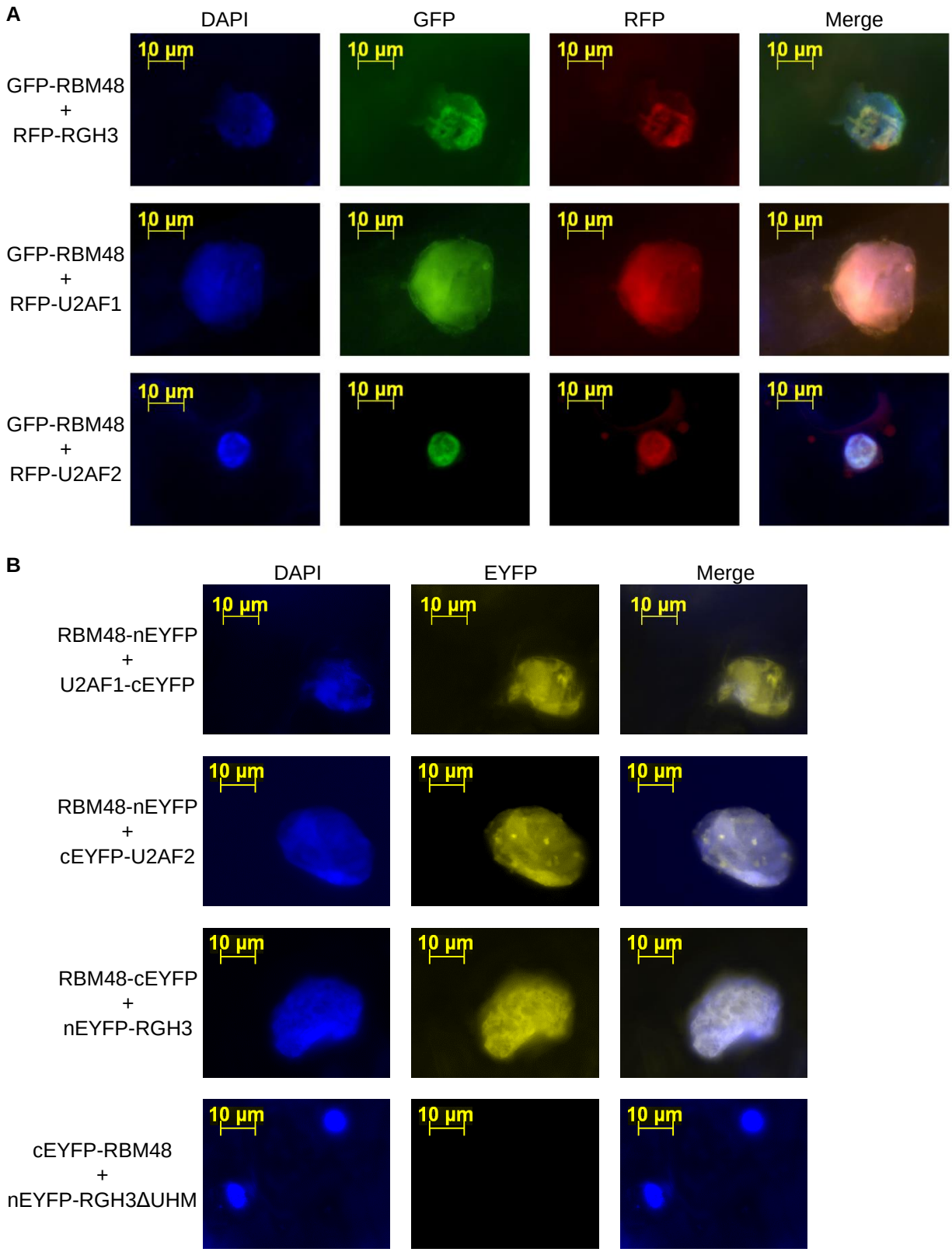

Figure 6. RBM48 co-localizes with U2AF and RGH3. (A) Transient co-expression of GFP-RBM48 with RFP-tagged RGH3, U2AF1, and U2AF2 in *N. benthamiana* leaves. (B) BiFC assays showing transient co-expression fusion proteins with N-terminal EYFP (nEYFP) or C-terminal EYFP (cEYFP). Yellow signal indicates reconstituted EYFP. DAPI staining of DNA marks nuclei in blue.

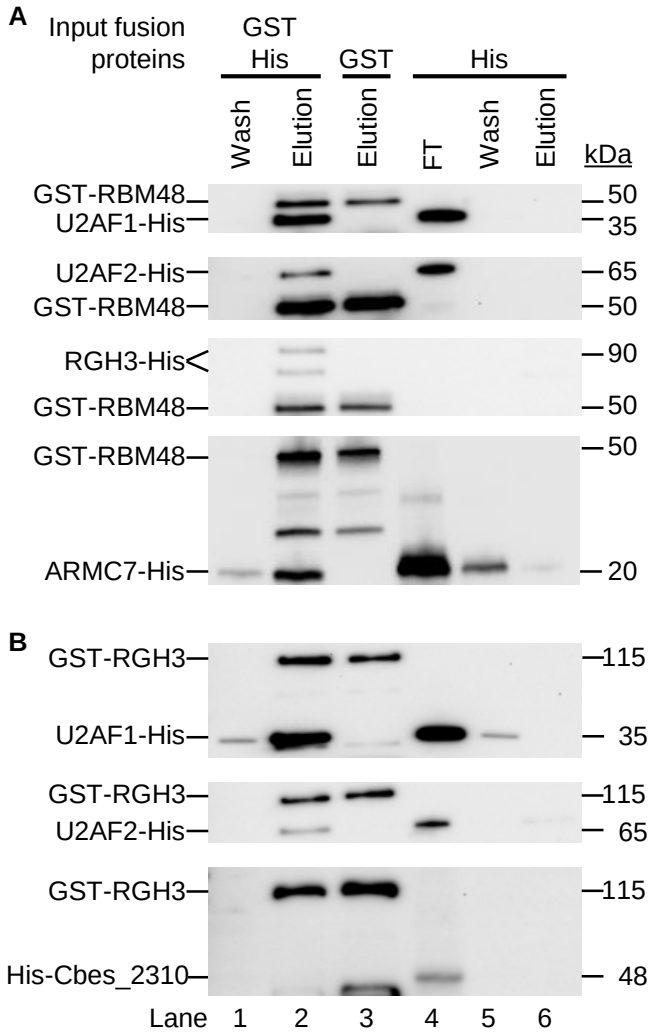

Figure 7. RBM48 interacts with U12 and U2 splicing factors. (A) Western blots of *in vitro* pull down assays with GST-His-tagged RBM48. GST-His-RBM48 immobilized on glutathione beads was incubated with bacterial lysate containing His-tagged U2AF1, U2AF2, RGH3, or ARMC7. (B) Western blots of *in vitro* pull down assays with GST-His-RGH3 immobilized on glutathione beads and bacterial lysates containing His-tagged U2AF1, U2AF2, or Cbes_2310. Recombinant proteins are detected with αHis monoclonal antibodies. Lanes 1-2 show the wash and elution of the pull down assay with both recombinant proteins. Lane 3 is the elution of the GST-His-tagged protein alone. Lanes 4-6 show the level of direct binding of the His-tagged protein with glutathione beads alone. FT is the flow through lysate after incubation with unbound beads.