

# Aberrant splicing in maize *rough endosperm3* reveals a conserved role for U12 splicing in eukaryotic multicellular development

Christine M. Gault<sup>a,b,1,2</sup>, Federico Martin<sup>a,b,1,3</sup>, Wenbin Mei<sup>c</sup>, Fang Bai<sup>b</sup>, Joseph B. Black<sup>b</sup>, W. Brad Barbazuk<sup>a,c,d</sup>, and A. Mark Settles<sup>a,b,d,4</sup>

<sup>a</sup>Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL 32611; <sup>b</sup>Horticultural Sciences Department, University of Florida, Gainesville, FL 32611; <sup>c</sup>Department of Biology, University of Florida, Gainesville, FL 32611; and <sup>d</sup>Genetics Institute, University of Florida, Gainesville, FL 32611

Edited by Julia Bailey-Serres, University of California, Riverside, CA, and approved January 30, 2017 (received for review September 29, 2016)

**RNA splicing of U12-type introns functions in human cell differentiation, but it is not known whether this class of introns has a similar role in plants. The maize ROUGH ENDOSPERM3 (RGH3) protein is orthologous to the human splicing factor, ZRSR2. ZRSR2 mutations are associated with myelodysplastic syndrome (MDS) and cause U12 splicing defects. Maize *rgH3* mutants have aberrant endosperm cell differentiation and proliferation. We found that most U12-type introns are retained or misspliced in *rgH3*. Genes affected in *rgH3* and ZRSR2 mutants identify cell cycle and protein glycosylation as common pathways disrupted. Transcripts with retained U12-type introns can be found in polysomes, suggesting that splicing efficiency can alter protein isoforms. The *rgH3* mutant protein disrupts colocalization with a known ZRSR2-interacting protein, U2AF2. These results indicate conserved function for RGH3/ZRSR2 in U12 splicing and a deeply conserved role for the minor spliceosome to promote cell differentiation from stem cells to terminal fates.**

minor spliceosome | cell differentiation | stem cell | myelodysplastic syndrome | maize endosperm

Most eukaryotic transcripts contain introns that are removed by dynamic ribonucleoprotein complexes known as spliceosomes (1). Spliceosomes include hundreds of RNA-splicing factors that influence splice site selection. Many eukaryotic lineages have two different spliceosomes (2). The major spliceosome removes the common U2-type introns, and the minor spliceosome removes rare U12-type introns. U2- and U12-type introns are recognized by different consensus sequences at the splice sites and branch point (3). The significance of maintaining separate splicing machinery for these rare introns is not well defined. U12-type intron splicing efficiency increases in response to cell stress signaling in HeLa cells (4). Reduction in U12-type splicing causes developmental defects in *Arabidopsis*, *Drosophila*, zebrafish, and humans (5–11).

Human ZRSR2 is required for the second transesterification reaction in U2 splicing and takes part in the initial assembly of the minor spliceosome in human cell extracts (12, 13). Somatic mutations in ZRSR2 are found in patients with myelodysplastic syndrome (MDS) (14–16). MDS is a blood cell differentiation disorder that causes an increase in undifferentiated blasts, abnormal myeloid cells, and a decrease in fully differentiated myeloid cell types (17). MDS subtypes with ZRSR2 mutations more frequently progress to acute myeloid leukemia, and consequently, loss of ZRSR2 is considered a driver toward cancer (17). RNA-sequencing (RNA-seq) analysis of ZRSR2 mutants found reduced U12 splicing with U2-type introns largely unaffected (18).

The hypomorphic *rough endosperm3* (*rgH3*) allele in maize disrupts the ZRSR2 ortholog, seed development, and plant viability (19). Endosperm cell differentiation is defective and delayed in *rgH3* mutants, allowing mutant cells to proliferate in tissue culture at late developmental stages when WT endosperm is unable to grow (19) (Fig. S1). Based on a survey of alternatively spliced

transcripts in maize, only a few genes have been identified with altered transcript isoform abundance in *rgH3* mutants (19).

## Results

**Missplicing of U12-Type Introns in *rgH3*.** We determined the genome-wide effect of *rgH3* on mRNA splicing with RNA-seq and isoform expression analysis from homozygous *rgH3* and WT sibling root and shoot tissues. Cufflinks (20) predicted 46 genes had altered isoform use in *rgH3* (Table S1). Localized differences in read coverage depth between *rgH3* and WT libraries were used to design semiquantitative RT-PCR assays for nine of the genes identified as differentially spliced (Fig. 1A and Fig. S2). Sequencing of the amplified products revealed that intron retention occurs more frequently in *rgH3* mutants at seven of these loci (Fig. 1B and Figs. S3 and S4). Some *rgH3* amplifications exhibited slow-migrating bands due to heteroduplex formation of differently sized RT-PCR products (Fig. S5). Six of the validated splicing differences have U12-type introns as defined by the ERISdb plant splice site database (21). The intron retained in

## Significance

The last eukaryotic common ancestor had two spliceosomes. The major spliceosome acts on nearly all introns, whereas the minor spliceosome removes rare, U12-type introns. Based on in vitro RNA-splicing assays, the RGH3/ZRSR2 RNA-splicing factor has functions in both spliceosomes. Here, we show that the maize *rgH3* mutant allele primarily disrupts U12 splicing, similar to human ZRSR2 mutants, indicating a conserved in vivo function in the minor spliceosome. These mutant alleles block cell differentiation leading to overaccumulation of stem cells in endosperm and blood, respectively. We found extensive conservation between maize and human U12-type intron-containing genes, demonstrating that a common genetic architecture controls at least a subset of cell differentiation pathways in both plants and animals.

Author contributions: C.M.G., F.M., W.B.B., and A.M.S. designed research; C.M.G., F.M., F.B., J.B.B., and A.M.S. performed research; C.M.G., F.M., W.M., W.B.B., and A.M.S. analyzed data; and C.M.G., F.M., W.B.B., and A.M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The RNA-seq data files reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo) (accession no. GSE57466).

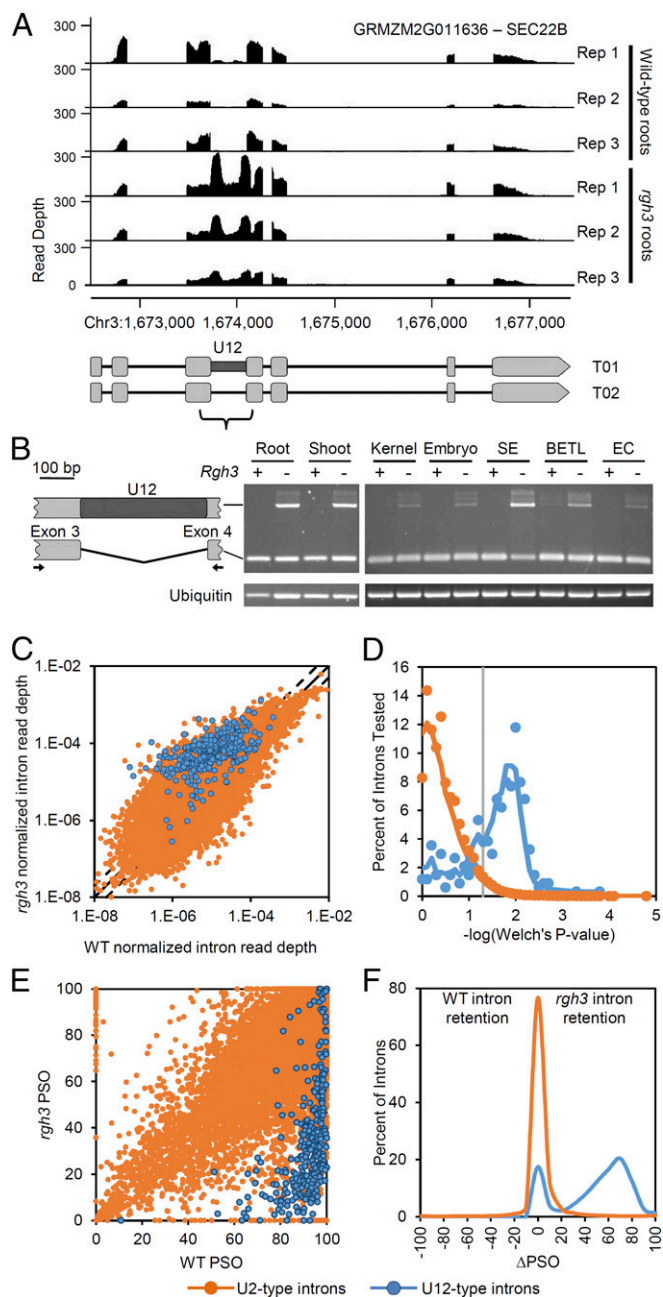
<sup>1</sup>C.M.G. and F.M. contributed equally to this work.

<sup>2</sup>Present address: Institute of Genomic Diversity, Cornell University, Ithaca, NY 14853.

<sup>3</sup>Present address: Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80511.

<sup>4</sup>To whom correspondence should be addressed. Email: [settles@ufl.edu](mailto:settles@ufl.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616173114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616173114/-DCSupplemental).



**Fig. 1.** Minor splicing is compromised in *rgh3*. (A) WT and *rgh3* root RNA-seq read depth for GRMZM2G011636. Transcript models show the U12-type intron with increased read depth in *rgh3*. A brace symbol indicates the region amplified in B. (B) RT-PCR of normal (*Rgh3* +) and *rgh3* mutant (*Rgh3* -) RNA from root, shoot, kernel, embryo, starchy endosperm (SE), basal endosperm transfer cell layer (BETL), and endosperm culture (EC) tissues. Schematic shows sequenced amplification products with forward and reverse primers indicated by arrows. (C–F) Intron read depth and PSO metrics with U2-type introns indicated in orange and U12-type introns indicated in blue. (C) Scatterplot showing intron read depth normalized for gene expression. Black diagonal line indicates equivalent read depth, and dotted lines are twofold differences. (D) Distribution of Welch's *t* test *P* values. Lines were fit with a sliding window average of 0.3 log units. Gray line, *P* = 0.05. (E) Scatterplot of PSO metrics. Introns with less than 10 exon-exon junction reads in both WT and *rgh3* are not plotted. (F) Distribution of splicing differences between WT and *rgh3* based on  $\Delta$ PSO metrics.

the seventh gene, GRMZM2G133028, is likely to be a U12-type intron. Although the 5'-splice site and branch point in GRMZM2G133028 are somewhat diverged from the U12-type

consensus, the six additional plant species analyzed within ERISdb have U12-type introns at the orthologous intron position in this conserved, plant-specific gene (Fig. S6).

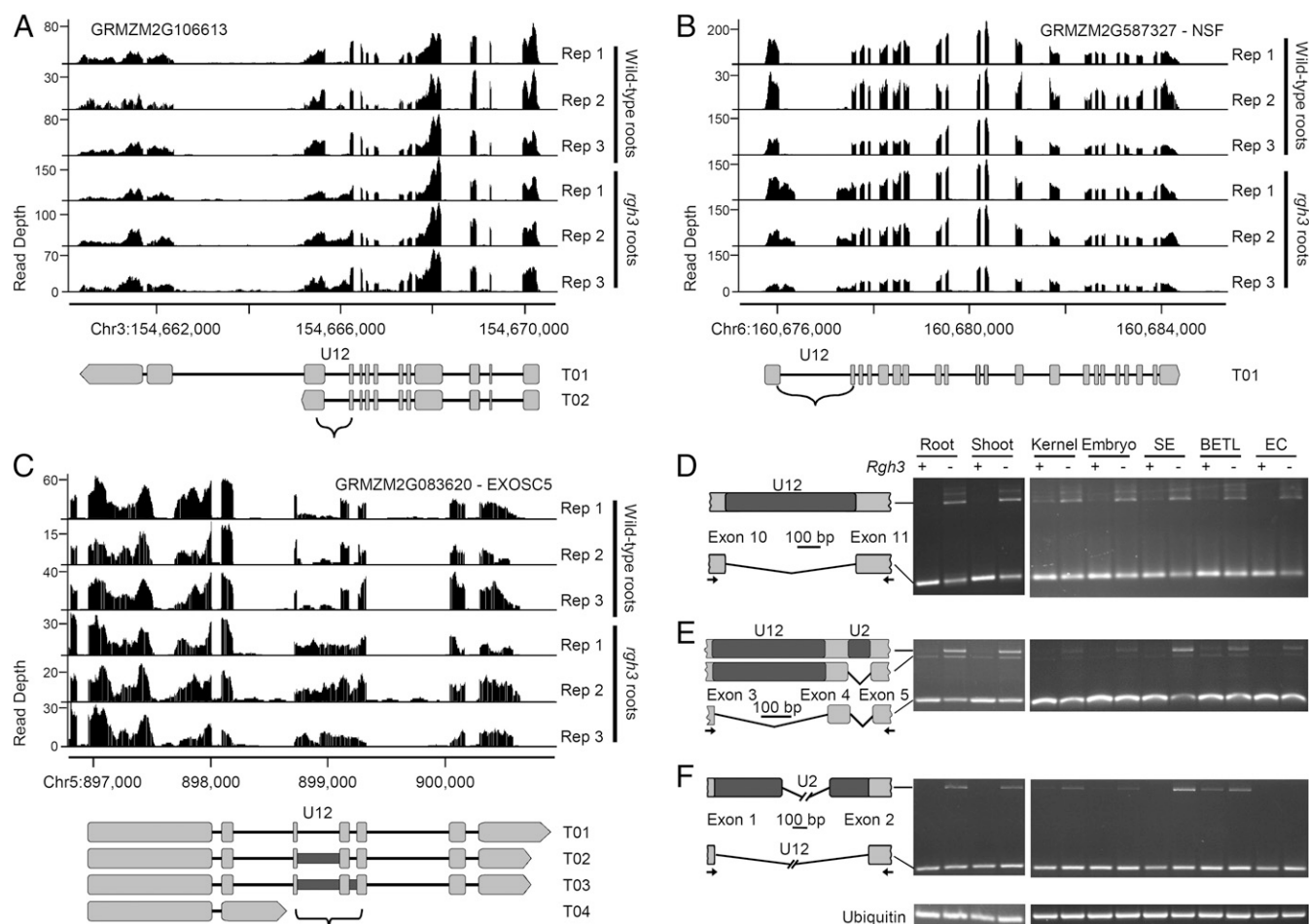
U12-type intron-containing genes account for <2% of the B73 filtered gene set. With 20% of Cufflinks predictions involving U12-type introns, we hypothesized that U12-type introns are misspliced on a global scale in *rgh3*. To allow for novel transcript isoforms to be detected, we analyzed intron read counts normalized by gene expression for all introns in the filtered gene set of the maize genome (Fig. 1C). Of the 446 nonredundant U12-type introns in ERISdb, 340 were expressed in the seedling libraries. Significantly more RNA-seq reads mapped to 240 U12-type introns in *rgh3* libraries vs. WT libraries (Fig. 1D and Dataset S1). This represents 71% of the maize U12-type introns tested. Only three U12-type introns showed significantly fewer reads in *rgh3* libraries, which is below the expected false-positive rate for the number of *t* tests completed. By contrast, 4% of the remaining 113,345 nonredundant introns within expressed genes had differences in the number of normalized reads mapped in *rgh3* vs. WT (Fig. 1D and Dataset S2). Globally, fewer U2-type introns showed significant differences than the expected number of false-positive *t* tests, indicating that *rgh3* specifically affects genes with U12-type introns.

To quantify the extent of intron splicing defects, we calculated percent spliced out (PSO) for individual introns using exon-exon junction and intron reads (Fig. 1E). Based on Fisher's exact tests of read counts, splicing defects were detected for 77% of U12-type introns (Dataset S1). The median difference between WT and *rgh3* PSO ( $\Delta$ PSO) was 62%, indicating extensive retention of U12-type introns in *rgh3* (Fig. 1F). More than 13% of all other introns also showed statistically significant increased retention in *rgh3*, but the median  $\Delta$ PSO for these introns was only 4%, indicating minor impacts on U2 splicing (Dataset S2).

An additional nine genes showing significant differences in U12-type intron read depth and PSO were randomly selected for RT-PCR (Fig. 2A–C and Fig. S7). All of these genes had splicing defects in *rgh3* with three patterns of altered splicing: intron retention (Fig. 2D and Fig. S8A–D), missplicing of the U12-type intron concomitant with retention of a downstream U2-type intron (Fig. 2E and Fig. S8E), and activation of cryptic, U2-type, 5'- and 3'-splice sites at the U12-type intron (Fig. 2F and Fig. S8E and F). There were no differences in splice site consensus sequences between the U12-type introns that are misspliced in *rgh3* and those with no significant difference in *rgh3* mutants (Fig. S9). The range of splicing defects found in *rgh3* is also observed in human bone marrow samples from MDS patients with *ZRSR2* mutations (18). We conclude that *rgh3* mutants, like human *ZRSR2* mutants, are impaired in minor spliceosome function.

**Predicted Functions of *rgh3* Misspliced Genes.** Pfam domains in the 230 maize genes with increased intron read depth in *rgh3* showed no significant enrichment or deenrichment compared with all 308 expressed U12-type intron-containing genes in the RNA-seq experiment (Dataset S3). These analyses suggest most biological processes that are dependent upon U12 splicing are affected in *rgh3*. Many U12-type introns are found within genes involved in DNA replication, DNA repair, transcription, RNA processing, and translation (22, 23). *Arabidopsis* U12-type introns were used for prior cross-kingdom comparisons, but only 58% of maize U12-type intron-containing genes have an *Arabidopsis* ortholog that also contains a U12-type intron (21).

Pfam domains were analyzed at a global level to identify U12-dependent biological processes affected in *rgh3* (Fig. 3A and Dataset S3). More than one-half of the domains with annotated roles in translation, endomembrane dynamics, and unknown functions were enriched in the 230 *rgh3* misspliced genes relative to all genes tested for intron read depth differences. When all U12-type intron-containing genes are compared genome-wide, more than one-half of domains with roles in cell cycle, RNA processing,



**Fig. 2.** Diverse *rgh3* splicing defects are associated with U12-type introns. (A–C) WT and *rgh3* root RNA-seq read depth for (A) GRMZM2G106613, (B) GRMZM2G083620, and (C) GRMZM2G587327. Transcript models show U12-type introns with a brace symbol indicating the regions amplified in D–F. (D–F) RT-PCR of normal (*Rgh3* +) and *rgh3* mutant (*Rgh3* –) RNA from root, shoot, kernel, embryo, starchy endosperm (SE), basal endosperm transfer cell layer (BETL), and endosperm culture (EC) tissues. Schematics show sequenced amplification products with forward and reverse primers indicated by arrows. (D) GRMZM2G106613 showed U12-type intron retention in *rgh3*. (E) GRMZM2G083620 had adjacent U12- and U2-type intron retention. (F) GRMZM2G587327 activated cryptic U2-type splice sites in *rgh3*.

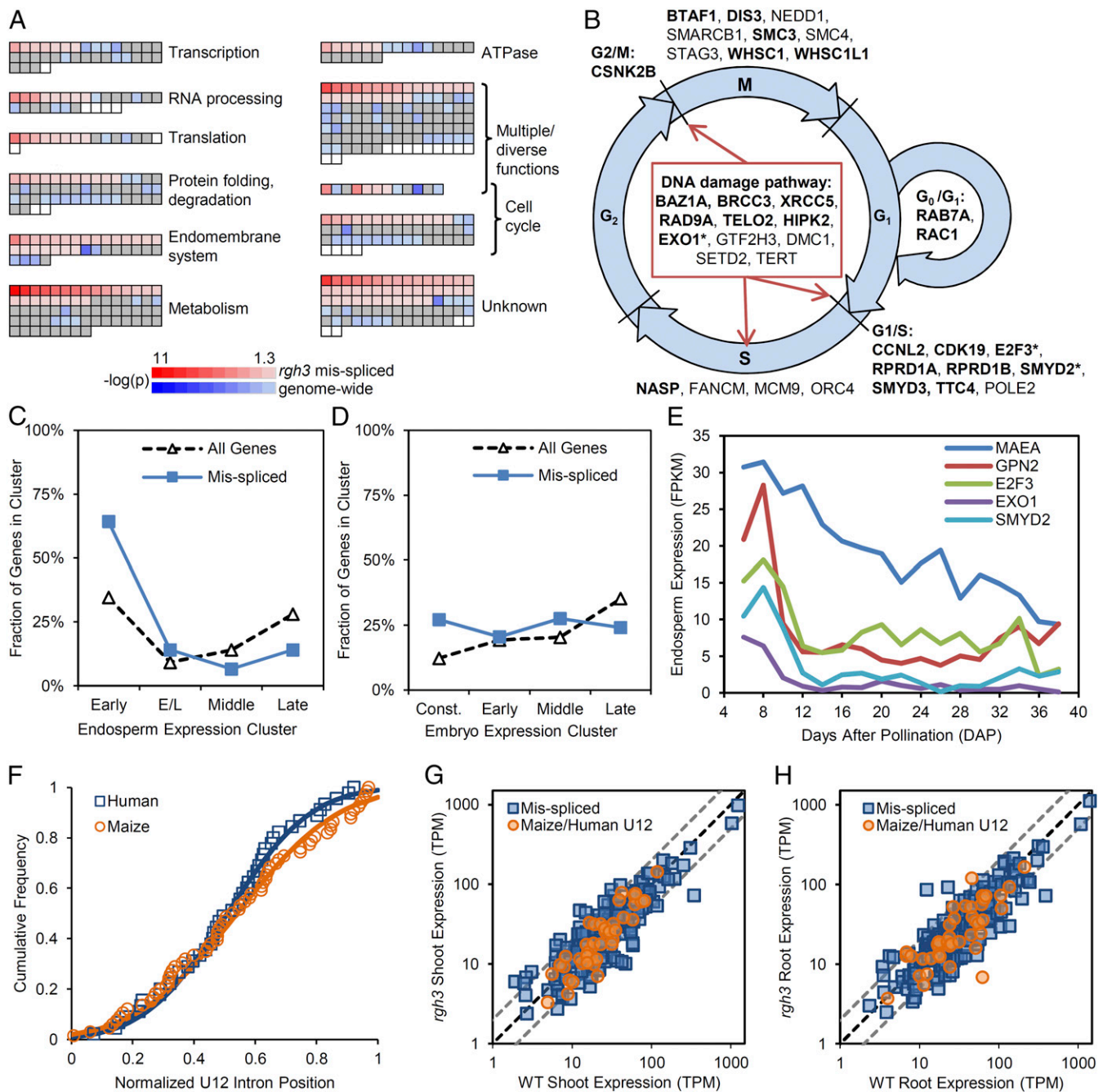
and protein folding/degradation also are enriched (Fig. 3A). These analyses support conserved functions for maize U12-type intron-containing genes with U12 spliceosome targets in other species.

Endosperm cells in *rgh3* show aberrant differentiation of embryo surrounding region (ESR) and basal endosperm transfer layer (BETL) cells (19), which is analogous to blood cell differentiation defects in MDS patients (18). BETL and ESR differentiation occurs early in seed development. To examine developmental expression of *rgh3*-affected genes, we reanalyzed a transcriptome profile of maize seed development (24). Based on reported clusters of expression profiles, the 230 misspliced genes in *rgh3* were overrepresented in early endosperm but not in early embryo development (Fig. 3C and D). Approximately 78% of misspliced genes have a peak of expression between 6 and 12 d after pollination (Fig. 3E). These data are consistent with U12-type intron-containing genes playing a role in endosperm cell differentiation.

To explore mechanistic similarities between U12-dependent cell differentiation pathways in maize and human, we identified human homologs of maize genes with U12-type introns. BLASTP searches of human RefSeq proteins identified 233 maize U12-type intron-containing genes with a human homolog (Dataset S4). The 233 maize genes correspond to 154 human genes due to differences in gene redundancy in the maize and human genomes. We found two biological processes that account for 35% of the U12-type intron-

containing maize genes with human homologs. First, 50 maize U12-type intron-containing genes have roles in cell cycle. Of these cell cycle genes, 29 are differentially spliced in *rgh3*, which correspond to 24 human homologs (Fig. 3B). Second, 33 maize genes with U12-type introns have predicted roles in protein glycosylation including the following: synthesis and transport UDP-xylose, synthesis of dolichyl-diphosphooligosaccharide, secretion of glycosylated protein complexes, and the unfolded protein response pathway (Table S2). Twenty-five of the protein glycosylation genes are misspliced in *rgh3* and correspond to 10 human homologs.

To find genes with altered splicing in both *rgh3* and human *ZRSR2* mutants, we identified the current gene symbols for human U12-intron-containing genes in U12DB (Dataset S5) (25). This revealed 36 human genes that are homologs of 57 maize genes with U12-type introns in both species. Maize/human homology enriches for U12 splicing defects in both *rgh3* and *ZRSR2* mutants (Table S3). Fifty of these maize genes were tested in the RNA-seq analysis, and 96% (48 of 50) had evidence of splicing defects in *rgh3*, as determined by intron read depth or junction read tests. Of the 758 human genes in U12DB, only 216 (28%) genes had significant splicing defects reported in *ZRSR2* mutants, but 47% (17 of 36) of human U12-type intron-containing genes with maize homologs have splicing defects (18).



**Fig. 3.** U12 splicing defects in *rgh3*-affected genes involved in cell differentiation and growth. (A) Heat map of Pfam domains found in U12-type intron-containing genes. Red domains are enriched in genes with *rgh3* U12 splicing defects relative to all maize genes tested for splicing defects. Blue indicates additional domains enriched in U12-type intron-containing genes relative to all maize genes. Gray and white indicate no enrichment of the domain relative either to genes tested for splicing defects or to all maize genes, respectively. (B) Cell cycle schematic showing human homolog gene symbols for maize U12-type intron-containing genes. Bold indicates U12 splicing defects in *rgh3*. Asterisks indicate splicing defects in human *ZRSR2* mutants. (C–E) Endosperm expression of *rgh3* misspliced genes (24). Cluster analysis of all genes (open triangles) and *rgh3* misspliced genes (blue squares) is plotted for embryo (C) and endosperm (D). E/L, early/late developmental expression. Const., constitutive expression. (E) Example endosperm expression profiles for maize/human homologs with U12-type introns. (F) Cumulative frequency plot for U12-type intron positions in maize proteins (orange circles) with human homologs (blue squares) containing U12-type introns. Orange (maize) and blue (human) lines plot the expected normal distribution of each protein set. (G and H) Scatterplots of expression levels for misspliced genes with human homologs that also have U12-type introns (orange circles) and all other misspliced genes (blue squares). Dotted black line shows 1:1 ratio of WT to *rgh3* expression. Dotted gray lines show a twofold ratio change.

Mutations in three of these genes, GPN2, MAEA, and TAPT1, have documented roles in animal and plant development. The *Arabidopsis* homolog of GPN2 encodes the QOT1 protein and is required for embryos to complete periclinal divisions to establish epidermal and internal cell layers (26). The mouse ortholog of

MAEA is required for final differentiation of erythrocytes (27). The vertebrate TAPT1 gene is required for skeletal patterning and normal function of the primary cilium (28). The *Arabidopsis* TAPT1 ortholog, *PODI*, is required for apical–basal patterning of the early embryo and endomembrane protein sorting (29).

Our analysis also revealed divergent loss of U12-type introns for genes with conserved biological functions (Datasets S4 and S5). The protein glycosylation enzymes, ALG12 and PIGB, have U12-type introns in both maize and human. U12-type introns are also found in other protein glycosylation genes including the following: ALG3, ALG6, ALG8, PIGN, and PIGP in human as well as the maize homologs of DDOST, MPDU1, MGAT1, RFT1, and PIGM. The DNA origin of replication complex subunit, ORC3, in humans contains a U12-type intron, whereas the maize ORC4 homolog contains a U12-type intron. For ubiquitin-specific peptidases, U12-type introns are found in human USP7, USP10, and USP14 as well as in maize homologs of USP36 and USP42. Similarly, molybdenum cofactor (Moco) biosynthesis genes contain U12-type introns in maize. In humans, two Moco cofactor-requiring enzymes, aldehyde oxidase (AOX1) and xanthine dehydrogenase (XDH), contain U12-type introns. Loss of Moco in maize does not affect endosperm development but does cause seedling lethality (30, 31). There are at least 25 additional examples where different members of protein complexes or biochemical pathways have U12-type introns in human and maize including TRAPP and adaptor related protein complexes, Rab, importins, DNA polymerase, activating signal cointegrator-1, TFIIA, RNA polymerase III, the exosome, ribosomal proteins, and the autophagy pathway. These homologies suggest a common genetic architecture of minor spliceosome targets in human and maize where splicing defects in U12-type introns disrupt conserved biological processes and result in stem cell-like phenotypes in *ZRSR2* mutant MDS cells and *rgl3* endosperm cells.

**Conservation of U12-Type Intron Positions in Maize and Human.** Divergence of gene sets with U12-type introns in a given species is almost exclusively due to loss of U12-type introns or mutation toward a U2-type intron (22, 32). Protein alignments of the human and maize isoforms were used to identify U12-type intron positions for maize/human homologs (Fig. S10 and Dataset S4). Approximately one-half of the maize/human homologs (19 of 36 human genes and 26 of 57 maize genes) have at least one U12-type intron in a conserved position within the protein coding sequence (Fig. S11 and Table S3). The U12-type introns are randomly distributed within the coding sequences of both maize and human homologs (Fig. 3F). The intron positions fit the normal distribution (Shapiro-Wilk  $P > 0.05$ ) with very little skew (human =  $-0.04$ ; maize =  $-0.03$ ) but have a low kurtosis (human =  $-0.53$ ; maize =  $-0.91$ ). The near-uniform distribution suggests a diversity of coding sequence impacts with most defective transcripts expected to trigger nonsense-mediated decay (NMD). However, *rgl3* splicing defects do not appear to significantly alter transcript abundance. For the *rgl3* misspliced genes, 86 and 79% of shoot and root transcripts accumulate within a twofold range of WT, respectively (Fig. 3G and H). Near-equivalent transcript levels were also observed for misspliced genes with a U12DB human homolog.

RT-PCR from purified nuclei and polysomes indicates that some misspliced transcripts in *rgl3* are likely to be translated. Intron retention and misspliced transcripts copurified with polysomes when the U12-type intron is the last intron of the transcript, such as in maize homologs of the E2F3 cell-cycle transcription factor and TRAPPC2 (Fig. 4A and B). The U12 splicing defects in GRMZM2G033430 and PQLC2 also copurify with polysomes in *rgl3* mutants (Fig. 4C and Fig. S124). These splice variants may not be NMD targets, because the termination codons are relatively close to the last exon-exon junction. By contrast, transcripts with multiple exon-exon junctions downstream of the U12 splicing defect as well as control transcripts expected to be retained in the nucleus were enriched in nuclei and excluded from polysomes (Fig. 4D–F and Fig. S12B–E).

Missplicing of E2F3 was identified as a candidate for mediating human myeloid malignancies observed in *ZRSR2* mutant cells (18). U12-type intron retention in the maize E2F3 homologs

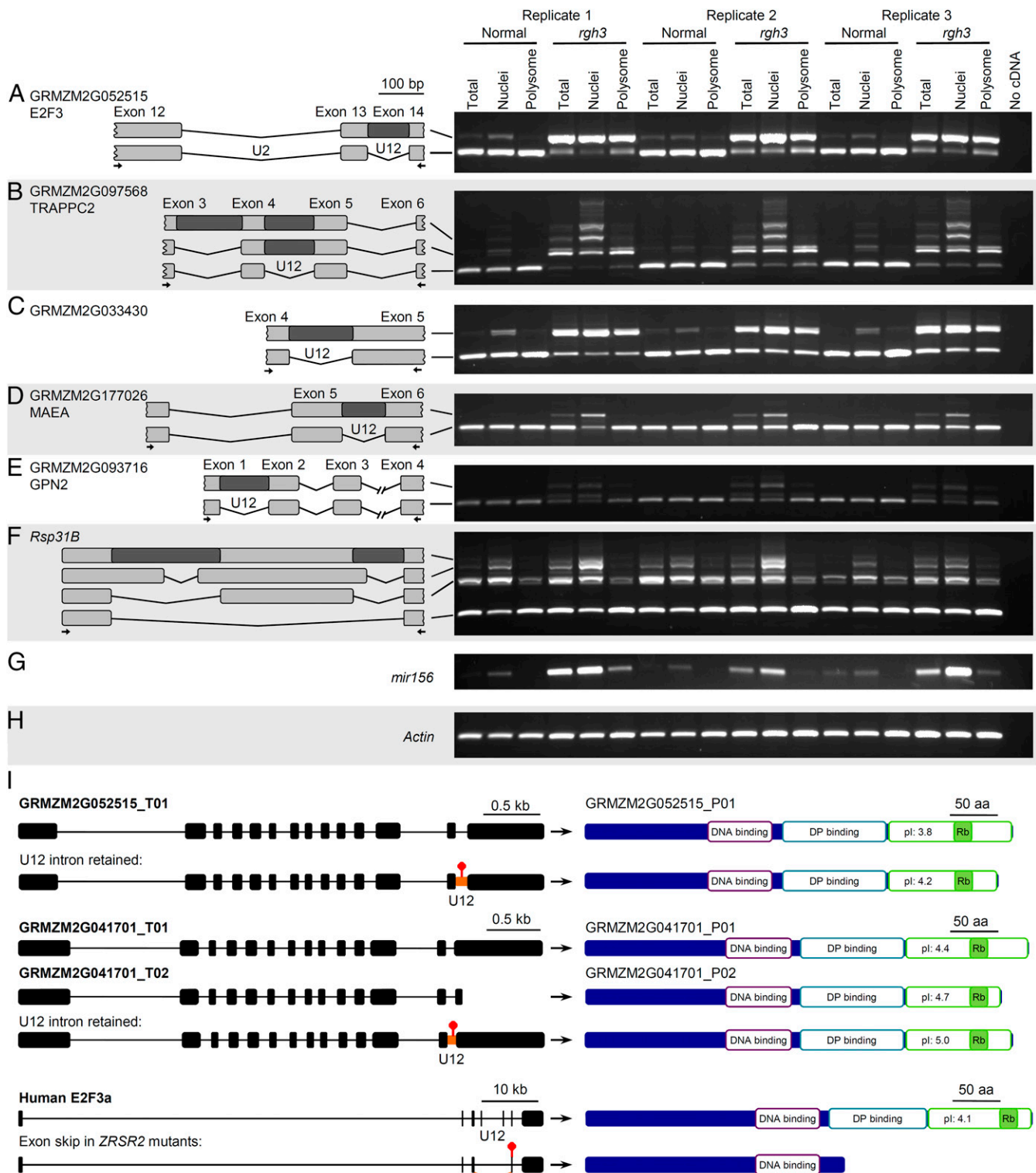
truncate the C-terminal transactivation domain similar to the endogenous GRMZM2G041701\_T02 isoform (Fig. 4I). The predicted E2F proteins from the intron retention transcripts contain all domains necessary for cell proliferation and the endocycle in plants (33). By contrast, the splicing defects observed in human E2F3 are primarily expected to be NMD targets, except for a transcript that skips exons 3–4, which flank the U12-type intron (18). If translated, the human exon skip transcript would produce a nonfunctional E2F3 protein lacking the DP-dimerization and transactivation domains. This difference in U12-type intron position may contribute to the contrasting cell proliferation phenotypes with *rgl3* being proliferative in culture and *ZRSR2* mutants arresting the cell cycle.

**Mutant RGH3 Proteins Disrupt Localization with U2AF2.** It is surprising that *rgl3* disrupts a larger proportion of U12-type introns than *ZRSR2* mutations in MDS patients. The *rgl3* mutant is predicted to encode a weak allele, whereas most *ZRSR2* mutations are loss-of-function alleles (14–16, 19). The *rgl3-umu1* allele has a *Mutator* transposon insertion that partially splices from the transcript to delete 12 aa and insert 47 aa coded by the transposon sequence in the N-terminal acidic domain of the RGH3 protein (Fig. 5A). The normal *Rgl3* allele produces multiple splice variants. Only the *Rgl3 $\alpha$*  isoform encodes a full-length ortholog of ZRSR2 (19). Both the mutant and splice isoforms coding for protein truncations are likely to be expressed as protein in maize, because RGH3 antibodies cross-react with multiple proteins that migrate similarly to in vitro transcribed/translated cDNA clones of the alternatively spliced variants (Fig. S13). We investigated subcellular localization of RGH3 protein variants to gain additional insight into the nature of the *rgl3* mutant allele.

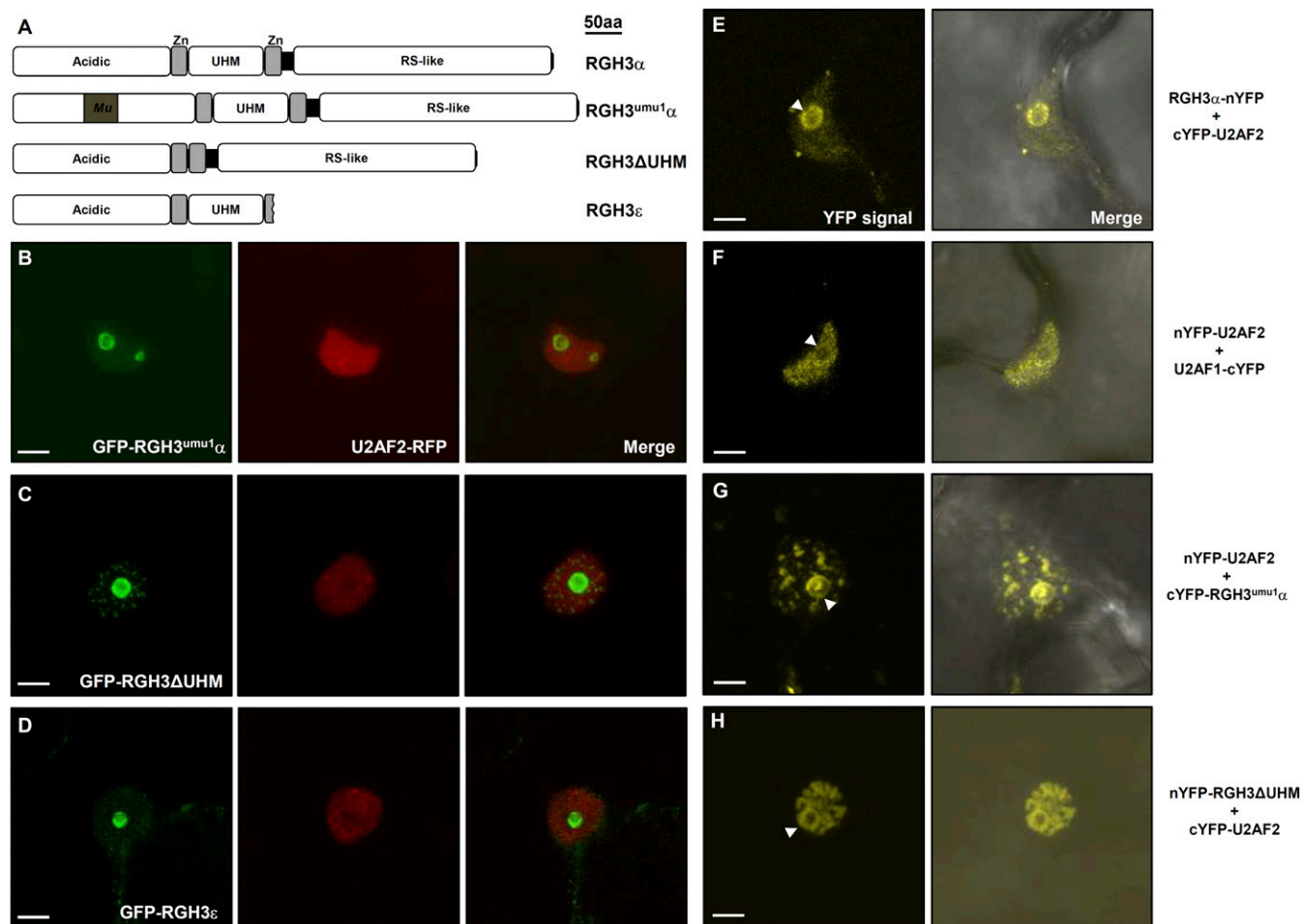
Human ZRSR2 interacts with the major spliceosome subunit, U2AF2, through a U2AF homology motif (UHM) that is related to RNA recognition motifs but mediates protein-protein interactions (13, 34). These interactions are likely conserved in maize (19). Colocalization is observed when maize RGH3 $\alpha$  and U2AF2 are transiently expressed in *Nicotiana benthamiana* as GFP and RFP fusions (Fig. S144). U2AF2 is the large subunit of U2 auxiliary factor (U2AF), and maize U2AF2 also colocalizes with the small subunit of U2AF, U2AF1 (Fig. S14B). These data are consistent with maize RGH3 $\alpha$  and U2AF2 acting in the same subnuclear compartment.

The mutant, RGH3<sup>umu1</sup> $\alpha$ , protein has aberrant subnuclear localization with low levels of diffuse signal in the nucleoplasm (Fig. 5B) instead of localizing to spliceosomal speckles as seen for RGH3 $\alpha$  (Fig. S144). When RGH3<sup>umu1</sup> $\alpha$  is coexpressed with U2AF2, the proteins typically localize to different subnuclear compartments. Some cells showed partial overlap of RFP and GFP fusions, indicating reduced colocalization (Fig. S14C). An in-frame deletion of the UHM domain (RGH3 $\Delta$ UHM) showed intermittent colocalization like RGH3<sup>umu1</sup> $\alpha$  (Fig. 5C and Fig. S14D). Splice isoforms coding for protein truncations of the RS domain (RGH3 $\epsilon$ ) or the UHM and RS domain (RGH3 $\beta$ , RGH3 $\gamma$ ) did not show overlap with U2AF2 (Fig. 5D and Fig. S14E and F).

Bimolecular fluorescent complementation (BiFC) assays of U2AF2 with RGH3 $\alpha$ , U2AF1, RGH3<sup>umu1</sup> $\alpha$ , and RGH3 $\Delta$ UHM all resulted in YFP signal in the nucleus (Fig. 5E–H and Fig. S15). Similar to colocalization experiments, U2AF2 and RGH3 $\alpha$  showed YFP signal in nuclear speckles and the nucleolus, whereas U2AF1 and U2AF2 had YFP signal in nuclear speckles. By contrast, BiFC signals from U2AF2 with RGH3<sup>umu1</sup> $\alpha$  or RGH3 $\Delta$ UHM appear aggregated in larger subnuclear foci. Reconstitution of YFP in BiFC assays is irreversible with transient interactions able to give stable YFP signal (35). These data show that U2AF1, U2AF2, and RGH3 colocalize as predicted from human protein-protein interaction studies. Mutations or truncations affecting the acidic, UHM, or RS-domain all disrupt the dynamic colocalization of



**Fig. 4.** A subset of *rgh3* misspliced transcripts may be translated. (A–H) RT-PCR analysis using RNA extracted from total, nuclei, and polysome fractions of normal and *rgh3* samples. U12-type intron regions were amplified from the following: (A) GRMZM2G052515, (B) GRMZM2G097568, (C) GRMZM2G033430, (D) GRMZM2G177026, and (E) GRMZM2G093716. The larger two mRNA variants of *Rsp31B* (F) and *mir156* premiRNA (G) are expected to be excluded from polysomes, whereas actin (H) serves as loading control. All transcript diagrams are drawn to the same scale as in A with U12-type introns indicated. Arrows show forward and reverse primers used for RT-PCR. (I) Transcript and protein diagrams for maize *E2F3* genes and their human *E2F3a* homolog. Retention of the maize U12-type introns introduce premature termination codons (red octagon) that would have minimal effect on E2F protein domains. Retention and cryptic splicing of the U12-type intron in human are predicted NMD targets, whereas skipping exons 4–5 is predicted to produce a nonfunctional protein.



**Fig. 5.** Multiple RGH3 protein domains are necessary for colocalization with U2AF2. (A) Protein domain schematic of RGH3 isoforms and UHM domain deletion tested in colocalization assays. (B–D) Transient colocalization of U2AF2 with mutant *RGH3<sup>umu1</sup>α* allele (B), *RGH3ΔUHM* domain deletion (C), and the *RGH3ε* isoform (D). (E–H) Transient expression of BiFC constructs: cYFP-U2AF2 with *RGH3α*-nYFP (E), nYFP-U2AF2 with U2AF1-cYFP (F), nYFP-U2AF2 with cYFP-*RGH3<sup>umu1</sup>α* (G), and cYFP-U2AF2 with nYFP-*RGH3ΔUHM* (H). White arrowheads point to nucleolus. [Scale bar: 5 μm (in all microscopy images).]

RGH3 with U2AF2 equivalently. Combined with the RNA-seq results, these localization experiments suggest that *rgH3-umu1* is more likely a strong allele and indicate an important role for the acidic domain of ZRSR2/RGH3 in U12 splicing.

## Discussion

**RGH3/ZRSR2 Are U12 Splicing Factors in Vivo.** Our data provide an independent genetic analysis of ZRSR2 function in a distantly related species from humans. RGH3/ZRSR2 has a conserved role in splicing U12-type intron-containing genes. *ZRSR2* and *rgH3* mutants exhibit U12-type intron retention, activation of cryptic 5'- and 3'-splice sites within U12-type introns, and retention of U2-type introns that are adjacent to misspliced U12-type introns.

A near-exclusive *in vivo* function for ZRSR2/RGH3 in U12 splicing contradicts biochemical experiments that conclusively show ZRSR2 copurifies and is required in both spliceosomes (12, 13). It is possible that copurification of ZRSR2 and U2AF in human cell extracts is due to independent binding of common pre-mRNA species. However, ZRSR2 interacts with U2AF2 in yeast two-hybrid assays (13). Yeast lacks U12-type introns, suggesting direct protein–protein interactions are more likely between these spliceosome subunits. In maize, colocalization of RGH3 with U2AF2 appears to be an indicator of RGH3 function in the minor spliceosome. Potentially, interactions between the major and minor spliceosomes promote efficient splicing of either class of intron.

**Minor Splicing as a Regulatory Process.** Minor splicing factors are at low abundance, and U12 splicing can be a rate-limiting step to produce protein-coding mRNA (36). For example, U6atac levels in HeLa cells affect splicing efficiency as well as expression level of genes with U12-type introns, and the minor spliceosome was proposed to regulate cellular responses to stimuli (4). Under this model, minor spliceosome activity determines the balance of coding mRNA vs. NMD or translation into alternative protein isoforms (4). We found little evidence of splicing defects resulting in expression level changes of maize U12-type intron-containing genes. Similarly, *Drosophila* U6atac mutants have little impact on U12-type intron-containing gene expression levels (37). In maize, reduced U12 splicing generally leads to predicted NMD targets being retained in the nucleus. A smaller subset of U12-type intron retention transcripts in *rgH3* are associated with polysomes and are likely to be translated into alternative protein isoforms.

Although different molecular mechanisms seem to act downstream of U12 splicing efficiency in maize and humans, there are a substantial number of homologous genes with U12-type introns. We found 233 maize U12-type intron-containing genes with easily identified human homologs representing 154 human genes. Nearly 25% of the maize–human homolog pairs contain U12-type introns in both species (57 of 233 for maize genes or 36 of 154 human homologs). There are also overlapping functions among non-homologous genes subject to U12 splicing. These nonhomologous

overlapping genes are subunits of the same protein complexes, members of conserved gene families, or members of the same metabolic pathways. These overlaps suggest selection for specific biological processes, such as cell cycle and protein glycosylation, to be dependent upon U12 splicing efficiency and support the idea that the minor spliceosome could be regulatory.

**Minor Splicing Is Required for Cell Differentiation.** *ZRSR2* and *rgl3* mutants both disrupt cell differentiation programs. MDS patients with *ZRSR2* mutations accumulate myeloid blast precursors (18). Maize *rgl3* endosperm retains proliferative capacity and shows cell fate switching to aleurone in the basal endosperm transfer cell layer and the embryo-surrounding region (19). Mutations in other minor spliceosome factors lead to developmental defects in many species. In humans, Taybi-Linder syndrome is caused by mutations in *U4atac*, resulting in severe bone abnormalities and microcephaly (6, 7). Reduced U12 splicing efficiency may also be the primary molecular cause of spinal muscular atrophy, which affects the peripheral nervous system (38, 39). Zebrafish mutations in *RNPC3* lead to defects in endodermal organ development and aberrant intestinal epithelium morphology (8). *Drosophila* mutants in *U6atac* and U12 snRNA are lethal in third-instar larvae, when adult metamorphosis occurs (5). Artificial microRNA (amiRNA) down-regulation of *Arabidopsis* U12 splicing factors show defective leaf morphology and arrest inflorescence development (9–11). These phenotypes all suggest that reduced U12 splicing is not immediately lethal to the cell but rather disrupts essential developmental processes.

Despite the many U12 splicing phenotypes reported, no unifying developmental function has been ascribed to minor splicing. By focusing on individual tissues such as endosperm or blood, a common function for *Rgl3* and *ZRSR2* in cell differentiation becomes clear (18, 19). Extending to a more general interpretation of U12 splicing mutant phenotypes, the data suggest a role for U12 splicing to promote differentiation of a subset of cell types in both plants and animals. It is unlikely for U12 splicing to be needed in all differentiation processes, because the minor spliceosome has been lost in multiple eukaryotic lineages (2).

**Conservation of Minor Splicing During Evolution.** The losses of minor splicing during evolution raise the possibility that the minor spliceosome evolved in a convergent manner to function in plant and animal cell differentiation. However, there appears to be more selective pressure to maintain U12 splicing in multicellular eukaryotes. Although the minor spliceosome is missing in some lineages with multicellular development, a larger fraction of unicellular eukaryotic genomes have lost U12 splicing (2, 40). The few unicellular species that retain the minor spliceosome, such as *Acanthamoeba castellanii*, tend to have amoeboid cellular organization (2, 40). The *A. castellanii* genome sequence revealed a high frequency of horizontal gene transfer, including potential eukaryote-to-eukaryote gene transfer that would confound deep evolutionary comparisons (41).

Volvocine green algae illustrate convergent evolutionary innovation in multicellular development. Volvocines do not have a minor spliceosome and independently evolved multicellular species relative to higher plants (2, 42). Recent genome comparisons within this clade found that cell cycle genes are expanded in multicellular species with mutation of retinoblastoma as the primary change needed for unicellular species to become multicellular (41). In addition, genes encoding glycosylated extracellular matrix proteins are expanded in *Volvox carteri*, which has two distinct cell types. Both processes are highly represented in maize U12-type intron-containing genes. However, neither retinoblastoma nor the volvocine-type extracellular matrix proteins have U12-type introns in any species (21, 25).

Importantly, the U12-type introns affected in *rgl3* and *ZRSR2* mutants were maintained since the divergence of plants and animals

(32). The number of U12-type introns in the last eukaryotic common ancestor is not known, but all lineages with a minor spliceosome have a small fraction of genes with U12-type introns. Assuming random loss of introns, the number of homologous genes that still have U12-type introns in human and maize is large, especially with one-half of the homologs having divergent U12-type intron positions. These observations argue for selective pressure to maintain a subset of genes with U12-type introns to carry out essential roles in cell differentiation.

## Materials and Methods

**RNA-Seq.** Normal and *rgl3* kernels in the W22 inbred background were sown in soil (29 °C/20 °C, 12 h/12 h, day/night). Individual seedlings were harvested when about 5 cm high. Normal seedlings were genotyped to select homozygous WT individuals. Total mRNA was extracted from three biological replicates of *rgl3* and WT roots and shoots using TRIzol Reagent (Life Technologies) with RNase-free glycogen (Fermentas) as an RNA carrier. Nonstrand-specific TruSeq (Illumina) cDNA libraries were built from 2 µg of RNA input with a 200-bp median insert length. Twelve libraries were pooled using the KAPA Library Quantification Kit (Kapa Biosystems) and sequenced on the HiSeq 2000 platform with 100-bp paired-end reads.

**Read Mapping.** Bases with a quality score <20 were trimmed with the FASTQ/A utility in the FASTX toolkit ([hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Low-quality reads were discarded if >20% of bases had a quality score <20. Reads from repetitive elements in the MIPS database (43) were removed using Bowtie (44), version 0.12.7. Single-nucleotide polymorphisms (SNPs) between W22 and B73 RefGen\_v2 genome were identified by mapping WT libraries to the B73 RefGen\_v2 genome (45) (Dataset S6). Mosaik 2, version 2.1.33, was run with a hash size of 14, up to 100 hash positions per seed, and allowing 2% mismatch (46). Duplicate reads were removed using Picard, version 1.54 ([broadinstitute.github.io/picard/](http://broadinstitute.github.io/picard/)). W22 SNPs were identified with FreeBayes, version 0.9.5, assuming 2% pairwise differences with B73 and requiring a 99% confidence statistic that an individual variant exists (47). FreeBayes parameters included at least 10 mapped reads with >40% of reads supporting the variant and at least one read having a minimum quality score of 20.

GSNAP, version 2012-5-24, with an IIT file containing W22/B73 SNPs, was used to uniquely align all RNA-seq libraries with B73 RefGen\_v2 (48). Alignment parameters allowed 2% mismatch and mate pairs to deviate 800 bp from the expected paired-end length of 200 bp. GSNAP was guided by the ZmB73\_5b filtered gene set annotations but was permitted to discover novel splicing events. Transcript isoforms were assembled using Cufflinks, version 1.3.0 (20). Each library was assembled independently with a minimum intron size of 10 bp and an overlap radius of 10 bp. Isoform expression analysis was restricted to the ZmB73\_5b filtered gene set. Differential splicing was tested for roots and shoots independently using Cuffdiff with the bias detection and correction algorithm and upper-quartile normalization.

**Tests for Altered Splicing of Individual Introns.** Nonredundant introns were identified from the ZmB73\_5b annotation. HTSeq, version 0.6.1p1 (49), tallied intron read counts, which were normalized by the fragments per kilobase of transcript per million mapped reads (FPKM) of the gene model in each library. Differences in intron reads were determined by a two-tailed *t* test with the Welch modification for degrees of freedom with no correction for multiple testing. Genes that were not expressed in every RNA-seq library were excluded from the analysis. Genomic coordinates for U12-type introns were identified by BLASTN searches against the B73 RefGen\_v2 genome using sequences from the ERISdb database (21, 45).

PSO was calculated from exon-exon junction reads and reads mapping to shortest nonredundant intron for each unique 5'- and 3'-splice site (50). Read counts were calculated with the intersectBed command from Bedtools, version 2.24.0 (51). Fisher's exact test was calculated for each intron with the summed reads across all WT and all mutant libraries and a false-discovery rate of 0.05 using the Benjamini-Hochberg method.

**RT-PCR.** RNA from 12 d after pollination (DAP) whole kernels was isolated by freezing *rgl3* and normal kernels in liquid nitrogen with extraction using TRIzol Reagent followed by RNeasy RNA cleanup with on-column DNase I digestion (Qiagen). Embryo, starchy endosperm, and basal endosperm transfer cell-enriched tissue was hand dissected from 12 DAP whole kernels at -18 °C (52). RNA from these tissues was extracted with the Arcturus PicoPure RNA isolation kit (Applied Biosystems) with on-column DNase I digestion. Endosperm cultures were established as described (19). RNA was extracted from endosperm culture, seedling roots, and seedling shoots using the RNeasy Plant



Mini Kit (Qiagen) with on-column DNase I digestion. Total RNA was reverse-transcribed using the SuperScript III First-Strand Synthesis System (Invitrogen) with oligo(dT)<sub>20</sub> as primer.

Gene-specific primers were designed to distinguish multiple splice variants from individual genes using one primer pair. For genes predicted by Cufflinks to have splicing differences in *rgb3*, annotated alternative transcripts along with raw read coverage depth, visualized by the Integrated Genome Browser (53), were used to target specific regions for RT-PCR. Amplification conditions consisted of the following: 1-min denaturation at 95 °C; 30–31 cycles of 95 °C for 1 min, 55 °C for 1 min, 72 °C for 30 s, and a final extension step of 72 °C for 5 min. RT-PCR products were cloned into the pCR4-TOPO vector for Sanger sequencing.

**Conserved Protein Domain Enrichment Analysis.** All predicted protein isoforms in the maize filtered gene set were queried with HMMER 3.0 against the Pfam database using an inclusion threshold of 0.01 (54, 55). Hypergeometric *P* values were calculated for three samples and populations (Dataset S3): (i) *rgb3* misspliced genes relative to all expressed U12-type intron-containing genes, (ii) *rgb3* misspliced genes relative to all expressed maize genes, and (iii) all U12-type intron-containing genes relative to all genes in the filtered gene set. A gene was considered expressed if at least one intron was tested for normalized intron read depth differences between *rgb3* and WT samples. Pfam domains were manually categorized into biological functions based on the annotation of each domain.

**Human Homolog Analysis.** The B73 RefGen\_v3 maize protein isoforms for the 408 genes in ERISdb with U12-type introns were downloaded from MaizeGDB (21, 56). Each protein isoform was queried against the human RefSeq protein database using BLASTP. Human homologs were defined as the best match with a minimum bit score of 80 (Dataset S4). National Center for Biotechnology Information (NCBI) Gene database unique identifiers were retrieved for each human RefSeq accession along with the corresponding HUGO Gene Nomenclature Committee (HGNC) human gene symbols (57). Human gene symbols containing U12-type introns were identified using the “intron FASTA” query at U12DB to download the complete set of U12-type introns (25). ENSEMBL gene identifiers were parsed from the FASTA file, and identifiers that were still current were converted to NCBI Gene database unique identifiers and human gene symbols using DAVID (58, 59). Predicted protein sequences from retired ENSEMBL gene identifiers were used to query the human RefSeq protein database using BLASTP. Intron sequences that did not have an ENSEMBL identifier were queried against the NCBI human genomic plus transcript database using BLASTN. The complete list of U12DB intron identifiers with current NCBI Gene database unique identifiers and human gene symbols is given in Dataset S5. The human gene symbols from maize–human homologs were used to cross-reference the human gene symbols identified from U12DB to determine maize and human genes that both contained a U12-type intron (Table S3).

To determine U12-type intron positions relative to maize protein sequences, the 5'-exon at ERISdb was queried with BLASTX against the maize protein isoforms from the B73 RefGen\_v3 annotation. Human protein isoforms were downloaded from the Consensus CDS protein set at the NCBI CCDS database (60). The 5'-exon sequence given at U12DB was used for pattern match searches of the consensus transcripts. The relative amino acid position of the U12-type intron in the maize and human homologs were determined using Clustal Omega (61). All maize protein isoforms related to each human protein were aligned along with protein isoforms that were truncated at the last complete codon in the exon 5' of the U12-type intron. The U12-type intron position was scored as identical if the C-terminal residues of the maize and human truncated isoforms were at the same position

in the alignment. Normalized U12-type intron positions were determined based on the amino acid residues for each protein isoform and then averaged by gene. When a protein had more than one U12-type intron, normalized positions were calculated independently for each intron. U12-type intron retention models were constructed for E2F gene family members using the first annotated transcript and protein isoform for each gene.

**Isolation of Polysomes.** Normal and *rgb3* seedlings were collected at 3- to 5-cm stage. For each biological replicate, seedlings from two self-pollinated families were combined, flash frozen, and pulverized in liquid nitrogen. Total, nuclei, and polysomal RNA fractions were extracted from 2.5–3 mL of ground tissue using the conventional isolation of plant polysomes protocol as described (62). Initial pellets to clarify the tissue extract were saved as the nuclear fraction, and 500  $\mu$ L of the clarified supernatant was saved as the total RNA fraction. RNA from all fractions was extracted with Qiagen RNeasy kits and treated with Ambion TURBO DNA-free kit (Thermo Fisher Scientific) following the manufacturer's instructions. RT-PCR was completed from 1  $\mu$ g of RNA using M-MLV reverse transcriptase (Promega) to synthesize cDNA as described (63). PCR cycles were optimized for each gene.

**Subcellular Localization.** The RGH3-GFP and U2AF2 fluorescent protein fusion constructs were previously described (19). *Rgb3 $\Delta$ UHM* was cloned from *Rgb3x* by amplifying an N-terminal fragment from the start codon to the end of the first zinc finger using primers Rgb3 $\Delta$ UHM-1 and Rgb3 $\Delta$ UHM-2 (Table S4). A C-terminal fragment was amplified from the second zinc finger to the C-terminal stop codon using primers Rgb3 $\Delta$ UHM-3 and Rgb3 $\Delta$ UHM-4. The fragments were ligated by overlap extension PCR (64). U2AF2 and U2AF1 were amplified from leaf cDNA of normal 14-d-old seedlings. The RT-PCR products were subcloned into pENTR vector (Invitrogen) and then cloned into pB7-WGF2, pB7-RWG2, or pB7-FWG2 (65, 66) as described (19). BiFC constructs were created by transferring coding sequences from pENTR vector clones to pSAT4-DEST or pSAT5-DEST vectors (67, 68) by LR clonease reactions following manufacturer's instructions (Invitrogen). The recombinant transcription cassettes were digested with I-CeuI or I-SceI and ligated into the pPZP-RCS2 binary vector (67). Transient expression in *Nicotiana benthamiana* and microscopy analyses were completed as described with some modifications (19). Binary vectors were transformed into *Agrobacterium tumefaciens* strain GV3101. Protein expression was visualized 24–48 h after transient transformation. YFP was excited at 514 nm and detected with an emission band of 525–565 nm.

**Protein Analyses.** Rabbit polyclonal N-terminal RGH3 antibodies was raised against the synthetic peptide: SAQEVLDKVAQETPNFGTE (Bio Synthesis). Total protein from seed tissue was extracted as described (69) except that the fresh tissue extracts were cleared with a single filtration step. For the in vitro transcription/translation, ORFs from described *Rgb3* variants (19) were amplified with Phusion polymerase (Finnzymes) using primers containing an N-terminal 6 $\times$ His tag and restriction sites for *Sgf1* (Promega) and *Pme1* (New England Bio Labs). PCR products were digested and ligated into pF3AWG vectors (Promega) and expressed using the TnT SP6 High-Yield Wheat Germ Protein Expression System (Promega) following the manufacturer's instructions.

**ACKNOWLEDGMENTS.** We thank Yuqing Xiong and Byung-Ho Kang for technical assistance in dissecting endosperm tissues. This work was supported by National Science Foundation (NSF) Graduate Research Fellowship DGE-0802270 (to C.M.G.), NSF Awards MCB-1412218 and IOS-1031416, the University of Florida Genetics Institute Center for RNA Research, and the Vasil–Monsanto Endowment.

- Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ (2013) Dynamic integration of splicing within gene regulatory pathways. *Cell* 152(6):1252–1269.
- Dávila López M, Rosenblad MA, Samuelsson T (2008) Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. *Nucleic Acids Res* 36(9):3001–3010.
- Turunen JJ, Niemelä EH, Verma B, Frilander MJ (2013) The significant other: Splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA* 4(1):61–76.
- Younis I, et al. (2013) Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *eLife* 2:e00780.
- Otake LR, Scamborova P, Hashimoto C, Steitz JA (2002) The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in *Drosophila*. *Mol Cell* 9(2):439–446.
- Ederly P, et al. (2011) Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science* 332(6026):240–243.
- He H, et al. (2011) Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* 332(6026):238–240.
- Markmiller S, et al. (2014) Minor class splicing shapes the zebrafish transcriptome during development. *Proc Natl Acad Sci USA* 111(8):3062–3067.
- Kim WY, et al. (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(12):3951–3962.
- Jung HJ, Kang H (2014) The *Arabidopsis* U11/U12-65K is an indispensable component of minor spliceosome and plays a crucial role in U12 intron splicing and plant development. *Plant J* 78(5):799–810.
- Xu T, Kim BM, Kwak KJ, Jung HJ, Kang H (2016) The *Arabidopsis* homolog of human minor spliceosomal protein U11-48K plays a crucial role in U12 intron splicing and plant development. *J Exp Bot* 67(11):3397–3406.
- Shen H, Zheng X, Luecke S, Green MR (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(21):2389–2394.
- Tronçère H, Wang J, Fu XD (1997) A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. *Nature* 388(6640):397–400.

14. Yoshida K, et al. (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478(7367):64–69.
15. Damm F, et al.; Groupe Francophone des Myélodysplasies (2012) Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood* 119(14):3211–3218.
16. Thol F, et al. (2012) Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* 119(15):3578–3584.
17. Cazzola M, Della Porta MG, Malcovati L (2013) The genetic basis of myelodysplasia and its clinical relevance. *Blood* 122(25):4021–4034.
18. Madan V, et al. (2015) Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat Commun* 6:6042.
19. Fouquet R, et al. (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(12):4280–4297.
20. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–515.
21. Szcześniak MW, Kabza M, Pokrzywa R, Gudyś A, Makalowska I (2013) ERISdb: A database of plant splice sites and splicing signals. *Plant Cell Physiol* 54(2):e10.
22. Burge CB, Padgett RA, Sharp PA (1998) Evolutionary fates and origins of U12-type introns. *Mol Cell* 2(6):773–785.
23. Chang WC, Chen YC, Lee KM, Tarn WY (2007) Alternative splicing and bioinformatic analysis of human U12-type introns. *Nucleic Acids Res* 35(6):1833–1841.
24. Chen J, et al. (2014) Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol* 166(1):252–264.
25. Alioto TS (2007) U12DB: A database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res* 35(Database issue):D110–D115.
26. Lahmy S, et al. (2007) QQT proteins colocalize with microtubules and are essential for early embryo development in *Arabidopsis*. *Plant J* 50(4):615–626.
27. Soni S, Bala S, Kumar A, Hanspal M (2007) Changing pattern of the subcellular distribution of erythroblast macrophage protein (Emp) during macrophage differentiation. *Blood Cells Mol Dis* 38(1):25–31.
28. Symoens S, et al. (2015) Genetic defects in TAPT1 disrupt ciliogenesis and cause a complex lethal osteochondrodysplasia. *Am J Hum Genet* 97(4):521–534.
29. Li HJ, et al. (2011) POD1 regulates pollen tube guidance in response to micropylar female signaling and acts in early embryo patterning in *Arabidopsis*. *Plant Cell* 23(9):3288–3302.
30. Porch TG, Tseung CW, Schmelz EA, Settles AM (2006) The maize *Viviparous10/Viviparous13* locus encodes the *Cnx1* gene required for molybdenum cofactor biosynthesis. *Plant J* 45(2):250–263.
31. Suzuki M, et al. (2006) The maize *viviparous15* locus encodes the molybdopterin synthase small subunit. *Plant J* 45(2):264–274.
32. Lin CF, Mount SM, Jarmolowski A, Makalowski W (2010) Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol Biol* 10:47.
33. Magyar Z, et al. (2012) *Arabidopsis* E2FA stimulates proliferation and endocycle separately through RBR-bound and RBR-free complexes. *EMBO J* 31(6):1480–1493.
34. Kielkopf CL, Rodionova NA, Green MR, Burley SK (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* 106(5):595–605.
35. Hu CD, Chinenov Y, Kerppola TK (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* 9(4):789–798.
36. Patel AA, McCarthy M, Steitz JA (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J* 21(14):3804–3815.
37. Pessa HK, et al. (2010) Gene expression profiling of U12-type spliceosome mutant *Drosophila* reveals widespread changes in metabolic pathways. *PLoS One* 5(10):e13215.
38. Lotti F, et al. (2012) An SMN-dependent U12 splicing event essential for motor circuit function. *Cell* 151(2):440–454.
39. Doktor TK, et al. (2016) RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns. *Nucleic Acids Res* 45(1):395–416.
40. Russell AG, Charette JM, Spencer DF, Gray MW (2006) An early evolutionary origin for the minor spliceosome. *Nature* 443(7113):863–866.
41. Clarke M, et al. (2013) Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol* 14(2):R11.
42. Hanschen ER, et al. (2016) The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat Commun* 7:11370.
43. Nussbaumer T, et al. (2013) MIPS PlantsDB: A database framework for comparative plant genome research. *Nucleic Acids Res* 41(Database issue):D1144–D1151.
44. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
45. Monaco MK, et al. (2014) Gramene 2013: Comparative plant genomics resources. *Nucleic Acids Res* 42(Database issue):D1193–D1199.
46. Lee WP, et al. (2014) MOSAIK: A hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS One* 9(3):e90581.
47. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907.
48. Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26(7):873–881.
49. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.
50. Katz Y, Wang ET, Airolidi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* 7(12):1009–1015.
51. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
52. Xiong Y, Li Q, Kang B, Chourey P (2011) Discovery of genes expressed in basal endosperm transfer cells in maize using 454 transcriptome sequencing. *Plant Mol Biol Report* 29(4):835–847.
53. Nicol JW, Helt GA, Blanchard SG, Jr, Raja A, Loraine AE (2009) The Integrated Genome Browser: Free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25(20):2730–2731.
54. Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform* 23(1):205–211.
55. Finn RD, et al. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res* 44(D1):D279–D285.
56. Andorf N, et al. (2016) MaizeGDB update: New tools, data and interface for the maize model organism database. *Nucleic Acids Res* 44(D1):D1195–D1201.
57. Gray KA, Seal RL, Tweedie S, Wright MW, Bruford EA (2016) A review of the new HGNC gene family resource. *Hum Genomics* 10:6.
58. Huang D-W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
59. Huang D-W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1–13.
60. Farrell CM, et al. (2014) Current status and new features of the Consensus Coding Sequence database. *Nucleic Acids Res* 42(Database issue):D865–D872.
61. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
62. Mustroph A, Juntawong P, Bailey-Serres J (2009) Isolation of plant polysomal mRNA by differential centrifugation and ribosome immunopurification methods. *Methods Mol Biol* 553:109–126.
63. Bai F, Reinheimer R, Durantini D, Kellogg EA, Schmidt RJ (2012) TCP transcription factor, BRANCH ANGLE DEFECTIVE 1 (BAD1), is required for normal tassel branch angle formation in maize. *Proc Natl Acad Sci USA* 109(30):12225–12230.
64. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: Gene splicing by overlap extension. *Gene* 77(1):61–68.
65. Karimi M, Inzé D, Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci* 7(5):193–195.
66. Karimi M, Depicker A, Hilson P (2007) Recombinational cloning with plant gateway vectors. *Plant Physiol* 145(4):1144–1154.
67. Tzfira T, et al. (2005) pSAT vectors: A modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. *Plant Mol Biol* 57(4):503–516.
68. Citovsky V, et al. (2006) Subcellular localization of interacting proteins by bimolecular fluorescence complementation in plants. *J Mol Biol* 362(5):1120–1131.
69. Abdalla KO, Thomson JA, Rafudeen MS (2009) Protocols for nuclei isolation and nuclear protein extraction from the resurrection plant *Xerophyta viscosa* for proteomic studies. *Anal Biochem* 384(2):365–367.