Evolutionary Conservation and Essential Function of Human and Maize RNA Binding Motif Protein 48 (RBM48) in U12-Type Intron Splicing

Amy E. Siebert[1,2,*], Jacob Corll[1,3,*], J. Paige Gronevelt[1,4], Laurel Levine[1], Linzi M. Hobbs[1], Catalina Kenney[1], Ruth Davenport[3], A. Mark Settles[5,6], W. Brad Barbazuk[3,6,7], Randal J. Westrick[1,8,9], Gerard J. Madlambayan[1] and Shailesh Lal[1]

1 Departments of Biological Sciences and Bioengineering, Oakland University, Rochester Hills, MI, 48309, USA
2 Versiti Blood Research Institute, Milwaukee, WI, 53226, USA
3 Department of Biology, University of Florida, Gainesville, FL, 32611, USA
4 Department of Internal Medicine Hematology and Oncology, University of Michigan, Ann Arbor, MI, 48109, USA
5 Department of Horticultural Sciences, University of Florida, Gainesville, FL, 32611, USA
6 The Genetics Institute, University of Florida, Gainesville, FL, 32611, USA
7 The Interdisciplinary Center for Biotechnology Research (ICBR) University of Florida, Gainesville, FL, 32611, USA
8 Center for Data Science and Big Data Analytics, Oakland University, Rochester Hills, MI, 48309, USA
9 Life Sciences Institute, University of Michigan, Ann Arbor, MI, 48109, USA

*Authors contributed equally

Corresponding Author
Shailesh Lal
Email: Lal@oakland.edu

Abstract

U12-type (minor) introns are found in most multicellular eukaryotes and constitute ~0.5% of all introns in species with a minor spliceosome required for their splicing. However, the biological relevance of U12-type introns is not well understood. It is known that mutations resulting in aberrant U12-type intron splicing cause developmental defects in both plants and animals. We recently reported that maize RNA Binding Motif Protein 48 (RBM48) is an essential splicing factor for U12-type introns. Maize *rbm48* mutants display aberrant genome-wide U12-type intron splicing. This leads to severe defects in endosperm development, resulting in non-viable seeds. In this report, we use CRISPR/Cas9-mediated ablation of *RBM48* in human K-562 cells to establish the evolutionary conservation of *RBM48* dependent U12-type intron splicing between maize and humans. Comparative RNA-seq analysis performed on RBM48 deficient human cell lines and maize endosperm defined a subset of orthologous minor U12-type containing genes (MIGs) displaying aberrant splicing of U12-type introns in both species. Mutations in the majority of these MIGs have been reported to cause developmental defects in both plants and animals. Thus, a comparison of RNA-seq data between distantly related species containing mutations in RBM48 identifies candidate genes likely to mediate mutant phenotypes of U12-type splicing defects. Our results elucidate deeply conserved post-transcriptional processing mechanisms that are required for normal growth and development of eukaryotes with a minor spliceosome.

Introduction

Accurate splice site recognition and precise removal of introns during pre-mRNA processing is fundamental for gene expression in eukaryotes (Simpson, Filipowicz 1996; Lorkovic et al. 2000; Ru, Wang, Brendel 2008). The vast majority of introns are spliced by the major spliceosome and are categorized as U2-type (or major) introns (Staley, Guthrie 1998; Lee, Rio 2015). A second group of introns are spliced by the minor spliceosome and are categorized as U12-type (or minor) introns (Turunen et al. 2013). U12-type introns constitute ~0.5% of all introns in species with a minor spliceosome (Turunen et al. 2013). In contrast to U2-type, U12-type introns lack 3' polypyrimidine tracts, have differences in their 5' and 3' terminal dinucleotides, and also have highly conserved 5' splice site and branch point sequences (Hall, Padgett 1994; Will, Luhrmann 2005). These U12-type introns exist side-by-side with U2-type introns within Minor Intron-containing Genes (MIGs) (Turunen et al. 2013). Most of the 758 MIGs identified in humans contain only a single U12-type intron, however, MIGs harboring multiple U12-type introns also exist (Alioto 2007).

Defects in pre-mRNA splicing by cis- or trans-acting mutations have been linked to approximately 60% of human diseases (Lim et al. 2011; Sterne-Weiler, Sanford 2014; Chabot, Shkreta 2016). While many of the identified mutations impact splicing of U2-type introns or major spliceosome components (Cooper, Wan, Dreyfuss 2009), mutations linked to anomalous splicing of U12-type introns have recently been demonstrated to cause developmental defects in both plants and animals. (Kim et al. 2010; Edery et al. 2011; He et al. 2011; Jung, Kang 2014; Markmiller et al. 2014; Xu et al. 2016). Mutations in factors specific to the minor spliceosome have also been shown to impact the splicing of U12-type introns in a subset of target MIGs responsible for development in both humans and maize (Fouquet et al. 2011; Madan et al. 2015; Gault et al. 2017). For example, in human subtypes of myelodysplastic syndrome (MDS) that harbor mutations in the spliceosomal gene *ZRSR2*, the aberrant differentiation and proliferation of myeloid precursors that defines this hematological disorder is associated with impaired splicing and retention of U12-type introns (Madan et al. 2015). In maize, the corresponding *ZRSR2* ortholog is disrupted by the hypomorphic *rough endosperm3* (*rgh3*) allele which results in aberrant endosperm differentiation and proliferation that is also associated with impaired splicing and retention of U12-type introns (Gault et al. 2017). A comparison of MIGs affected in the human *ZRSR2* and maize *rgh3* mutants identifies a deeply conserved role for cell differentiation pathways involved in promoting stem cells to terminal fates (Gault et al. 2017).

The major and minor spliceosomal protein complexes that catalyze the splicing of U2-type and U12-type introns respectively, are complex molecular machines displaying remarkable

similarity (Turunen et al. 2013). While most of the core and auxiliary protein components are postulated to be shared between the two complexes, to date, only eight have been identified to be specific to the minor spliceosome (Turunen et al. 2013). We recently reported a conserved RNA Binding Motif Protein (RBM48) in eukaryotes as a minor spliceosomal factor required for the efficient splicing of U12-type introns in maize (Bai et al. 2019). The maize *rbm48* mutants share a stark phenotypic similarity with *rgh3* mutants in their developmental defects of endosperm, as they also exhibit increased cell proliferation and decreased cell differentiation (Bai, et al. 2019). Furthermore, both *in vitro* protein-protein interaction and *in vivo* co-localization studies indicate RBM48 interacts with core U2 Auxiliary Factor (U2AF) and RGH3, suggesting the components of both major and minor spliceosomal complexes interact during pre-mRNA processing (Bai et al. 2019). Through comparative transcriptomics, it was found that more than 75% of mis-spliced MIGs overlapped between the *rbm48* and *rgh3* mutants (Gault et al. 2017; Bai et al. 2019). This large overlap confounded our efforts to identify genes contributing to the mutant maize phenotype.

Comparative transcriptomic analysis of orthologous U12 splicing factor mutants between distantly related species, such as the comparison between the U12-type splicing factor mutants *ZRSR2* and *rgh3* (Gault et al. 2017) could provide a novel strategy to not only identify candidate genes impacting mutant phenotypes, but also to elucidate conserved gene regulation and cellular processes required for normal growth and development. Herein, we generated a CRISPR/Cas9 mediated functional knockout of the orthologous *RBM48* gene in the human K-562 chronic myeloid leukemia cell line. Using comprehensive transcriptome profiling, we demonstrate that RBM48 is required for efficient splicing of U12-type introns indicating evolutionary conservation of function between plants and humans. Comparative analyses identified a subset of conserved MIGs that are affected in both maize and human RBM48 knockout mutants suggesting evolutionary selection to maintain common post-transcriptional RNA processing mechanisms since the divergence of plants and animals.

Results

Targeting human *RBM48* for Cas9-mediated cleavage in K-562 cells

The CRISPR/Cas9 genome editing system was implemented to generate a functional knockout of *RBM48* in K-562 cells. Two single-guide RNAs (sgRNA) were designed to target the proximal (sgRNA#1) and distal (sgRNA#2) regions of the RNA Recognition Motif (RRM) domain within the *RBM48* genomic locus. The position, sequence, and approximate cleavage site of each sgRNA is displayed schematically in Figure 1A. To determine the efficacy of Cas9-mediated DNA cleavage at each targeting site, genomic DNA (gDNA) from cell populations at harvest following puromycin selection (day 0), and 10 and 16 days post-selection was analyzed by Surveyor nuclease assay (Fig. 1B). Heteroduplex formation due to insertion-deletion (indel) polymorphisms generated within the *RBM48* genomic locus was confirmed on day 0 in sgRNA#1- and sgRNA#2-targeted cell populations by detection of 309 and 272 base pair (bp) and 332 and 266 bp cleavage fragments, respectively. No detectable fragments were observed in digested vector control (VC) amplicons. Indels were detectable in each population at 10 days post-selection. By day 16, digested fragments were undetectable from sgRNA#1-targeted DNA. By contrast, cleavage fragments from DNA targeted by sgRNA#2 were stable and appeared to intensify over this 16 day duration.

Growth of the transfected cell populations was monitored over the 16 day post-selection time interval (Fig. 1C and D). Over the first 8 days of culture, cell growth was reduced in all 3 cohorts as a result of the transfection procedure (Fig. 1C). While the VC and sgRNA#2 cells began to recover by 10 days post-selection, a lag in proliferation was still evident in the sgRNA#1-targeted K-562 cell population as compared to VC cells (q=0.002). By day 16 however, the sgRNA#1 cells regained vigor as the number of viable cells from both Cas9-targeted and VC populations were comparable (Fig. 1D).

The indels generated within the two *RBM48*-targeted cell populations were further investigated by Sanger sequencing and Tracking of Indels by DEcomposition (TIDE) analysis on day 16 post-selection cells. Sequencing chromatograms flanking the protospacer adjacent motif (PAM) site of sgRNA#1 (Fig. 1E; upper panel) and subsequent TIDE analysis (Fig. 1F; left), showed that the sgRNA#1-targeted cells were a nearly homogenous population predominantly harboring wild-type *RBM48* by day 16 post-selection. By contrast, multiple sequence trace peaks immediately following the cleavage site of Cas9 were observed with sgRNA#2 (Fig. 1E; lower panel). The frequency of wild-type *RBM48* sequences within the heterogeneous sgRNA#2 cell population by TIDE analysis was only 19.8%, while the total Cas9 cleavage efficiency was approximately 72% with nearly 30% comprising a 1 bp insertion (Fig. 1F; right). From these

observations, we conclude that sgRNA#1 was inefficient in generating stable targeted-indels which led to the rise of unmodified *RBM48* wild-type cells as the dominant proliferative population. However, the sgRNA#2 targeting site was highly efficient at generating stable indels without apparent growth defects, enabling us to readily propagate this cell population. Therefore, the VC and sgRNA#2-targeted cell populations were established and maintained as independent K-562 cell sublines, hereafter referred as VC and *RBM48* KO, respectively.

Loss of full length RBM48 in the *RBM48* KO subline

Loss of *rbm48* in maize impairs splicing of U12-type introns (Bai et al. 2019). To elucidate whether human RBM48 has a conserved role in U12-dependent splicing, we performed mRNA-seq analysis on six unique passages of both *RBM48* KO and corresponding VC populations. Prior to analyzing the mRNA-seq data for splicing defects, *RBM48* expression was compared between the heterogeneous KO and VC populations to determine the impact of CRISPR/Cas9 mutagenesis. In agreement with the sgRNA#2 TIDE analysis in Figure 1F, a disparity in summed read depth was clearly visible within the targeting region of *RBM48* exon 3 in the KO population in comparison to VC (Fig. 2A). However, RT-qPCR assays for expression of all three *RBM48* transcripts (Fig. 2B) showed no significant differences between the KO and VC populations over six passages indicating that Cas9-targeting did not affect overall transcription of the *RBM48* gene. Through further analysis, 12 indel-harboring transcript sequences were identified that were maintained within all six passages of the *RBM48* KO subline. Figure 2C displays the indel positions and their predicted translational consequences in relation to NM_032120.4 and NP_115496.2, respectively. Within the heterogeneous KO subline, we found evidence for two mutant transcripts resulting in loss of a single amino acid and ten mutant transcripts resulting in frameshifts and truncation of the C-terminal region of the RBM48 protein product in all three coding variants.

To test whether sgRNA#2 targeting impacts wild-type RBM48 protein levels, we generated a third K-562 cell subline that carried a Cas9-mediated C-terminal Myc epitope tag as a tool to monitor protein truncation or loss of expression by sgRNA#2 targeting as described in Figure S1. The *RBM48* C-terminal region was selected to avoid a potential synthetic interaction due to overlap between the regulatory regions of *RBM48* and the adjacent *PEX1* gene. CRISPR/Cas9 targeting in the *RBM48-Myc* subline using sgRNA#2 (KO) resulted in indel generation similar to, but with lower efficiency than that observed in the original *RBM48* KO subline (Fig. 3A and B). Western blot analyses of three independent passages of the *RBM48-Myc* VC and KO cells revealed a significant reduction in RBM48-Myc levels in the sgRNA#2-targeted

KO cell populations compared to VC (Fig. 3C; p=0.0051). We concluded that *RBM48* KO cells had drastically reduced expression of full length RBM48 protein and that this subline could be used to determine the effects of the *rbm48* mutation on transcript processing.

Disruption of U12-dependent splicing in *RBM48* KO cells

To assess whether the *RBM48* RNA processing role is conserved in maize and humans, the transcriptome profiling of the *RBM48* KO and VC populations were analyzed for intronic splicing defects. Intronic and flanking exon-exon junction read counts for each individual U2- and U12-type intron expressed in the K-562 cell sublines were assessed for significance using Student's t-test (Dataset S1). A minimum of 10 total exon-exon junction reads in both VC and KO populations were required to test an intron. Of the 607 introns tested, 313 U12-type introns (52%) were significantly retained in the *RBM48* KO population in comparison to VC (q<0.05). Gene Ontology (GO) term enrichment analysis identified cellular localization as the most significantly enriched term ($p\leq$ 0.05) among MIGs with U12-type intron retention (Table S1). Four significantly affected U12-type introns in the *DIAPH1*, *MAPK1*, *MAPK3*, and *TXNRD2* genes were selected for validation by RT-qPCR. Based on our mRNA-seq data, each of these genes exhibited significantly increased read depth (p<0.05) in the U12-type intronic regions in *RBM48* KO populations compared to VC (Fig. 4). We designed primers to detect both U12-type intron retention and total transcript levels by RT-qPCR for each of the four selected MIGs (Fig. 4). No significant differences were found in the total transcript expression levels between the *RBM48* KO and VC populations. By contrast, expression levels of all four U12-type introns were significantly increased, suggesting intron retention in these transcripts. These findings indicate that in humans, RBM48 plays a role in post-transcriptional RNA processing and U12-dependent splicing, but that U12-type intron retention does not necessarily reduce the transcript levels of all genes where the U12-type introns are retained.

We previously reported that maize *rbm48* mutants display retention in a greater proportion of U12-type introns than U2-type introns (Bai et al. 2019). The extent of U12-type splicing defects in *RBM48* KO appears to be similar to defects observed in maize *rbm48* mutants. The distributions of log2 fold-change (log2fc) of all U2-type and U12-type introns in human *RBM48* KO and maize *rbm48* mutants are plotted in Figure 5A. In both species, U12-type introns are shifted to greater overall log2fc values than U2-type introns suggesting that loss of RBM48 in humans and maize specifically increases retention or mis-splicing of U12-type intron sequences.

Conservation of mis-spliced genes shared between U12 splicing mutants of human and maize

The aberrant splicing of multiple U12-type introns poses a challenge to unambiguously identify genes causing defective phenotypes. In our effort, we performed comparative analysis and searched for MIGs that display orthologous defects of U12-type intron splicing between the RBM48 mutants of human and maize. Gault *et al.* (2017) found 36 human MIGs that share homology with 57 maize MIGs. These candidate orthologs have conserved domain structure and high sequence similarity of greater than 80 bit score using reciprocal blastp. The imbalance of human to maize MIGs is a result of differential gene duplication and divergent retention of duplicated genes since the split of animal and plant lineages (Wei et al. 2007). Therefore, several single copy human MIGs share homology with two or more gene copies in maize. Of these 36 human MIGs, 58% are significantly mis-spliced in *RBM48* KO cells. As shown in Figure 5B, we found that 18 of these MIGs share mis-splicing with 22 homologous maize MIGs when RBM48 is knocked out or mutated, respectively (Table 1). To validate these RNA-seq comparisons, we selected four genes, *SMYD2*, *TAPT1*, *WDR91* and *ZPR1*, that showed significant U12-type intron retention in both human *RBM48* KO and maize *rbm48* mutants (Fig. 6). As shown in Figure 6A, the *RBM48* KO cells exhibited increased RNA-seq read depth in U12-type intronic regions compared to VC. When analyzed by RT-PCR, each of these genes also display stronger band intensity of the U12-type intron retained transcript in *RBM48* KO mutants (Fig. 6B). Moreover, corresponding RT-PCR assays of maize homologs show increased U12-type intron retention products in maize *rbm48* mutants relative to wild-type (Fig. 6B).

A similar example of U12-type intron retention in conserved MIGs has been observed between human *ZRSR2* and maize *rgh3* mutant samples (Madan et al. 2015; Gault et al. 2017). Previously, we reported strong overlap of U12-type intron targets between maize *rbm48* and *rgh3* mutants as well (Bai et al. 2019). Interestingly, of the 36 human and 57 maize homologous MIGs, 9 human MIGs are significantly mis-spliced in our *RBM48* KO and the *ZRSR2* mutants (Madan et al. 2015) corresponding to 11 significantly mis-spliced homologous MIGs in the maize *rbm48* (Bai et al. 2019) and *rgh3* (Gault et al. 2017) mutants (Table 1).

Discussion

Components of the minor spliceosome play important roles in promoting normal multicellular processes in both plants and animals. However, the mechanisms by which they impact cell development and differentiation are not well understood. Mutations in minor spliceosomal factors cause pleiotropic molecular defects affecting a large proportion of MIGs (Jafarifar et al. 2014; Jung, Kang 2014; Markmiller et al. 2014; Gault et al. 2017; Kwak et al. 2017; Verma et al. 2018). This has made it difficult to delineate the causative genes underlying the pleiotropic phenotypes of MIG spliceosomal factor mutants, even if they exhibit highly similar characteristics.

Based on the role of maize RBM48 in U12-type intron splicing, we hypothesized human RBM48 would also impact U12-type intron splicing. We investigated RBM48 function by generating a CRISPR/Cas9 knockout in the human K-562 cell line and found that this role is conserved between humans and maize. Approximately 52% of U12-type introns analyzed in the *RBM48* KO subline were significantly retained, with their corresponding MIGs associated with several different pathways that are essential to growth and differentiation. We further determined that there are 18 orthologous MIGs mis-spliced in both human and maize RBM48 mutants. These observations identified a small group of MIGs as candidates for potential trans-species drivers of mutant phenotypes.

With the exception of *FRA10AC1*, mutations that either alter or abolish gene expression of these 18 conserved MIGs have been documented in the literature. In each case, the MIG mutations result in phenotypes displaying a broad spectrum of developmental defects in either humans or plants, several of which are reported in both species. For example, mutations in human ALG12 lead to Congenital Disorders of Glycosylation type Ig (CDG-Ig). CDG-Ig patients display a wide range of pleiotropic effects, including but not limited to hypotonia, developmental delay, and abnormal blood clotting (Grubenmann et al. 2002; Zdebska et al. 2003; Haeuptle, Hennet 2009). A mutation in the homologous Arabidopsis ALG12 gene disrupts ER-mediated degradation of brassinolide receptors conferring a dwarf mutant phenotype to the plant (Hong et al. 2009). Similarly, mutations in the chromatid cohesin *SMC3* and homologous *TTN* genes lead to pleotropic developmental defects in both humans and plants, respectively (Liu Cm et al. 2002; Vega et al. 2005).

Mutations in U12 splicing factors are reported to inhibit cell differentiation and promote cell proliferation in both plants and animals. For example, somatic mutations of human *ZRSR2* cause suppression of myeloid cell differentiation and increased proliferation of myeloid progenitor cells (Madan et al. 2015). Similarly, maize mutants in both *rgh3* and *rbm48* display

suppression of endosperm differentiation and increased proliferation of endosperm in tissue culture (Gault et al. 2017; Bai et al. 2019). Within several of the conserved MIGs identified herein, mutations have been linked to abnormal cell proliferation or cancer in humans. For example, mutations in BRCC3, which encodes a cell cycle regulator involved in G2/M phase transition, has been linked to myelodysplastic syndrome (MDS), as well as breast and cervical cancers (Boudreau et al. 2007; Huang et al. 2015; Zhang, Zhou 2018). Mutations in E2F3 have been associated with retinoblastoma, bladder and lung cancer (Ziebold et al. 2001; Wang et al. 2017a; Liu et al. 2018).

Despite these observed similarities, cell type- and lineage-specific differences play a large role in phenotypic effects of U12-type intron retention. Differential effects on *in vitro* cell viability are observed with targeted loss of *RBM48* in genome-wide CRISPR/Cas9 dropout screens where it is deemed essential in some human cell lines and dispensable in others (Hart et al. 2015; Wang et al. 2015; Aguirre et al. 2016; Tzelepis et al. 2016; Hart et al. 2017; Wang et al. 2017b). In a recent meta-analysis of 17 CRISPR screens conducted using three different large-scale libraries, *RBM48* is found to be essential for survival in 4 out of 7 leukemia-derived cell lines (Hart et al. 2017). This includes the OCI-AML3, OCI-AML5, HL-60 and HAP1 lines. However, viability is unaffected in MV4-11, KBM-7 and K-562 cells deficient in *RBM48*, which is in agreement with our results in the *RBM48* KO K-562 subline. Interestingly, the essentiality of *ZRSR2* overlapped with *RBM48* in 6 of the 7 hematopoietic cell lines (Hart et al. 2017). These 7 lines each retain a degree of lineage plasticity and can be induced to further differentiate in response to defined stimuli. For example, the adherent HAP1 cell line was isolated from an attempt to generate induced pluripotent stem cells from KBM-7 suspension cells and as a result, no longer display any hematopoietic markers (Carette et al. 2011). These two nearly haploid lines also vary in their requirement of *RBM48* and *ZRSR2* for survival. These cell-type specific differences suggest that phenotypic effects of U12-type intron retention is likely dependent on the lineage-specific differentiation potential of each cell line with disruption of MIG-specific pathways resulting in lethality in some cell lines but not others.

It is well established that developmental gene expression is intricately regulated in a temporal, cellular, and tissue-specific manner beginning in the early stages of embryogenesis. Surprisingly, the inefficient splicing of U12-type introns often perturbs the coding potential but not the expression level of MIGs, many of which are essential for development. We found no changes in total transcript expression levels in our validation assays and similar observations have been made for maize *rgh3* and *rbm48* (Gault et al, 2017; Bai et al., 2019). Consequently, the mutant phenotypes observed for U12-dependent splicing defects are likely a direct result of the

disrupted post-transcriptional processing of MIG transcripts. Further elucidation of the mechanism(s) by which retention of U12-type introns result in mutant phenotypes will not only entail biochemical and genetic analysis of individual MIGs, but will also require comprehensive profiling of their expression in diverse cells and tissues at different stages of development, both at the levels of transcription and translation.

Materials and Methods

Design and Generation of RBM48-Targeting CRISPR/Cas9 Constructs and Donor Templates

Generation of vector and target-specific guide RNA (sgRNA) was designed as described (Westrick et al. 2017). The three *RBM48*-specific sgRNAs (Table S2) were separately cloned into the pX459 expression vector [pSpCas9(BB)- 2A-Puro; Addgene plasmid ID: 48139] (Ran et al. 2013). The single stranded DNA oligonucleotide necessary for generating a carboxy terminal (C-terminal) *RBM48* Myc-epitope tag within the *RBM48* genomic locus was obtained from Integrated DNA Technologies (IDT, Coralville, IL), is listed in Table S2 and described in Figure S1.

Transfection of K-562 Cells

K-562 cells (American Type Culture Collection (ATCC) CCL-243™, Manassas, VA) were cultured in Iscove's Modified Dulbecco's Medium (IMDM) (Hyclone, Logan, UT), supplemented with 10% (v/v) fetal bovine serum (FBS) (Hyclone) and incubated with 5% CO_2 at 37˚C. K-562 cells were seeded in triplicate into 6-well plates at a density of 2.4 x 10^5 cells per well in 3 mL IMDM containing 10% FBS for 24 hours (h) prior to transfection. Cells in each well were transfected with 0.5 μg pX459 construct expressing sgRNA#1 or sgRNA#2, or co-transfected with 0.5 μg pX459 construct with or without sgRNA TAG, and 0.5 μg homology directed repair (HDR) donor template for 24 h using 0.75 μl lipofectamine 3000 (Invitrogen/Thermo Fisher Scientific, Waltham, MA) diluted in Opti-Mem I media (Gibco/Thermo Fisher Scientific) following manufacturer's instruction. Transfected cells were selected for 48 h using 3 μg/ml puromycin dihydrochloride (Gibco/Thermo Fisher Scientific). Subsequently, the single transfectants were resuspended in fresh media and aliquots were isolated for cell viability assays and genotyping. Six-well culture plates were initially seeded for cell viability with approximately 1 x 10^5 cells per well in 3 ml culture media containing 10% FBS. Cell viability was monitored every 48 h for 16 days using trypan blue exclusion assay and counted by hemocytometer. Significant differences were determined by unpaired Student's *t*-test followed by adjustment for multiple comparisons using the Benjamini and Hochberg False Discovery Rate. The vector control (VC) and sgRNA#2-targeted *RBM48* knockout (KO) K-562 cell sublines were subcultured and six subsequent passages were used for further analyses. For co-transfectants, single-cell colonies isolated by serial dilution were expanded for two weeks prior to genotypic analysis. A single homogenous *RBM48-Myc* cell population was selected,

passaged twice, and similarly re-transfected with either pX459 or pX459 construct containing sgRNA#2.

Genomic DNA Extraction and Sanger Sequencing

Genomic DNA (gDNA) from harvested cell populations was extracted using the E.Z.N.A. Tissue DNA kit (Omega Bio-tek, Norcross, GA, cat# D3396-01) according to manufacturer's protocol. DNA was amplified using *RBM48* targeting site-specific primer pairs and amplicons were bidirectionally sequenced with the same or custom inner forward and reverse primers as listed in Table S3. All primers were obtained from IDT. Sanger sequencing was performed by Genewiz (South Plainfield, NJ). The sequencing chromatograms were visually analyzed and frequency of indels generated at CRISPR target sites were quantified using Tracking Indels by Decomposition (TIDE) analysis (Brinkman et al. 2014).

Surveyor Nuclease Assay

The genomic region spanning the CRISPR target sites were PCR amplified using primers listed in Table S3. Resultant PCR products were subjected to enable heteroduplex formation using the following temperature cycles: 95°C for 10 min; 95°C to 85°C ramping at -2°C/s; 85°C to 25°C at -0.3°C/s; 25°C for 1 min and 4°C hold. The resultant DNA products were subjected to SURVEYOR nuclease analysis following instructions provided by the manufacturer (IDT).

RNA Extraction, cDNA Synthesis and RT-PCR

Cells were pelleted via centrifugation and suspended in RNAlater® (Sigma) for 24 hours at 4°C prior to storage at -20°C until further use. Cells were homogenized in Trizol reagent (Invitrogen) followed by phase separation. Total RNA was extracted from the resulting supernatants using the RNeasy Plus Universal Kit (Qiagen) per the manufacturer's protocol with the inclusion of gDNA digestion using the RNase-Free DNase Set (Qiagen) as instructed by the manufacturer. gDNA-free total RNA was re-suspended in 30 µl RNase free water and the RNA concentration and purity was determined by Nanodrop 2000C (ThermoFisher Scientific). RNA integrity was confirmed by gel electrophoresis using 1% agarose with ethidium bromide. gDNA-free total RNA (1 µg) was reverse transcribed using SuperScript VILO cDNA Synthesis Kit (Life Technologies, Grand Island, NY) according to the protocol provided by the manufacturer. Reverse transcription (RT) was performed for 10 min at 25˚C, 60 min at 42˚C, and reactions were terminated by 85˚C incubation for 5 min. RT samples were used immediately or stored at -20˚C

until further use. After first strand cDNA synthesis, MIG transcripts were amplified using gene-specific primers positioned as shown in Figure S2 and listed in Table S4.

U12-Type Intron-Containing Template Preamplification for Generating Calibration Curves

As the expression of U12-type intron retained RNA templates is expected to be low in wild-type cells, calibration curves to determine PCR priming efficiencies were generated from the K-562 *RBM48* KO subline. cDNA (5 ng/50 μL reaction) was pre-amplified for 10 cycles with U12-type intron specific primers (Table S5) and 1x SsoAdvanced SYBR Green Supermix (Bio-Rad) using the thermal cycling profile described below minus dissociation curves. PCR amplicons were purified, quantified by NanoDrop 2000C and copy number determined using the online calculator available at https://cels.uri.edu/gsc/cndna.html. Diluted amplicons were then used for generating calibration curves by serial dilution.

Reverse Transcription Quantitative Polymerase Chain Reaction (RT−qPCR)

Quantitative real-time PCR of reverse transcribed cDNA (RT-qPCR) from VC and *RBM48* KO samples was performed in a 96-well format in the Bio-Rad CFX96 Real Time System. The characteristics of the genes of interest (GOIs) and reference genes (RGs) including name, accession number, product length, and the respective primer exon locations and sequences are summarized in Table S5. The concentration of cDNA refers to, and is derived from, the concentration of RNA that was originally synthesized in the cDNA synthesis reaction described above. The real-time PCR mixtures consisted of 1.0 μL cDNA template (corresponding to 1 ng cDNA/well), 400 nM specific sense primer, 400 nM specific antisense primer, RNase/DNase-free water, and 1x SsoAdvanced SYBR Green Supermix (Bio-Rad) in a final volume of 10 μL/well. The thermal profile of the PCR procedure following the SsoAdvanced supermix protocol was: 1) enzyme activation and initial DNA denaturation at 95°C for 30 seconds (sec); 2) 5 sec denaturation at 95°C, 20 sec annealing and extension at 60°C repeated for 40 cycles (amplification data collected at the end of each cycle); 3) dissociation curve consisting of 5 sec incubation at 65 – 95°C in 0.5°C increments. Dissociation curves were used to validate product specificity. Experimental control and Quality Check (QC) assays (Bio-Rad) were performed to evaluate RNA quality, genomic DNA contamination, and PCR reaction performance. Biological replicates from six independent experimental reproductions (6 passages per genotype), which passed control and QC assays, were amplified in triplicate for the target GOIs and RGs. Real-time PCR efficiencies were determined by calibration curves generated as described above for minor intron-containing templates (ranging from 2×10^3 – 2×10^6 copies/well) or from a 10-fold

dilution series of reverse transcribed cDNA (derived from K-562 *ATCC*® CCL-243™ cells) ranging from 5 pg to 50 ng cDNA/well for total GOIs and RGs. The quantification cycle (Cq) was plotted against the log amount of cDNA input and the relationship between Cq values and RNA concentration was calculated by linear regression to find a slope and intercept that predicts cDNA amounts and correlation coefficient (R^2). Amplification efficiencies (E) (Pfaffl 2001; Vandesompele et al. 2002) were calculated according to the equation E = (10(−1/slope) - 1) x 100 and are expressed as a percentage. The qPCR parameters providing the standard curve for each primer pair are summarized in Table S6. A cDNA positive control inter-run calibrator (1 ng/well) was included in every run and the ΔCqs were verified to be <0.01 between runs. Validation of RGs (6 total) was determined using RefFinder and selecting the three genes with the lowest geometric mean rank having a geNorm stability value <0.5 (Vandesompele et al. 2002; Hellemans, Vandesompele 2014). Gene expression was determined using the Bio−Rad CFX Manager™ software version 3.0 and calculated using the efficiency corrected model (Pfaffl 2001) of the $\Delta\Delta C_q$ method (Livak, Schmittgen 2001) modified for normalization by geometric averaging of multiple reference genes (Vandesompele et al. 2002). Results are expressed as the ratio of *RBM48* KO ΔCq expression to VC ΔCq expression (Relative Normalized Expression). Unpaired Student's *t*-test was utilized to determine significant differences between VC and KO populations.

Analysis of RBM48-Myc Protein Expression

Following puromycin selection of RBM48-Myc transfectants, heterogeneous sgRNA#2-targeted and vector control cell populations were cultured in fresh media containing 10% FBS and passaged once prior to harvesting cells from three subsequent passages into RNAlater® until further use. Total protein was extracted and analyzed as described previously (Siebert et al. 2011). Briefly, cell pellets were lysed in RIPA lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 1.0% Triton X-100, 0.25% sodium deoxycholate, 5.0 mM EDTA) in the presence of 10 µL/mL HALT protease inhibitor cocktail (Thermo Fisher Scientific) and homogenized. High speed supernatants of the extracts were prepared via centrifugation at 15,000 × g for 15 min at 4°C, and sample aliquots were snap frozen in liquid nitrogen and stored at -80°C until further use. 20 µg total protein, quantified via Bradford assay, was denatured and loaded onto 7.5% Mini-PROTEAN® TGX™ polyacrylamide gel (Bio-Rad) under denaturing conditions and electrophoresed for the separation of proteins. Proteins were wet transferred to PVDF membrane (Millipore, Bedford, MA) and subjected to western blotting using Myc-Tag (9B11) mouse monoclonal and proliferating cell nuclear antigen (PCNA, D3H8P) XP® Rabbit monoclonal

primary antibodies (Cell Signaling Technology, Danvers, MA). Following primary antibody incubation, membranes were washed and incubated with horseradish peroxidase-conjugated anti-mouse IgG (Cell Signaling Technology) and anti-rabbit IgG (Santa Cruz Biotechnology, Santa Cruz, CA) secondary antibodies. Myc-Tag and PCNA Blots were developed by enhanced chemiluminescence using SuperSignal™ West Femto (Thermo Fisher Scientific) or Clarity™ (Bio-Rad) ECL substrates, respectively, and imaged using the Bio-Rad ChemiDoc Touch imaging system. Western blots were subjected to quantification using the Image Studio Lite program version 3.1 (LI-COR Biosciences, Lincoln, NE) and RBM48-Myc protein band density was normalized to PCNA loading control. Statistical analyses were performed using the unpaired Student's *t*-test.

RNA-seq

Approximately 1 x 10^6 cells from 6 biological replicates of paired *RBM48* KO and VC K-562 cell populations were pelleted, re-suspended in Trizol reagent and stored at -80°C prior to RNA extraction. Each replicate was a subsequent cell passage. Homogenized cells were subjected to chloroform/isopropanol extraction. Following DNase treatment, the samples were further purified using RNeasy MinElute Cleanup Kit (Qiagen, cat# 74204). Sample quality was examined using the Agilent Tape-station 2200, and samples displaying RNA Integrity Number (RIN) value greater than 9.5 were deemed acceptable for experiments. The NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, cat# E7530S) was used to generate cDNA libraries and 2 lanes of HiSeq3000 paired-end 100 bp sequences were generated. Libraries and sequencing were outsourced to the University of Florida Interdisciplinary Center for Research.

A total of 750,648,520 reads (31,277,021 per sample average, range 28,153,614 – 37,903,702*)*. 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 --minimum-length=0 --adapter=GATCGGAAGAGCACACGTCT --quality-base=33. Adapter trimmed sequences were quality filtered/trimmed with Trimmomatic v0.22 (Bolger, Lohse, Usadel 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. Only reads remaining in pairs after trimmomatic were used for subsequent analysis. On average, 28,839,533 read pairs per sample (range 26,349,516 – 33,018,146) remained after sequence clean up.

Reads were aligned to the human genome sequence assembly (GRCh38) with HiSAT2 (Pertea et al. 2016; Kim et al. 2019) using the following parameters: --max-intronlen 100000 -q --

pen-noncansplice 6 --no-discordant --rna-strandness RF. Homo_sapiens.GRCh38.87. Annotation (gtf/gff) was used for intron counts and transcripts per million (TPM) normalization. A custom GFF file detailing all non-redundant introns from Homo_sapiens.GRCh38/hg38 was constructed and used to perform U12-type intron retention analysis. Introns within this GFF file were given a unique identifier and annotated as U12- or U2-type introns. Determination of intron type (U12 or U2) was based on a publicly available collection of human U12-type introns (Alioto 2007) available at https://www.crg.eu/en/programmes-groups/guigo-lab/datasets/u12db-database-orthologous-u12-type-spliceosomal-introns. Six-hundred and ninety-five U12-type introns were identified by Alioto (2007) and their sequences and coordinates relative to Human genome annotation version NCBI35/hg17 were available. Conversion of the NCBI35/hg17 iU12-type intron coordinates to Homo_sapiens GRCh38/gh38 was accomplished using a two-step process. The annotation "lift-over' utility available from the UCSC genome browser (https://genome.ucsc.edu/cgi-bin/hgLiftOver) was used to first convert the U12-type intron coordinates from NCBI35/hg17 to CRCh37/hg19, and then again from CRCh37/hg19 to Homo_sapiens.GRCh38.87.

Read counts/gene were determined with the HTSeq-Count utility in the HTSeq package (Ver 0.8.0) (Anders et al, 2015) using the start-stop coordinates of the entire locus and the following parameters: -m intersection-nonempty -s reverse) These counts were used to perform a TPM normalization of expression at each locus for each sample. Intron counts were determined with the HTSeq-Count utility (-m intersection-nonempty -s reverse) using the intron GFF file for features. Intron read counts for all introns were then normalized by the TPM values for the respective samples, and normalized intron counts were compared between *RBM48* KO and VC using student's *t*-tests ($p < 0.05$). Test statistics were corrected to a false discovery rate ≤ 0.05 using the Benjamini-Hochberg method.

Gene ontology (GO) term enrichment analyses were performed using String-DB with default parameters and the genes having greater than 0 TPM in either VC or KO (genes expressed in either K-562 cell subline) as a customized reference. MIGs analyzed were selected based on significant differences (FDR ≤ 0.05) in U12 intron reads and filtered for transcript expression in the populations. Non-coding MIGs were excluded. String-DB results were further filtered by enrichment cutoff of >2.0 and FDR<0.01. GO term redundancy was filtered with ReviGO using Simrel semantic similarity measurement with allowed similarity=0.5.

Comparative Analysis of MIG Homology and Intron Retention

VC and KO population intron reads and their respective flanking exon-exon junction reads were filtered for significant ($q < 0.05$) retention using Student's t-test and *post hoc* adjustment using the Benjamini and Hochberg False Discovery Rate (FDR). Log_2 fold change of U2- and U12-type intron retention between *RBM48* KO and *rbm48* mutants was assessed using raw read counts of KO/mutant populations compared to wildtype populations.

Blastp was performed using all maize MIGs against all human MIGs. The MIGs with blastp score > 80 were considered orthologous if they also share reciprocal blastp hits. We searched for homologous MIGs that share significant intron retention between human *RBM48* KO population and maize mutant *rbm48* (Bai et al. 2019). The position of the intron was determined by direct splice alignment of the protein sequence by SplicePredictor (Usuka and Brendel, 2000). The position of the intron was considered conserved if it spans within 5 amino acids residues between the two species.

Acknowledgments

This work was supported by the National Science Foundation (grant 1412218 to S.L., W.B.B. and A.M.S.); the National Heart Lung and Blood Institute at the National Institutes of Health (grant R01-HL135035 to R.J.W.); the National Cancer Institute at the National Institutes of Health (grant R15-CA182889 to G.J.M.); the Oakland University Research Excellence Fund (to S.L., G.J.M. and R.J.W.) and American Heart Association Innovative Research Grant (to R.J.W.).

References

Aguirre, AJ, RM Meyers, BA Weir, et al. 2016. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. Cancer Discov 6:914-929.

Alioto, TS. 2007. U12DB: a database of orthologous U12-type spliceosomal introns. Nucleic Acids Res 35:D110-115.

Bai, F, J Corll, DN Shodja, et al. 2019. RNA Binding Motif Protein 48 Is Required for U12 Splicing and Maize Endosperm Differentiation. Plant Cell 31:715-733.

Bolger, AM, M Lohse, B Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Boudreau, HE, CG Broustas, PC Gokhale, D Kumar, RR Mewani, JD Rone, BR Haddad, U Kasid. 2007. Expression of BRCC3, a novel cell cycle regulated molecule, is associated with increased phospho-ERK and cell proliferation. Int J Mol Med 19:29-39.

Carette, JE, M Raaben, AC Wong, et al. 2011. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. Nature 477:340-343.

Chabot, B, L Shkreta. 2016. Defective control of pre-messenger RNA splicing in human disease. J Cell Biol 212:13-27.

Cooper, TA, L Wan, G Dreyfuss. 2009. RNA and disease. Cell 136:777-793.

Edery, P, C Marcaillou, M Sahbatou, et al. 2011. Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. Science 332:240-243.

Fouquet, R, F Martin, DS Fajardo, CM Gault, E Gomez, CW Tseung, T Policht, G Hueros, AM Settles. 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.

Gault, CM, F Martin, W Mei, F Bai, JB Black, WB Barbazuk, AM Settles. 2017. Aberrant splicing in maize rough endosperm3 reveals a conserved role for U12 splicing in eukaryotic multicellular development. Proc Natl Acad Sci U S A 114:E2195-E2204.

Grubenmann, CE, CG Frank, S Kjaergaard, EG Berger, M Aebi, T Hennet. 2002. ALG12 mannosyltransferase defect in congenital disorder of glycosylation type lg. Hum Mol Genet 11:2331-2339.

Haeuptle, MA, T Hennet. 2009. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. Hum Mutat 30:1628-1641.

Hall, SL, RA Padgett. 1994. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. J Mol Biol 239:357-365.

Hart, T, M Chandrashekhar, M Aregger, et al. 2015. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell 163:1515-1526.

Hart, T, AHY Tong, K Chan, et al. 2017. Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 (Bethesda) 7:2719-2727.

He, H, S Liyanarachchi, K Akagi, et al. 2011. Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. Science 332:238-240.

Hellemans, J, J Vandesompele. 2014. Selection of reliable reference genes for RT-qPCR analysis. Methods Mol Biol 1160:19-26.

Hong, Z, H Jin, AC Fitchette, Y Xia, AM Monk, L Faye, J Li. 2009. Mutations of an alpha1,6 mannosyltransferase inhibit endoplasmic reticulum-associated degradation of defective brassinosteroid receptors in Arabidopsis. Plant Cell 21:3792-3802.

Huang, D, Y Nagata, V Grossmann, et al. 2015. BRCC3 mutations in myeloid neoplasms. Haematologica 100:1051-1057.

Jafarifar, F, RC Dietrich, JM Hiznay, RA Padgett. 2014. Biochemical defects in minor spliceosome function in the developmental disorder MOPD I. RNA 20:1078-1089.

Jung, HJ, H Kang. 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. Plant J 78:799-810.

Kim, D, JM Paggi, C Park, C Bennett, SL Salzberg. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37:907-915.

Kim, WY, HJ Jung, KJ Kwak, MK Kim, SH Oh, YS Han, H Kang. 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.

Kwak, KJ, BM Kim, K Lee, H Kang. 2017. quatre-quart1 is an indispensable U12 intron-containing gene that plays a crucial role in Arabidopsis development. J Exp Bot 68:2731-2739.

Lee, Y, DC Rio. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annual Review of Biochemistry, Vol 84 84:291-323.

Lim, KH, L Ferraris, ME Filloux, BJ Raphael, WG Fairbrother. 2011. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. Proc Natl Acad Sci U S A 108:11093-11098.

Liu Cm, CM, J McElver, I Tzafrir, R Joosen, P Wittich, D Patton, AA Van Lammeren, D Meinke. 2002. Condensin and cohesin knockouts in Arabidopsis exhibit a titan seed phenotype. Plant J 29:405-415.

Liu, N, Z Liu, W Zhang, Y Li, J Cao, H Yang, X Li. 2018. MicroRNA433 reduces cell proliferation and invasion in nonsmall cell lung cancer via directly targeting E2F transcription factor 3. Mol Med Rep 18:1155-1164.

Livak, KJ, TD Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-408.

Lorkovic, ZJ, DA Wieczorek Kirk, MH Lambermon, W Filipowicz. 2000. Pre-mRNA splicing in higher plants. Trends Plant Sci 5:160-167.

Madan, V, D Kanojia, J Li, et al. 2015. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. Nat Commun 6:6042.

Markmiller, S, N Cloonan, RM Lardelli, et al. 2014. Minor class splicing shapes the zebrafish transcriptome during development. Proc Natl Acad Sci U S A 111:3062-3067.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17:3.

Pertea, M, D Kim, GM Pertea, JT Leek, SL Salzberg. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650-1667.

Pfaffl, MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45.

Ran, FA, PD Hsu, J Wright, V Agarwala, DA Scott, F Zhang. 2013. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8:2281-2308.

Ru, Y, BB Wang, V Brendel. 2008. Spliceosomal proteins in plants. Curr Top Microbiol Immunol 326:1-15.

Simpson, GG, W Filipowicz. 1996. Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery. Plant Mol Biol 32:1-41.

Staley, JP, C Guthrie. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 92:315-326.

Sterne-Weiler, T, JR Sanford. 2014. Exon identity crisis: disease-causing mutations that disrupt the splicing code. Genome Biol 15:201.

Turunen, JJ, EH Niemela, B Verma, MJ Frilander. 2013. The significant other: splicing by the minor spliceosome. Wiley Interdiscip Rev RNA 4:61-76.

Tzelepis, K, H Koike-Yusa, E De Braekeleer, et al. 2016. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep 17:1193-1205.

Vandesompele, J, K De Preter, F Pattyn, B Poppe, N Van Roy, A De Paepe, F Speleman. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:RESEARCH0034.

Vega, H, Q Waisfisz, M Gordillo, et al. 2005. Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet 37:468-470.

Verma, B, MV Akinyi, AJ Norppa, MJ Frilander. 2018. Minor spliceosome and disease. Semin Cell Dev Biol 79:103-112.

Wang, JP, Y Jiao, CY Wang, ZB Xu, B Zhang. 2017a. Rb knockdown accelerates bladder cancer progression through E2F3 activation. Int J Oncol 50:149-160.

Wang, T, K Birsoy, NW Hughes, KM Krupczak, Y Post, JJ Wei, ES Lander, DM Sabatini. 2015. Identification and characterization of essential genes in the human genome. Science 350:1096-1101.

Wang, T, H Yu, NW Hughes, B Liu, A Kendirli, K Klein, WW Chen, ES Lander, DM Sabatini. 2017b. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell 168:890-903 e815.

Wei, F, E Coe, W Nelson, et al. 2007. Physical and genetic structure of the maize genome reflects its complex evolutionary history. PLoS Genet 3:e123.

Westrick, RJ, K Tomberg, AE Siebert, G Zhu, ME Winn, SL Dobies, SL Manning, MA Brake, AC Cleuren, LM Hobbs. 2017. Sensitized mutagenesis screen in Factor V Leiden mice identifies thrombosis suppressor loci. Proceedings of the National Academy of Sciences 114:9659-9664.

Will, CL, R Luhrmann. 2005. Splicing of a rare class of introns by the U12-dependent spliceosome. Biological Chemistry 386:713-724.

Xu, T, BM Kim, KJ Kwak, HJ Jung, H Kang. 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.

Zdebska, E, B Bader-Meunier, PO Schischmanoff, T Dupre, N Seta, G Tchernia, J Koscielak, J Delaunay. 2003. Abnormal glycosylation of red cell membrane band 3 in the congenital disorder of glycosylation Ig. Pediatr Res 54:224-229.

Zhang, F, Q Zhou. 2018. Knockdown of BRCC3 exerts an antitumor effect on cervical cancer in vitro. Mol Med Rep 18:4886-4894.

Ziebold, U, T Reza, A Caron, JA Lees. 2001. E2F3 contributes both to the inappropriate proliferation and to the apoptosis arising in Rb mutant embryos. Genes Dev 15:386-391.

Figure Legends

Figure 1. Functional CRISPR/Cas9-mediated knockout of *RBM48* in K-562 cells.
(A) Schematic of human *RBM48* gene structure (Refseq accession: NM_032120.4) displaying the design and position of the sgRNAs used for Cas9 targeting. Gray, green and blue boxes and black lines display exons and introns, respectively with the exon number indicated below each box. The open reading frame (ORF) is indicated in blue and the RNA Recognition Motif (RRM) in green. Arrowheads indicate position and direction of PCR primers used in Figure 1B, E and F for sgRNA#1 (red) and #2 (blue) and are listed in Table S3. The target site of each sgRNA is expanded above the structure. Black, blue and green text indicates sequences within intronic, ORF and RRM domain regions of *RBM48*. The black double arrowheads mark the predicted cleavage site of Cas9. PAM = protospacer adjacent motif. **(B)** Surveyor Nuclease Assay. Heterogeneous K-562 cell populations transfected with empty pX459 vector control (VC), or sgRNA#1 (left panel set, red underline) or sgRNA#2 (right panel set, blue underline), were amplified with designated primers and digested with Surveyor nuclease to determine indel formation within transfected cell populations at 0, 10 and 16 days post-selection. The expected sizes for uncut (U) amplicons spanning the regions targeted by sgRNA#1 and #2 were 581 and 599 bp respectively. C = cut with surveyor nuclease. NTC = no template control. The molecular weight standards are indicated on the left. **(C)** and **(D)** Cell viability assays of transfected K-562 cells. Viable cell counts monitored every 48 h from 72 h post-puromycin selection (Day 0) through Day 10 are shown in (C). *q<0.05. Viable cell counts on Day 16 are shown in (D). **(E)** Sanger sequencing chromatograms of the amplified region flanking the Cas9 cleavage site targeted by sgRNA#1 (top panel) and sgRNA#2 (bottom panel) at Day 16 post-selection. **(F)** TIDE analysis of the chromatograms displayed in Figure 1E. Genomic DNA amplified from VC cell populations was used as non-targeted K-562 input sequence for the analysis.

Figure 2. *RBM48* transcript expression from the KO K-562 cell population.
(A) RNA-seq read depth analysis of *RBM48* expression between the six summed libraries of vector control (VC) and knockout (KO) biological replicates. Blue arrowheads mark the position of primers used for RT-qPCR in (B). The sequence of the primers is listed in Table S5. The expanded bottom panel displays the region spanning the sgRNA#2 target site in exon 3 of the *RBM48* sequence. **(B)** RT-qPCR of relative *RBM48* transcript expression between *RBM48* KO and VC cells are displayed as the mean ± SEM from six subsequent passages analyzed by RNA-seq (n = 6) and normalized to reference genes *HPRT1*, *IPO8*, and *PGK1*. **(C)** *RBM48* KO expressed transcripts determined from RNA-seq. The left panel displays the sequences from the *RBM48* sgRNA#2-targeting region spanning positions 392-467 of the NM_032120.4 transcript variant. The number of nucleotide insertions/deletions (indels) within each sequence are indicated. The predicted translational product of each transcript in reference to amino acids 125-149 of the NP_115496.2 protein product is shown on the right. Orange text designates changes in protein sequence. Red asterisks mark premature stop codons.

Figure 3. Ablation of RBM48-Myc protein levels by sgRNA#2-targeting of Cas9.
(A) Schematic of *RBM48-Myc* displaying the sgRNA#2 targeting site. **(B)** TIDE analysis of indels present in sgRNA#2-targeted *RBM48-Myc* cells. **(C)** Western blot (left panel) and densitometric analysis (right panel) of RBM48-Myc protein levels between KO and VC cell populations. The relative intensities of RBM48-Myc are displayed as the mean ± SEM from three independent experiments (n = 3) and normalized to PCNA levels (loading control). *p<0.001.

Figure 4. RT-qPCR validation of U12-type intron retention in RNA-seq libraries.
(A-D) Summed read depth analysis of six biological replicates per VC and KO samples are shown at the top of each panel representing the regions of **(A)** *DIAPH1*, **(B)** *TXNRD2*, **(C)**

MAPK1, and **(D)** *MAPK3* with significant (q<0.05) retention of a U12-type intron in KO cell populations compared to VC. Brace symbol indicates the U12-type intron with increased read depth in KO cells. The bottom panels display RT-qPCR analysis of the relative expression (mean ± SEM; n = 6) of each respective total transcript (left) and corresponding U12-type intron-containing transcript (right) from KO samples compared to VC. Expression was normalized to *HPRT1*, *IPO8*, and *PGK1* expression. Primers for RT-qPCR analysis are available in Table S5. *p<0.05.

Figure 5. Comparative analysis of U12 intron retention between the RBM48 mutants of human and maize.
(A) Log_2 fold change (Log2fc) plots of all introns displaying significant (q<0.05) retention in human (left) and maize (right) RBM48 mutants between the VC and wildtype samples, respectively. All introns sampled have at least 10 exon-exon junction reads in both human and maize samples. Maize data is available from Bai et al. (2019). **(B)** The number of human and maize MIGs is indicated based on U12DB and ERISdb, respectively. The number of homologous MIGs between humans and maize that are significantly mis-spliced in both human and maize RBM48 mutants are indicated. Purple and orange color represents human and maize data, respectively.

Figure 6. Comparison of U12-type intron retention in homologous MIGs between humans and maize. (A) Display of summed read depth analysis of significantly (q<0.05) increased U12-type intron retention in KO compared to VC cells for *SMYD2*, *TAPT1*, *WDR91*, and *ZPR1*. The brackets mark the position of the U12-type introns. **(B)** Semi-quantitative RT-PCR comparison of retention in U12-type introns indicated by band intensity in human (left panel) *SMYD2*, *TAPT1*, *WDR91*, and *ZPR1* and their maize (right panel) homologs GRMZM2G457881, GRMZM2G347645, GRMZM2G158179, and GRMZM2G351582, respectively. The position and sequence of the primers used during RT-PCR analysis are displayed in Figure S2 and Table S4, respectively.

Table 1. Conserved MIGs Mis-spliced in Human and Maize RBM48 Knockout Mutants

HGNC Symbol	HGNC Gene Name	Human ID†	Maize Homolog	Maize ID†
ALG12*	ALG12 alpha-1,6-mannosyltransferase	Q9BV10	GRMZM2G152194*	A0A1D6E4A9
BRCC3	BRCA1/BRCA2-containing complex subunit 3	P46736	GRMZM2G096491	B4FWV0
BTAF1	B-TFIID TATA-box binding protein associated factor 1	O14981	GRMZM2G168096	A0A1D6MDZ3
DERL2*	derlin 2	Q9GZP9	GRMZM2G117388* GRMZM2G143817	Q4G2J6 Q4G2J5
E2F3*	E2F transcription factor 3	O00716	GRMZM2G041701* GRMZM2G052515*	A0A1D6Q2U1 A0A1D6ITE9
FRA10AC1	FRA10A associated CGG repeat 1	Q70Z53	GRMZM2G001444	C0HFD7
GPN2*	GPN-loop GTPase 2	Q9H9Y4	GRMZM2G093716*	A0A1D6HAG4
IPO9*	importin 9	Q96P70	GRMZM2G457415*	A0A1D6NYK5
MAEA*	macrophage erythroblast attacher	Q7L5Y9	GRMZM2G177026*	B6TF70
SLC66A1	solute carrier family 66 member 1	Q6ZP29	GRMZM2G024733	C0HEC6
SACM1L*	SAC1 like phosphatidylinositide phosphatase	Q9NTJ5	GRMZM2G047894 GRMZM2G171080* GRMZM2G418916*	K7W0E1 K7UVY0 A0A1D6HAJ0
SMC3	structural maintenance of chromosomes 3	Q9UQE7	GRMZM2G456570	A0A1D6MEC8
SMYD2*	SET and MYND domain containing 2	Q9NRG4	GRMZM2G457881*	B4FCH5
SMYD3	SET and MYND domain containing 3	Q9H7B4	GRMZM2G080462	C4IZK2
TAPT1	transmembrane anterior posterior transformation 1	Q6NXT6	GRMZM2G347645	A0A1D6GJ53
WDR91*	WD repeat domain 91	A4D1P6	GRMZM2G158179*	C0PKZ7
XRCC5	X-ray repair cross complementing 5	P13010	GRMZM2G137968	A0A1D6LC75
ZPR1	ZPR1 zinc finger	O75312	GRMZM2G351582	B4FUA0

† UniProtKB/Swiss-Prot ID

* Conserved MIGs significantly mis-spliced in human *RBM48* and *ZRSR2* and maize *rbm48* and *rgh3* mutants.

Figure 1

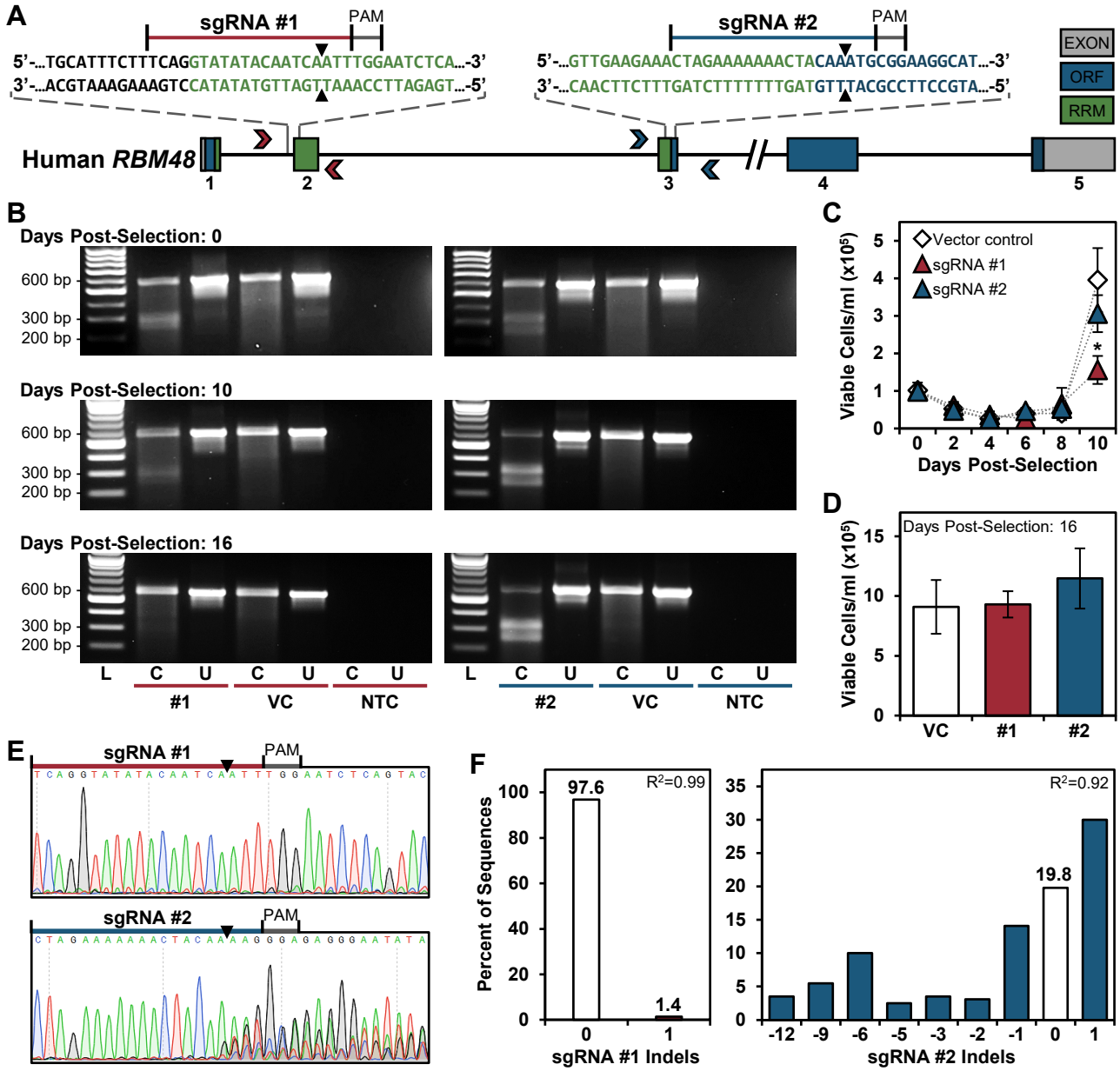

Figure 2

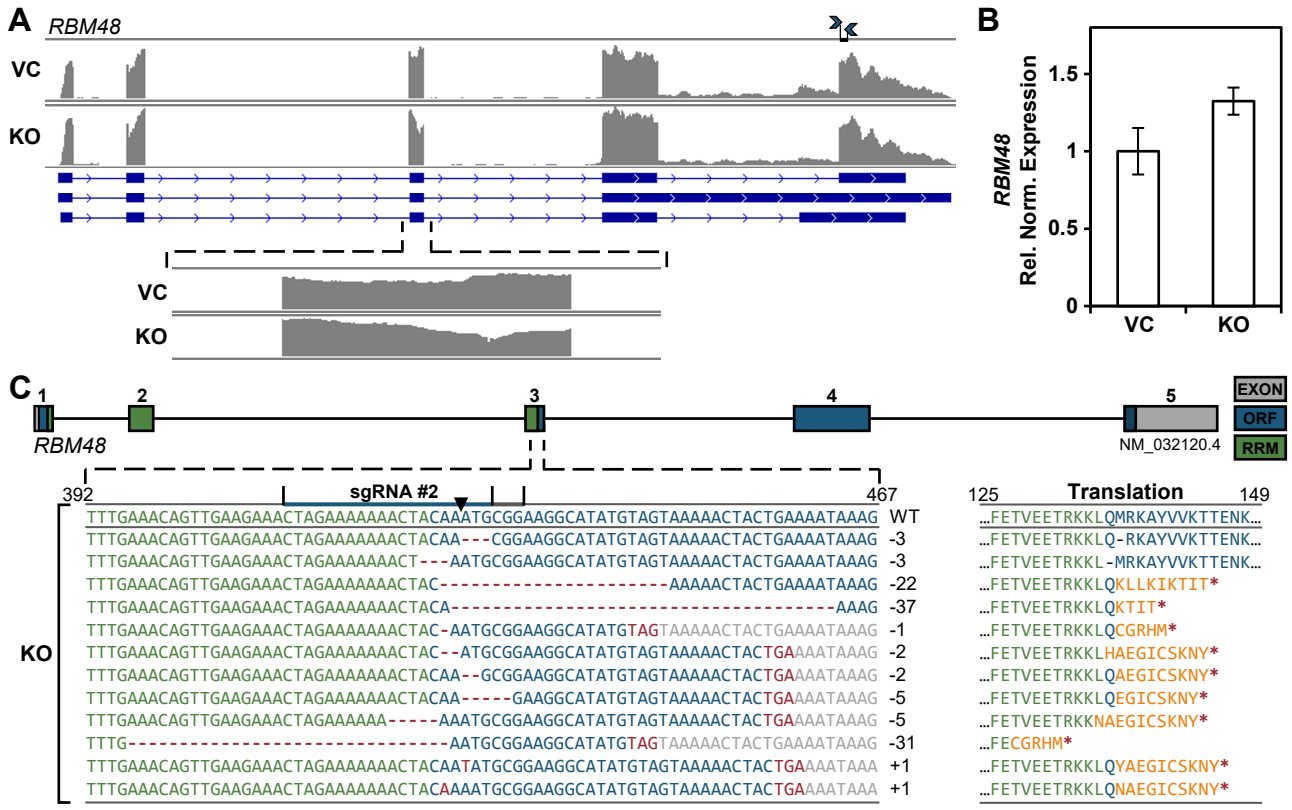

Figure 3

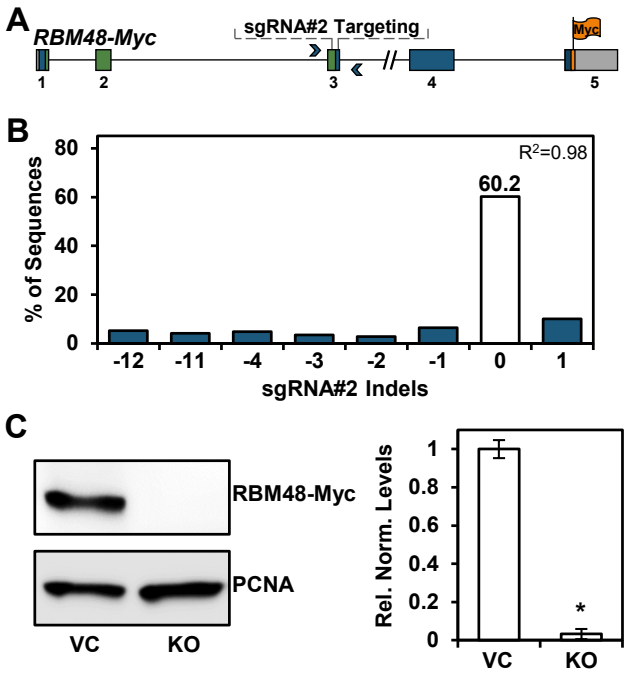

Figure 4

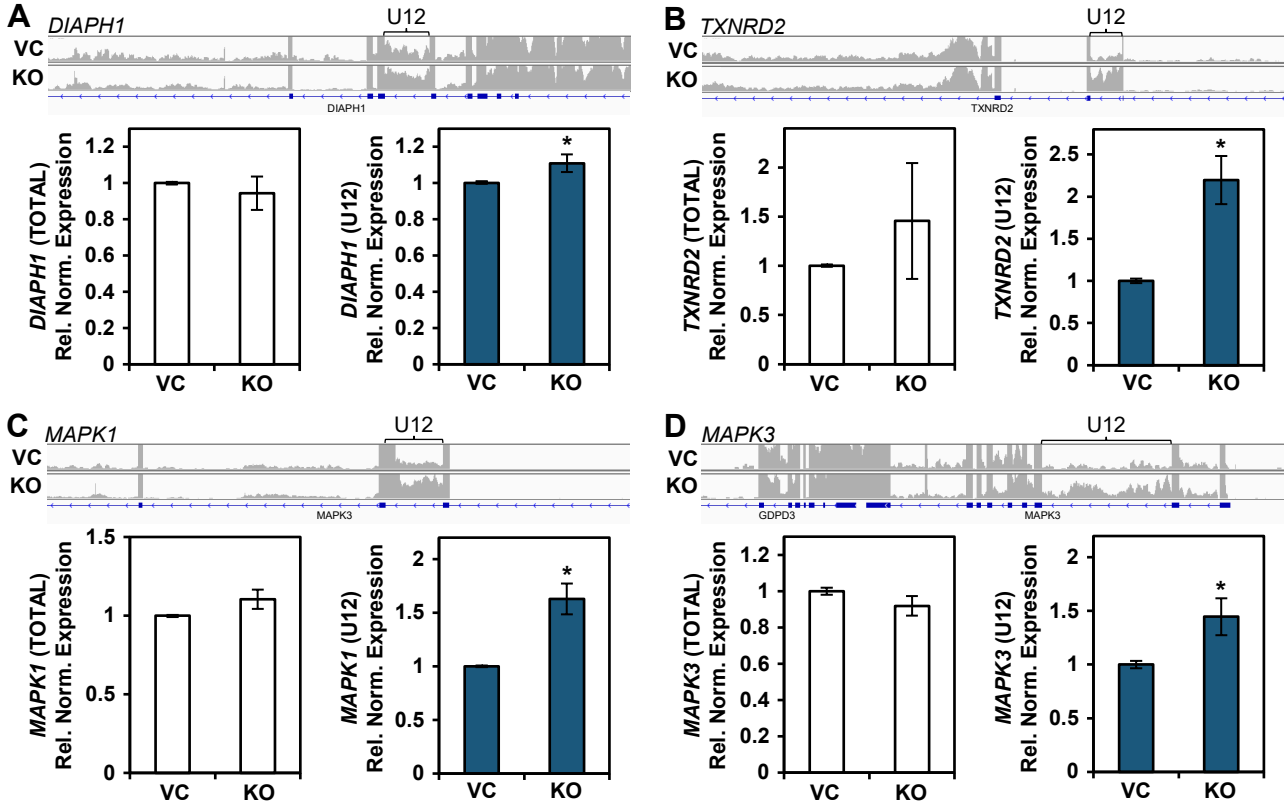

Figure 5

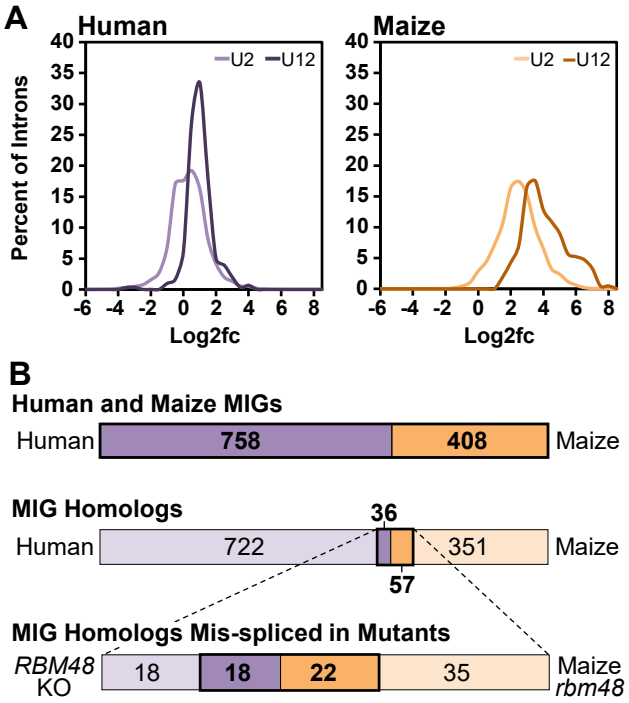

Figure 6