# Distinct functions for the paralogous RBM41 and U11/U12-65K proteins in the minor spliceosome

Antto J. Norppa [1], Iftekhar Chowdhury[2], Laura E. van Rooijen[3], Janne J. Ravantti[4],
Berend Snel [3], Markku Varjosalo[2] and Mikko J. Frilander [1,*]

[1]Institute of Biotechnology, University of Helsinki, Helsinki, Finland
[2]Molecular Systems Biology Research Group and Proteomics Unit, Institute of Biotechnology, University of Helsinki, Helsinki, Finland
[3]Theoretical Biology and Bioinformatics, Department of Biology, Faculty of Science, Utrecht University, 3584 CH Utrecht, the Netherlands
[4]Molecular and Integrative Biosciences Research Programme, University of Helsinki, Helsinki, Finland

*To whom correspondence should be addressed. Tel: +358 2941 59509; Email: mikko.frilander@helsinki.fi

## Abstract

Here, we identify RBM41 as a novel unique protein component of the minor spliceosome. RBM41 has no previously recognized cellular function but has been identified as a paralog of U11/U12-65K, a known unique component of the U11/U12 di-snRNP. Both proteins use their highly similar C-terminal RRMs to bind to 3′-terminal stem-loops in U12 and U6atac snRNAs with comparable affinity. Our BioID data indicate that the unique N-terminal domain of RBM41 is necessary for its association with complexes containing DHX8, an RNA helicase, which in the major spliceosome drives the release of mature mRNA from the spliceosome. Consistently, we show that RBM41 associates with excised U12-type intron lariats, is present in the U12 mono-snRNP, and is enriched in Cajal bodies, together suggesting that RBM41 functions in the post-splicing steps of the minor spliceosome assembly/disassembly cycle. This contrasts with U11/U12-65K, which uses its N-terminal region to interact with U11 snRNP during intron recognition. Finally, while RBM41 knockout cells are viable, they show alterations in U12-type 3′ splice site usage. Together, our results highlight the role of the 3′-terminal stem-loop of U12 snRNA as a dynamic binding platform for the U11/U12-65K and RBM41 proteins, which function at distinct stages of the assembly/disassembly cycle.

## Graphical abstract
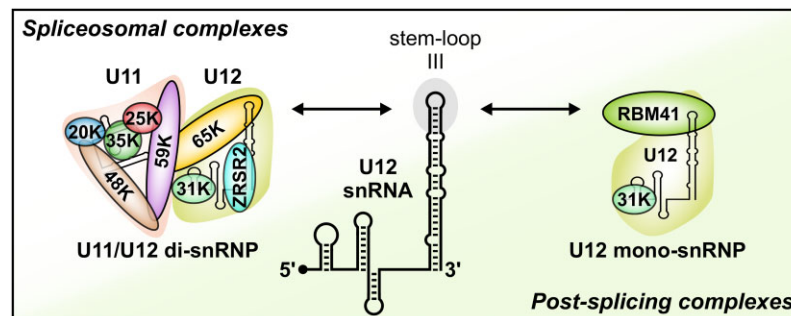


## Introduction

In the majority of metazoan species, the removal of spliceosomal introns from pre-mRNA is carried out by two parallel ribonucleoprotein machineries: the major spliceosome, which splices the U2-type (major) introns, and the minor spliceosome, responsible for splicing of the U12-type (minor) introns. The overall architecture of the major and minor spliceosomes is similar. Both are composed of five small nuclear ribonucleoprotein (snRNP) particles—U1, U2, U4, U5 and U6 in the major spliceosome or U11, U12, U4atac, U5 and U6atac in the minor spliceosome—and additional non-snRNP splicing factors (1–6). While the major spliceosome has been extensively characterized structurally (7), the structures of a minor spliceosome activated for the first step of splicing (B$^{act}$) (8) and the U11 snRNP (9) are presently the only high-resolution

cryo-EM structures of minor spliceosomal complexes available. However, it is generally accepted that the overall assembly pathway of the minor spliceosome is similar to that of the major spliceosome as both spliceosomes utilize similar snRNP complexes and a two-step splicing mechanism with branching and exon ligation reactions (2,3,10). A key mechanistic difference is in the recognition of U12-type introns, which is carried out by a preformed U11/U12 di-snRNP that cooperatively binds to both the 5′ splice site (5′ss) and branch point sequence (BPS) (11,12). In contrast to the earlier steps of the minor spliceosome assembly, very little is known about the post-catalytic events of the minor spliceosome cycle. In the major spliceosome, the DHX8 (hPrp22) helicase drives the release of the ligated exon product from the post-catalytic (P) complex and turns it to the intron lariat spliceosome (ILS)

which is subsequently disassembled by the DHX15 (hPrp43) helicase (13–15). Recycling of snRNPs for subsequent rounds of splicing is thought to take place in the Cajal body, which is also the cellular site for other processes related to snRNP biogenesis (16). There, the Cajal body-localized recycling factor SART3 is thought to function in both spliceosomes (17,18).

Both the major and the minor spliceosome contain four unique small nuclear RNAs (snRNAs) and a number of unique protein components that are not found in the other spliceosome, while the majority of proteins are likely to be shared (8,19–21). The first set of minor spliceosome-specific proteins (20K, 25K, 31K, 35K, 48K, 59K, 65K) were identified over 20 years ago by affinity purification and mass spectrometry analysis of the U11/U12 di-snRNP and U11 mono-snRNP fractions (19,20). Of these, U11/U12-65K, U11-59K and U11-48K form a chain of interactions connecting the U11 and U12 mono-snRNPs into a di-snRNP (22,23) and are essential for the stability of the di-snRNP (23–25). The U11 snRNP-associated U11-48K protein recognizes the 5′ splice site together with the U11 snRNA and interacts with the U11-59K protein (23), which is further engaged in an interaction with the N-terminal part of U11/U12-65K (22). The C-terminal RNA recognition motif (RRM) of U11/U12-65K binds the 3′-terminal stem-loop of U12 snRNA (26), but can also interact with the 3′-terminal stem-loop of the U6atac snRNA (27). ZRSR2, a component of the U11/U12 di-snRNP responsible for 3′ splice site recognition (20), has also been shown to almost exclusively affect splicing of U12-type introns in the cell (28), even though it may also have a separate role in the major spliceosome (29).

The notion that specific protein components of the minor spliceosome are needed only during the intron recognition phase and not in the later assembly steps has been challenged only very recently. The first such factor, the plant ortholog of RBM48, was originally identified as a U12-type intron splicing factor from a transposon screen in maize (30), and its specificity for U12-type introns was later confirmed in human cells (31). A cryo-EM structure of the catalytically activated minor spliceosome (B$^{act}$ complex) revealed that RBM48, together with additional three unique proteins ARMC7, SCNM1 and CRIPT, are all specific components of the B$^{act}$ complex (8). The first unique protein component of the U4atac/U6atac di-snRNP and U4atac/U6atac.U5 tri-snRNP, CENATAC, was initially identified as a human disease gene and was shown to be specifically required for the splicing of U12-type introns with AT-AN termini (32). Similarly, mutations in DROL1, the plant ortholog of TXNL4B, the main interactor of CENATAC (32), were shown to lead to splicing defects with U12-type introns with AT-AC termini in *Arabidopsis thaliana* (33).

In this work, we provide evidence that the specific protein components in the minor spliceosome are not limited to intron recognition and catalytic steps. We show that RBM41, a closely related paralog of the U11/U12-65K protein, is a novel specific component of the minor spliceosome. The two paralogous proteins have a similar shared dual RNA binding specificity *in vitro*, interacting with the same 3′-terminal stem-loops of U12 and U6atac snRNAs with approximately equal affinity. However, while the U11/U12-65K functions during the intron recognition step, our results provide strong evidence that RBM41 functions in post-catalytic steps of the U12-type intron splicing. This further suggests that the 3′-terminal stem-loop of the U12 snRNA has a role as a protein-binding platform that is dynamically recognized first by the U11/U12-65K

during the early steps of the spliceosome assembly followed by an exchange to RBM41 during or after the catalytic steps of splicing.

## Materials and methods

### Antibodies

The following antibodies were used in this study at the dilutions indicated for western blot (WB) and immunofluorescence (IF): anti-HA (BioLegend, 16B12, WB: 1:2000, IF: 1:1000), anti-GAPDH (Proteintech, 60004-1, WB: 1:5000), anti-PDCD7 (Proteintech, 12485-1-AP, WB: 1:500), anti-RBM41 (Atlas Antibodies, HPA042881, WB: 1:1000, IF: 1:200), anti-RNPC3 (Santa Cruz, sc-514951, WB: 1:500, IF: 1:100), anti-Sm (Invitrogen, MA5-13449, WB: 1:500), anti-SNRNP48 (Proteintech, 24297–1-AP, WB: 1:500), anti-V5 (Invitrogen, R960-25, WB: 1:5000), anti-ZCRB1 (Bethyl Laboratories, A304-697A, WB: 1:2000), Goat anti-Rabbit Alexa Fluor 488 (Thermo Fischer, A11008, IF: 1:500), Goat anti-Mouse Alexa Fluor 568 (Abcam, Ab175473, IF: 1:500), Goat Anti-Mouse IgG HRP, light chain-specific (Jackson ImmunoResearch, 115–035-174, WB: 1:10 000), Goat Anti-Mouse IgG HRP (Millipore, AP308P, WB: 1:10000), Donkey Anti-Rabbit IgG HRP (Amersham, NA934V, WB: 1:10 000). Mouse IgG2a kappa Isotype Control (Invitrogen, 14-4724-82) served as a negative control antibody in IP experiments. eBioscience Avidin-HRP (Invitrogen, 18-4100, WB: 1:500) was used to detect biotinylated proteins. Protein G-HRP (Millipore, 18-161, WB: 1:5000) was used in western blotting after IP to avoid detection of heavy and light chains from the IP antibody.

### Plasmid construction

RBM41 cDNA sequence, corresponding to Ensembl transcript ENST00000372479.7, was amplified from HEK293 cDNA in two fragments. Gibson assembly was used to assemble the fragments into vector pCI-neo in-frame with an N-terminal V5 epitope tag, resulting in pCI-neo-V5-RBM41. For BioID cell line construction, full-length RBM41 (1–413), RBM41 N-terminal fragment (1–258), RBM41 C-terminal fragment (259–413) and full-length 65K were cloned into MAC-tag-N vector (Addgene #108078) using Gateway cloning as described (34). Human DHX8 cDNA (MGC clone 5529639) was obtained from the Genome Biology Unit at the University of Helsinki and cloned into pCI-neo in-frame with an N-terminal V5 tag. Point mutations were introduced into plasmids using site-directed mutagenesis with Phusion polymerase (Thermo).

### Expression and purification of recombinant proteins

C-terminal fragments of RBM41 (amino acids 267–413) and U11/U12-65K (amino acids 380–517) were expressed as GST-tagged proteins in *E. coli* Rosetta cells. Protein expression was induced with 0.5 mM IPTG for 3 h at 37°C and cells were harvested by centrifugation. Cell pellets were resuspended in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 5 mM DTT, 1× cOmplete protease inhibitor cocktail (Roche) and lysed by sonication (30 s ON / 30 s OFF, 5 min total sonication time, Sonopuls HD 2070 with MS 73 microtip). After removal of cell debris (25 000 g, 15 min), GST-tagged proteins were captured from the lysate by incubation with glutathione agarose

(Pierce) for 1 h at 4°C. The resin was washed 3 times with 20 volumes of Wash/Cleavage buffer (50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM DTT) and GST tag cleaved by overnight digestion with PreScission Protease (Cytiva). Purified proteins were dialyzed against EMSA buffer (20 mM HEPES–KOH, 100 mM KCl, 1.5 mM MgCl$_2$, 5% glycerol, pH 8.0) and protein concentrations measured by Pierce BCA protein assay (Thermo Scientific).

### Electrophoretic mobility shift assay

Recombinant proteins (0.25–25 μM) were incubated for 1 h on ice with 5 nM [γ-$^{32}$P]-ATP-labeled RNA oligonucleotides in binding buffer (20 mM HEPES–KOH (pH 8.0), 100 mM KCl, 1.5 mM MgCl$_2$, 5% glycerol, 0.1 μg/μl BSA, 1 μg/μl yeast RNA (Roche), 1 U/μl RiboLock RNase inhibitor (Thermo Scientific)) in 10 μl final volume. After addition of 2.5 μl of 5× loading buffer (20 mM HEPES–KOH, 100 mM KCl, 1.5 mM MgCl$_2$, 50% glycerol, 0.1% bromophenol blue, 0.1% xylene cyanol), 4 μl of each binding reaction was loaded onto a native polyacrylamide gel (6%, 80:1 acrylamide:bisacrylamide ratio, 0.5× TBE, 5% glycerol). Electrophoresis was carried out at 120 V for 1.5 h at 4°C. After drying, the gel was exposed on an imaging plate and scanned using the FLA-5100 phosphorimager. Bound and free RNA bands were quantified using AIDA software (Raytest). Dissociation constants were determined by nonlinear regression using GraphPad Prism (one site – specific binding) from three independent replicates.

### Cell culture, transfection and nuclear extract preparation

HEK293 cells were grown in DMEM, 10% FBS, 1% penicillin–streptomycin and 2 mM L-glutamine at 37°C, 5% CO$_2$. All plasmid transfections were carried out using Lipofectamine 2000 (Thermo Fisher) according to manufacturer's instructions. The growth conditions for HeLa S3 suspension cells and the subsequent nuclear extract preparation have been described by de Wolf *et al.* (32).

### CRISPR/Cas9-mediated knockout of RBM41

HEK293 cells were transfected with pSpCas9(BB)-2A-Puro vectors with sgRNA sequences targeting RBM41 exon 2 or 3. 24 h after transfection, puromycin was added at 3 μg/ml to enrich for transfected cells and puromycin treatment continued for 72 h. After puromycin selection, monoclonal cell lines were obtained by limiting dilution in 96-well plates. Genomic DNA was extracted from single-cell clones using the NucleoSpin Tissue kit (Macherey–Nagel), and the targeted areas amplified by PCR using primers in the introns flanking exons 2 and 3 (Supplementary Table S1). Clones were screened for editing using the Surveyor Mutation Detection Kit (IDT). Positive clones were verified by TOPO cloning of PCR products followed by sequencing, as well as by direct sequencing of the PCR products and deconvolution of sequencing traces using the DECODR tool (35).

### Northern and western blotting

Northern blotting was carried out exactly as described (36) using LNA or DNA oligonucleotide probes listed in Supplementary Table S1. For western blotting, protein samples were resolved on 4–12% NuPAGE gels and transferred to Amersham Hybond P membrane using the Amersham TE 22 Mighty Small Transphor tank. After blocking for 1 h in 5% milk/TBS-T, membranes were incubated for 1 h with primary antibodies, followed by 3 × 5 min washes with TBS-T. After 1 h incubation with an HRP-conjugated secondary antibody, membranes were washed 5 × 5 min with TBS-T. Chemiluminescence was detected using the SuperSignal West Atto (Thermo Scientific) or Amersham ECL Prime detection reagent and the LAS-3000 imager (Fuji).

### RNA immunoprecipitation and protein co-immunoprecipitation

For RNA immunoprecipitation with V5-tagged proteins, pCI-neo vectors for expressing V5-tagged proteins or empty pCI-neo vector (4 μg) were transfected into HEK293 cells in 6-well plate format using Lipofectamine 2000. For RNA immunoprecipitation with endogenous proteins in cell lysate, typically ~20 × 10$^6$ cells were used for one IP. 24 h after transfection, cells were washed with 1× PBS, scraped into lysis buffer (20 mM HEPES, pH 7.9, 137 mM NaCl, 10% glycerol, 1% NP-40 + 1× cOmplete protease inhibitor cocktail + 0.5 U/μl RiboLock) and sonicated (5 × 30 s, Bioruptor Twin, High setting). Cell debris was removed by centrifugation at 16000 g for 15 min. The supernatant was incubated o/n with rotation at 4°C with 2 μg IP antibody and for 1 h with Dynabeads Protein G (Invitrogen). Alternatively, the antibody was prebound to Dynabeads and incubated with lysate for 1 h at 4°C. After five washes with 200 μl lysis buffer, co-immunoprecipitated RNA was isolated by treatment of beads with Proteinase K, phenol:chloroform:isoamyl alcohol (pH 4.8) extraction and ethanol precipitation. For nuclear extract RIP experiments, 60 μl HeLa S3 nuclear extract was used for one IP. Nuclear extract was preincubated for 10 min at 37°C in 13 mM HEPES (pH 7.9), 2.4 mM MgCl$_2$, 40 mM KCl, 2 mM DTT, 20 mM creatine phosphate, 0.5 mM ATP. After preincubation, the nuclear extract was diluted 1:5 and adjusted to 20 mM HEPES (pH 7.9), 137 mM KCl, 10% glycerol, 0.1% NP-40, 0.5 U/μl RiboLock, 1× cOmplete protease inhibitor. The remaining steps were carried out as described above.

Protein co-immunoprecipitation experiments were carried out similarly to RIP experiments, except that the IP antibodies were crosslinked to Dynabeads Protein G with bis(sulfosuccinimidyl)suberate (BS$_3$, Thermo) and complexes were eluted from the beads by heating in 1× NuPAGE LDS Sample Buffer for 5 min at 95°C.

### Glycerol gradient centrifugation

HeLa S3 nuclear extract (200 μl), prepared essentially as described in Tarn and Steitz (37), was adjusted to 13 mM HEPES (pH 7.9), 2.4 mM MgCl$_2$, 40 mM KCl, 2 mM DTT, 20 mM creatine phosphate, 0.5 mM ATP, and pre-incubated for 10 min at 30°C in a final volume of 300 μl. After incubation, 210 μl of Gradient buffer (20 mM HEPES (pH 7.9), 40 mM KCl, 2 mM DTT, 2.4 mM MgCl$_2$) was added, samples were centrifuged briefly (20 000 g, 1 min) and the supernatant was loaded on top of 10–30% glycerol gradient in Gradient buffer. Gradients were centrifuged in a Sorvall TH-641 rotor at 29 000 rpm, 4°C for 18 h and fractionated using BioComp Piston Gradient Fractionator. Gradient preparation, centrifugation and fractionation was carried out by the HiLIFE Biocomplex unit at the University of Helsinki. For RNA extraction, 20% of each fraction was treated with Proteinase K,

phenol:chloroform extracted and precipitated with ethanol. The remaining 80% of each fraction was precipitated with TCA for western blot analysis.

## BioID analysis

For each cell line, Flp-In™ T-REx™ 293 cells were grown in 5 × 15 cm plates to ∼70% confluency. MAC-tagged protein expression and biotinylation was induced by addition of 2 μg/ml of tetracycline and 50 μM biotin. Cells were harvested 24 h after induction by pipetting up and down with PBS-EDTA, centrifugation at 1200 g for 5 min, and snap freezing the pellet in liquid nitrogen. Three independent replicates of 5 × 15 cm dishes were prepared for each cell line. BioID analysis was carried out essentially as described previously (34,38).

## Immunofluorescence

HEK293 or Flp-In™ T-REx™ 293 cells were grown on poly-ʟ-lysine coated coverslips in a12-well plate. Cells were fixed with 4% paraformaldehyde/1× PBS for 10 min at room temperature. After two washes with 1× PBS, cells were permeabilized with 0.2% Triton X-100/1× PBS for 10 min and blocked with 1% BSA/1× PBS for 30 min at room temperature. Incubation with primary antibodies, diluted in blocking buffer, was carried out overnight at 4°C. Following 3 × 5 min washes with 1× PBS, cells were incubated with Alexa Fluor 488 or 568 conjugated secondary antibodies and DAPI (0.5 μg/ml) for 1 h at room temperature. Antibodies used for immunofluorescence and their respective dilutions are listed in the Antibodies section above. After final washes with 1× PBS (3 × 5 min), coverslips were mounted onto slides with Pro-Long Diamond (Invitrogen). Imaging was carried out using a Leica DM5000B microscope.

## cDNA synthesis and RT-PCR

RNA was treated with RQ1 RNase-free DNase (Promega) to remove any genomic DNA contamination. cDNA synthesis was carried out using Maxima H Minus RT (Thermo) and random primers according to the manufacturer's instructions. For PCR, Phire polymerase (Thermo) and primers listed in Supplementary Table S1 were used.

## High-throughput sequencing

Total RNA isolated using Trizol extraction followed by an additional acidic phenol (pH 5.0) extraction. RNAseq libraries were constructed using Illumina TruSeq Stranded Total RNA kit (Illumina) Human Ribo-Zero rRNA depletion kit (Illumina). Paired-end 150 + 150 bp sequencing was performed at the Institute for Molecular Medicine Finland FIMM Genomics unit with Illumina NovaSeq 6000 using partial S4 flow cell lane. The STAR aligner (39) was used for mapping the paired sequence reads to the genome (hg38/GRCh38). Transcript annotations were obtained from GENCODE (v29). The length of the genomic sequence flanking the annotated junctions (sjdbOverhang parameter) was set to 161. The Illumina adapter sequences AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC and AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA-GATCTCGGTGGTCGCCGTATCATT were, respectively, clipped from the 3′ of the first and the second pairs in the read libraries (using clip3pAdapterSeq parameter). Statistics of the RNAseq data are presented in the Supplementary Table S5.

## Differential alternative splicing analysis and intron retention analysis

Differential alternative splicing (AS) analysis was done using Whippet (v0.11) (40). Both merged aligned reads (bam files) and AS event annotations from GENCODE (v29) were used to build the index reference for AS events. To detect the significantly differential events, probability cutoff of Pr > 0.9 and Percentage Spliced In deviation cutoff of |ΔΨ| > 0.05 were used. Differential intron retention was analyzed with IRFinder-S using SUPPA2 wrapper (41,42). A custom list of human U12-type intron coordinates (Supplementary Table S8) combining high-confidence U12-type introns from IntEREst (43), IAOD (44) and MIDB (45) databases was used in the annotation of U12-type introns and their host genes.

## Phylogenetic profiling-based co-evolution analysis

To conduct the phylogenetic profiling analysis of RBM41, we utilized a diverse set of eukaryotic proteomes. This dataset, consisting of 167 eukaryotic species, was previously compiled to represent the eukaryotic tree of life, and the species were selected based on their representation in the tree. We used automatic orthologous groups (OG) based on previous work (46,47), generated using methods such as Orthofinder (48), eggNOG (49) hmm profile database, and OGs from Vosseberg *et al.* (47). However, as RBM41 was not accurately represented in these automatically generated OGs, we manually created the OG for RBM41. This was achieved by performing a blast search with RBM41 against our in-house eukaryotic dataset and creating a Hidden Markov model (HMM) of RBM41, which was then used to perform a Hmm search (50) against the same dataset. To determine the OG, a phylogenetic analysis was carried out with the top 100 entries of the Hmmer search using mafft E-INS-I (51) and IQtree (52). In this phylogeny, a cluster with representatives from diverse eukaryotic groups indicated the RBM41 Orthologous group. Besides the support-value, the species overlap of this cluster with the other putative orthologous groups in the tree solidifies it as resulting from an ancient duplication and having been present in LECA. To compare the phylogenetic similarity, we computed cosine distances between the phylogenetic profile of each automatically generated OG and RBM41. As we suspected that RBM41 was part of the minor spliceosome, we also compared the phylogenetic profile of RBM41 to the profiles of other known spliceosome proteins obtained from Vosseberg *et al.* (53).

# Results

## RBM41 is a paralog of the U11/U12-65K protein and a putative component of the minor spliceosome

Human RBM41 is a ∼47 kDa protein with a single annotated domain, an RNA recognition motif (RRM) near the C-terminus of the protein (at positions 309–387) (Figure 1A). While no specific cellular function has been assigned to RBM41, it has been listed as a paralog of the U11/U12-65K protein (*RNPC3*), a structural component of the U11/U12 di-snRNP complex in the minor spliceosome (22,53). The paralog assignment is based on the local sequence similarity between the C-terminal regions of the two proteins, that encompass the core RRM and its N-terminal expansion

**Figure 1.** RBM41 is a paralog of the U11/U12-65K protein. (**A**) Domain structures of human U11/U12-65K and RBM41 proteins. (**B**) Pairwise sequence alignment of RBM41 and U11/U12-65K. Local sequence alignment was carried out using Matcher and visualized using ESPript 3.0 (80). Identical residues are shown in white text with blue background and similar residues in blue text with white background. Protein secondary structure elements extracted from NMR structures (U11/U12-65K: 5OBN, RBM41: 2CPX) are shown below the alignment. (**C**) Structure of the U11/U12-65K C-terminal RRM (5OBN) showing identical and similar residues between RBM41 and U11/U12-65K. (**D**) AlphaFold-predicted structure of human RBM41 colored for sequence conservation. Conservation is based on a multiple sequence alignment of RBM41 orthologues from 15 animal species (Supplementary Figure S3). Conservation was mapped to the structure with ESPript 3.0 and structure rendered using PyMOL. (**E**) Phylogenetic profile of RBM41 compared to the known minor spliceosome-specific proteins and minor and major spliceosomal snRNAs.

(Figure 1B, C) which in the U11/U12-65K protein is essential for the stability of the C-terminal RRM and its interaction with the U12 snRNA (26). The RRM sequences of 65K and RBM41 conform with the general RRM consensus (54) except for an aromatic to nonaromatic amino acid substitution F/Y352Q at position 3 of the RNP1 motif (Figure 1B). The same substitution is present of in the homologous N-terminal RRMs of the U1A/U2B″/SNF family of spliceosome components, which are characterized by a YQF triad of RNP2 tyrosine and RNP1 glutamine and phenylalanine (55,56; Supplementary Figure S1). These three residues are displayed on the β-sheet surface of the RRM and engage in stacking interactions with RNA nucleobases (55,56). Conservation of the corresponding residues in the RBM41 RRM (Y312/Q352/F354; Supplementary Figure S1) suggests that RBM41 may employ a similar mode of RNA binding as utilized by the U1A/U2B″/SNF proteins.

The homology between RBM41 and U11/U12-65K raises the question of whether RBM41 also functions in the minor spliceosome. We carried out a phylogenetic profiling-based co-evolution analysis in 167 eukaryotic species to identify proteins that show a similar phylogenetic presence/absence profile as RBM41 and may therefore function in the same molecular process (Figure 1E). Notably, three minor spliceosome-specific proteins, CENATAC, U11-48K (*SNRP48*) and U11/U12-31K (*ZCRB1*) were found

among the proteins showing the strongest co-occurrence with RBM41 (Supplementary Table S3, Supplementary Figure S2), suggesting a role in the minor spliceosome. This was further supported by earlier reciprocal coevolution analysis for CENATAC (32), which similarly identified RBM41 as one of the top co-occurring proteins among the known minor spliceosome components. Upon comparing RBM41's phylogenetic profile with that of other known minor spliceosome proteins (Figure 1E, Supplementary Figure S2), it becomes evident that RBM41 adheres to the typical presence/absence profile of these proteins. However, the presence of RBM41 appears to be even less common than that of other minor spliceosome proteins.

In contrast to the homologous C-terminal RRMs, the N-termini of the two proteins do not share sequence similarity. This suggests a functional difference between RBM41 and U11/U12-65K, that uses the N-terminus to promote the formation of the U11/U12 di-snRNP via an interaction with the U11-59K protein (22; Figure 1A). RBM41 N-terminus is highly conserved among animal orthologs (Supplementary Figure S3, Supplementary Table S2, Figure 1D) but lacks annotated domains and adopts a predominantly helical conformation in an AlphaFold prediction (57) (Figure 1D). Together the sequence analysis suggests similar RNA binding properties, but otherwise divergent functions for the two paralogous proteins.

## RBM41 RRM interacts with the U12 and U6atac snRNAs in vitro

Previously, two independent systematic high-throughput studies of RNA-binding protein specificities (58,59) reported nearly identical consensus RNA motifs for RBM41 (Figure 2A) which are preferentially located in a loop sequence within an RNA stem-loop context. These RNA motifs bear a striking similarity to the loop sequences of the U12 snRNA 3′-terminal stem-loop and U6atac snRNA 3′-terminal stem-loop, both of which are bound by the U11/U12-65K protein (Figure 2A,B; 22,27). Given the apparent RNA-binding similarities between the RBM41 and U11/U12-65K proteins, we compared their RNA-binding characteristics *in vitro*, using recombinant C-terminal RRMs containing additional N- and C-terminal regions known to be essential for RNA binding (25,26). RNA binding properties were analyzed using electrophoretic mobility shift assay (EMSA) with untagged RRMs and RNA oligonucleotides corresponding to the apical hairpins of the U12 snRNA and U6atac 3′-terminal stem-loops (Figure 2C, D; 22,25,26).

Our EMSA analyses revealed that U11/U12-65K C-RRM and RBM41 RRM have similar overall binding characteristics, binding to both U12 and U6atac hairpins (Figure 2C, compare lanes 1–12 and 13–23). In contrast, no complex formation was observed with either RRM when a control hairpin (complementary to the U12 hairpin) was used as a ligand (Figure 2C, lanes 24 and 25). A further determination of the dissociation constants ($K_d$) revealed that the RBM41 RRM has a low micromolar affinity to both U12 ($K_d = 2.14$ μM) and U6atac hairpins ($K_d = 3.78$ μM; Figure 2D). The 65K C-RRM showed approximately 2-fold higher affinity to both hairpins (U12: $K_d = 1.04$ μM; U6atac: $K_d = 2.07$ μM). Furthermore, both RRMs bound the U12 hairpin with ∼2-fold higher affinity compared to the U6atac hairpin (Figure 2D). Thus, both RBM41 and U11/U12-65K C-terminal RRMs show dual snRNA binding specificity *in vitro* with only slight differences in RNA affinity.

## RBM41 specifically associates with minor spliceosomal snRNPs

We next asked if RBM41 also associates with spliceosomal snRNAs *in vivo*. We transfected HEK293 cells with V5-tagged 65K or RBM41 followed by RNA immunoprecipitations (RIP) with anti-V5 antibody and northern blot analysis of minor and major spliceosomal snRNAs (Figure 3A). Consistent with its function in the U11/U12 di-snRNP, V5-65K strongly co-immunoprecipitated the U11 and U12 snRNAs and to a lesser extent U6atac and U4atac snRNAs (Figure 3A, lane 8). In contrast, V5-RBM41 co-immunoprecipitated the U12 snRNA, U4atac and U6atac snRNAs, but not the U11 snRNA (Figure 3A, lane 7). Similar results were obtained with an antibody against the endogenous RBM41 in HeLa nuclear extract (Figure 3B, lanes 1–4) or HEK293 total cell lysate (Figure 3B, lanes 5–8), showing that these interactions are not artifacts resulting from RBM41 overexpression. U1, U2, U4, U5 and U6 snRNAs were variably detected slightly above control IP levels with transiently overexpressed V5-RBM41 (Figure 3A,D) and endogenous RBM41 (Figure 3B), likely due to non-specific association of RBM41 with these highly abundant snRNPs (see below).

Next, we tested the role of the RBM41 RRM and the large N-terminal region lacking any identifiable domains for their association with snRNPs. We carried out RIP experiments using wild-type V5-RBM41, V5-RBM41 lacking the RRM domain (V5-RBM41(1–258)) or the entire N-terminal region (V5-RBM41(259–413)), as well as V5-RBM41 constructs with alanine substitutions of Gln352 or Phe354 of the conserved YQF triad in the RRM (V5-RBM41-Q352A and V5-RBM41-F354A) (Figure 3C). The Q352A and F354A mutations and deletion of the RRM led to dramatic loss of the U12, U4atac and U6atac interactions (Figure 3D, lanes 8–11), while the V5-RBM41(259–413) construct still showed robust co-immunoprecipitation of all three snRNAs (Figure 3D, lane 12). Notably, while major spliceosome-specific snRNAs (U1, U2, U4, U6) and the shared U5 snRNA were also detected above control IP levels in the V5-RBM41 anti-V5 IP (Figure 3D, lane 8), these were unaffected by the RRM mutations or the truncations, indicating that these IP signals represent non-specific background rather than specific interactions. Taken together, our RIP experiments show that RBM41 specifically associates with minor spliceosomal snRNPs in the cell and suggest that the snRNP association of RBM41 is primarily mediated through the RRM binding with its target snRNAs.

## RBM41 and U11/U12-65K partition into distinct snRNP complexes

While our *in vitro* binding experiments demonstrated similar RNA-binding properties for RBM41 and U11/U12-65K, the distinct snRNA IP profiles and the lack of sequence similarity outside the RRM suggested functional divergence of the two proteins within the minor spliceosome. As a complementary method to study snRNP complex association of RBM41 and U11/U12-65K, we carried out glycerol gradient fractionation of HeLa nuclear extract and analyzed the sedimentation behavior of the two proteins by western blot and spliceosomal snRNAs by northern blot (Figure 4A). Consistent with our RIP experiments, RBM41 and U11/U12-65K were largely found in different fractions, indicating association with different molecular complexes. RBM41 peak co-migrated with the U12 mono-snRNP (Figure 4A, fractions 8–10), whereas the U11/U12-65K peak co-migrated with the U11/U12 di-snRNP (fractions 12–14). Though the U4atac or U6atac snRNAs did not show clear co-migration with RBM41, a minor fraction of these snRNAs was present in RBM41 peak gradient fractions. Taken together with our RIP data, while the two proteins share similar RNA binding specificity *in vitro*, in the cell they partition into distinct snRNP complexes.

To define the composition and interactions of the RBM41 in the U12 mono-snRNP related to the 65K in U11/U12 di-snRNP, we used proximity-based labeling (BioID) to map the proximity interactors of the two proteins. A panel of RBM41 constructs and a U11/U12-65K construct (Figure 4B), each carrying an N-terminal MAC-tag (consisting of BirA* biotin ligase, hemagglutinin (HA) and StrepIII tags), were integrated into the Flp-In™ T-REx™ 293 cell line, enabling both inducible transgene control (tetracycline) and inducible biotinylation of proteins coming into proximity with the bait protein. Wild-type U11/U12-65K, wild-type RBM41, four RNA binding deficient RBM41 mutants (Y312A, Q352A, F354A, Y312A + F354A) and two RBM41 truncation constructs (1–258 and 259–413) were analyzed (Figure 4B). Western blot analysis with anti-HA antibody confirmed correct expression of MAC-tagged bait proteins (Supplementary Figure S4A), while biotinylation was visualized with avidin-HRP
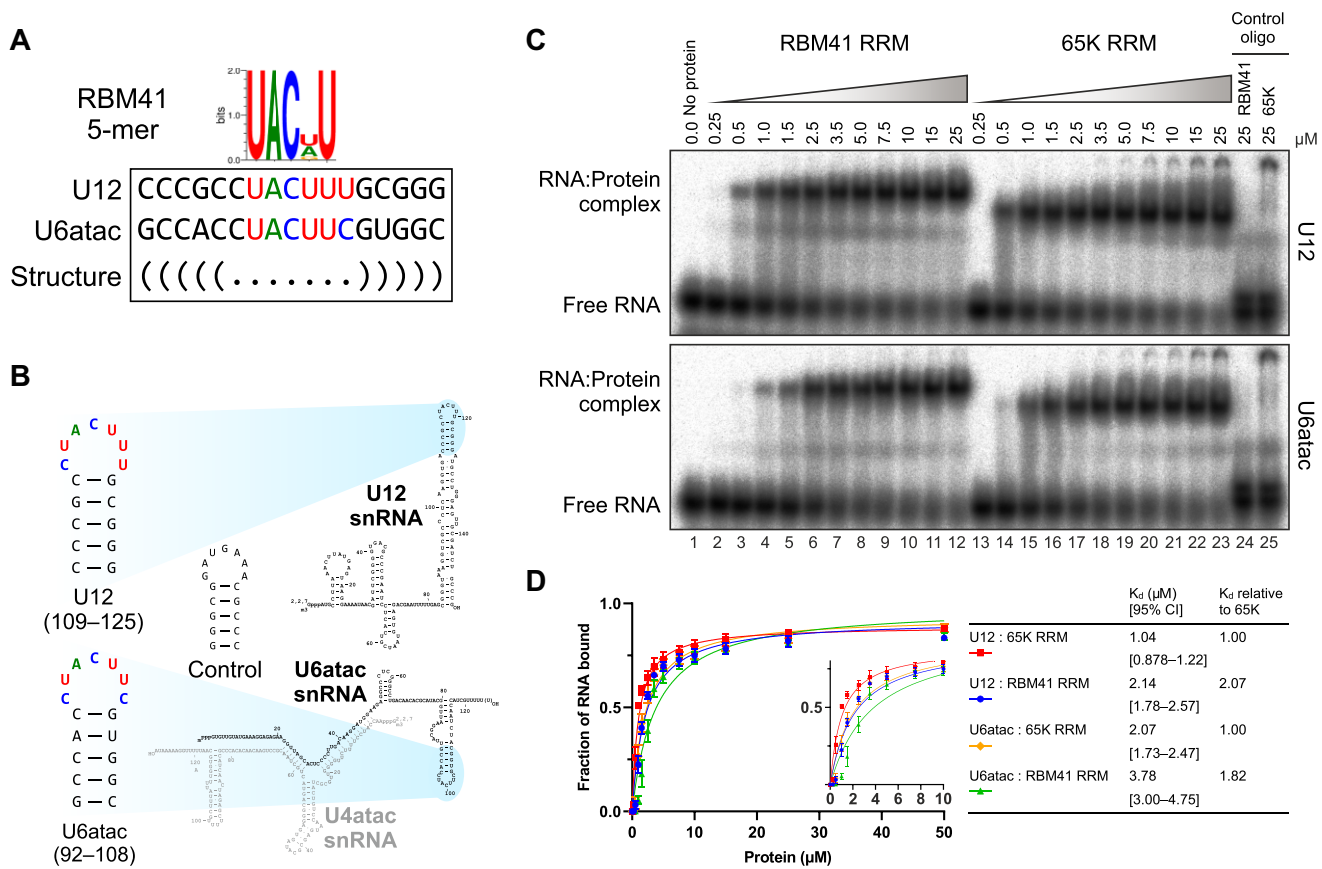
**Figure 2.** RBM41 interacts with the U12 and U6atac snRNAs *in vitro*. (**A**) Consensus RNA motifs bound by RBM41 *in vitro* and matching sequences in the U12 and U6atac snRNAs. The consensus motif (obtained from ENCODE database (81), entry ENCSR637HFY) determined by Ray *et al.* (58) using the RNAcompete method is shown. (**B**) RNA hairpins used in EMSA experiments and their location in the U12 and U6atac snRNAs. (**C**) EMSA analysis of RBM41 and U11/U12-65K RRM binding to U12 (top panel) and U6atac snRNA (bottom panel) hairpins. EMSA was carried out using recombinant RBM41 RRM (residues 267–413) or 65K C-terminal RRM (residues 380–517) and $^{32}$P-labeled U12, U6atac or negative control RNA hairpins shown in panel B. (**D**) Binding curves and dissociation constants for the interaction of RBM41 and 65K RRMs with U12 and U6atac hairpins. The inset shows a low protein concentration range (0–10 μM) of the same binding curves.

(Supplementary Figure S4B). Immunofluorescence with anti-HA was used to detect localization of the MAC-tagged proteins (Supplementary Figure S5A). MAC-RBM41-WT and MAC-U11/U12-65K proteins showed a predominantly nuclear localization, with MAC-RBM41-WT showing a more prominent cytoplasmic subpopulation. In contrast, MAC-RBM41-(1–258) was almost uniformly distributed between cytoplasm and nucleus, suggesting that deletion of the C-terminal RRM impairs nuclear import of RBM41.

MAC-tagged protein expression was induced with tetracycline in the presence of biotin for 24 h before harvesting of cells. Three independent replicates were analyzed for each cell line using an established BioID pipeline that has been described in detail (34,38). Consistent with its role in U11/U12 di-snRNP, MAC-U11/U12-65K interacted with all 8 minor spliceosome-specific proteins of the U11/U12 di-snRNP complex (Figure 4C, Supplementary Table S4), with its top interactors being U11-59K (PDCD7), U11/U12-20K (ZMAT5), U11-48K (SNRNP48) and U11-35K (SNRNP35). In contrast, MAC-RBM41-WT only interacted with one of the U11/U12 di-snRNP proteins, U11/U12-31K (ZCRB1) (Figure 4C, Supplementary Table S4). This interaction was dependent on a functional RBM41–U12 snRNA interaction as the construct encoding the RRM-only version of MAC-RBM41-(259–413) was still able to support the in-

teraction with U11/U12-31K while the Y312A, F354A and Y312A + F354A RRM mutants and the RRM deletion construct (MAC-RBM41-(1–258)) did not. The Q352A mutant likely retains some residual binding for U12 snRNP, as the BioID signal with U11/U12-31K was reduced but not completely eliminated (see also Figure 5E). Additionally, the loss of the RBM41 N-terminal domain resulted in a strong interaction with ZRSR2/1, which functions in 3′ss recognition of U12-type introns.

A further support for the direct or indirect interaction between RBM41 and U11/U12-31K within the U12 mono-snRNP is obtained from glycerol gradient centrifugation which shows that RBM41 and U11/U12-31K proteins cosediment in the same U12 mono-snRNP fractions (Figure 4A, lanes 8–9). Similarly, co-IP experiments show that the endogenous U11/U12-31K is preferentially associated with U12 snRNA, unlike the U11/U12-65K which shows an even co-IP efficiency for both U11 and U12 snRNAs (Figure 4D, cf. lanes 3 and 6). Association of the U11/U12-31K with the mono-snRNP is consistent with direct recognition of 2′-O-methylated A8 residue of the U12 snRNA (60). Conversely, the inverted co-IP pattern with U11-48K and U11-59K is consistent with their role as components of both U11 mono-snRNP and the U11/U12 di-snRNP (Figure 4D, lanes 4 and 5). Together, our RIP, glycerol gradient and BioID data show that RBM41
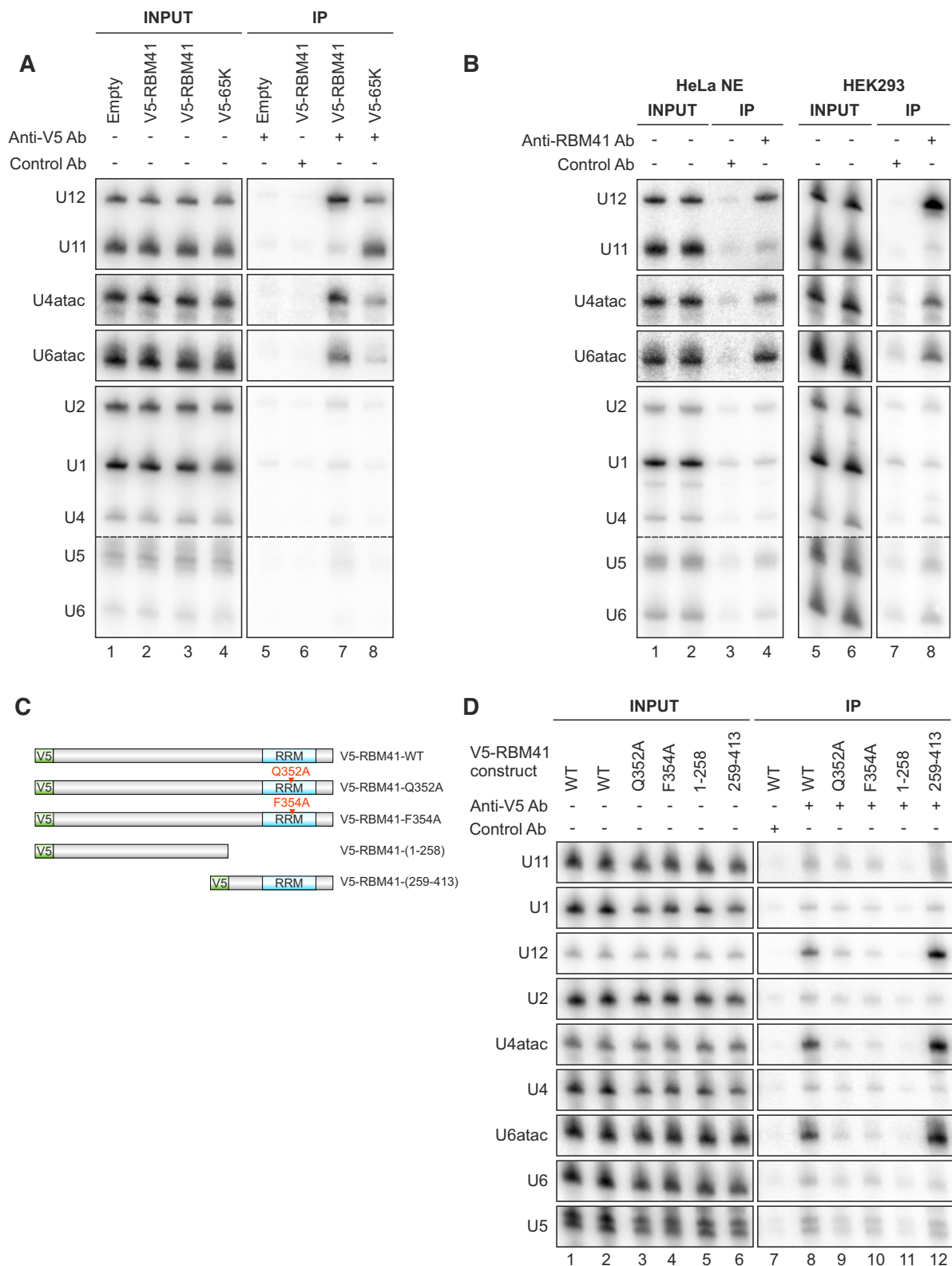
**Figure 3.** RBM41 specifically associates with minor spliceosomal snRNPs. (**A**) RNA immunoprecipitation with V5-tagged RBM41 and 65K. V5-RBM41 or V5-65K expression vector or empty vector were transfected into HEK293 cells. 24 h later, RNA immunoprecipitation with anti-V5 antibody or control antibody was carried out in native conditions and co-immunoprecipitated RNA analyzed by northern blot using the indicated probes. (**B**) RNA immunoprecipitation with endogenous RBM41. RIP was carried out in native conditions in either HeLa nuclear extract (left) or HEK293 total lysate (right) using an antibody against endogenous RBM41 or control antibody. (**C**) V5-RBM41 constructs used for RNA immunoprecipitation in panel D. (**D**) Effect of truncations and RRM mutations on the snRNP association of RBM41.V5-tagged RBM41 constructs shown in *C* were transfected into HEK293 cells and RNA immunoprecipitation carried out using anti-V5 or control antibody.
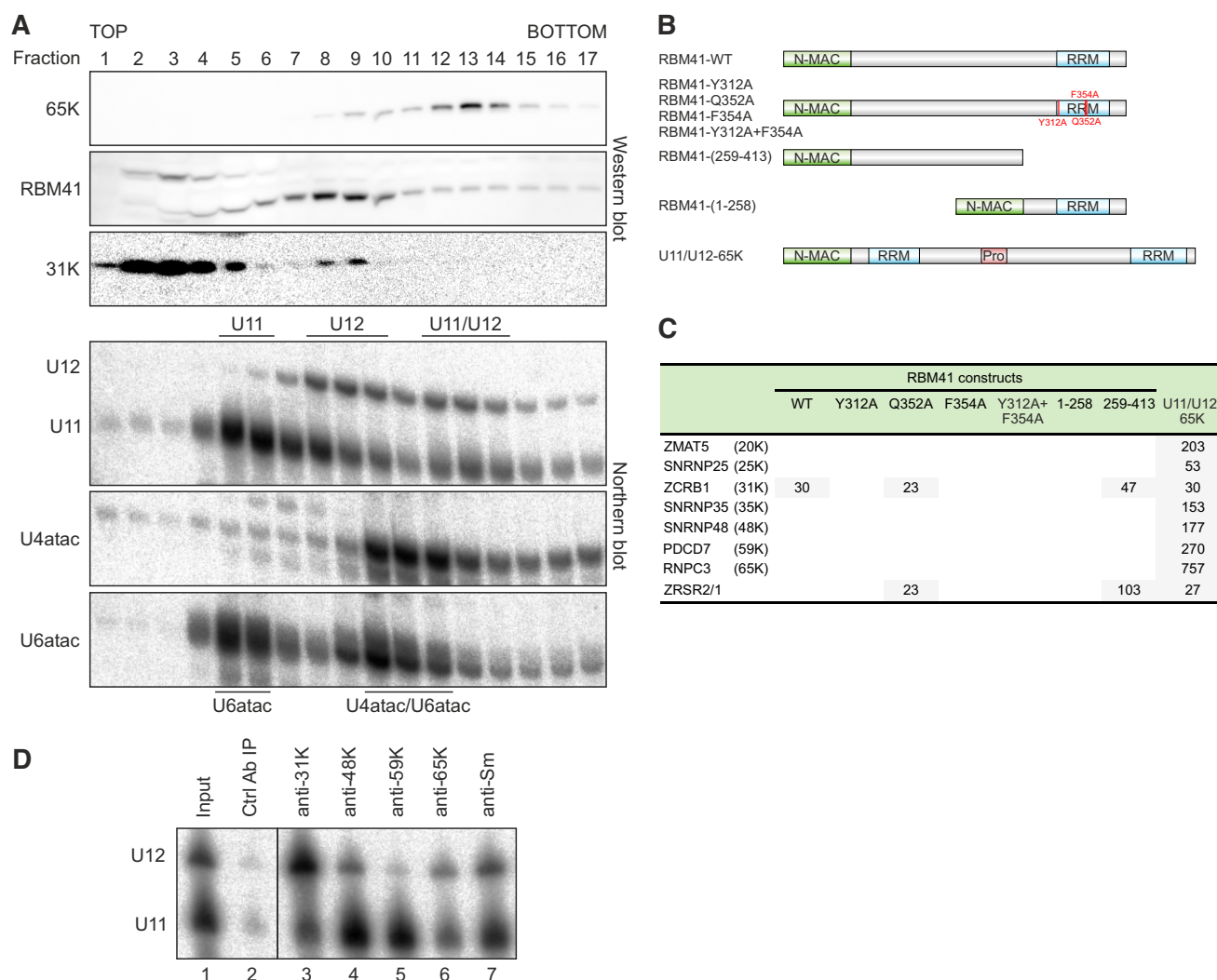
**Figure 4.** RBM41 and U11/U12-65K partition into distinct snRNP complexes. (**A**) Glycerol gradient analysis of RBM41 and U11/U12-65K in HeLa nuclear extract. Nuclear extract was loaded on top of a 10–30% glycerol gradient. After ultracentrifugation, the gradient was fractionated, protein and RNA isolated and analyzed by western and northern blot using the antibodies and probes indicated on the left. Location of the U11, U12 and U6atac mono-snRNPs, U11/U12 di-snRNP and U4atac/U6atac di-snRNP are inferred based on the snRNA profiles. (**B**) Domain structures of MAC-tagged RBM41 and 65K constructs used for BioID. N-terminal MAC tag is not drawn to scale. (**C**) Spectral count fold changes for U11/U12 di-snRNP proteins in BioID datasets. (**D**) Immunoprecipitation of U11 and U12 snRNAs by anti-31K, anti-48K, anti-59K, anti-65K and anti-Sm antibodies in HEK293 total lysate followed by Northern blot detection of the U11 and U12 snRNAs.

Table C:

| | RBM41 constructs | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | WT | Y312A | Q352A | F354A | Y312A+F354A | 1-258 | 259-413 | U11/U12-65K |
| ZMAT5 (20K) | | | | | | | | 203 |
| SNRNP25 (25K) | | | | | | | | 53 |
| ZCRB1 (31K) | 30 | | 23 | | | | 47 | 30 |
| SNRNP35 (35K) | | | | | | | | 153 |
| SNRNP48 (48K) | | | | | | | | 177 |
| PDCD7 (59K) | | | | | | | | 270 |
| RNPC3 (65K) | | | | | | | | 757 |
| ZRSR2/1 | | | 23 | | | | 103 | 27 |

and U11/U12-65K partition into distinct snRNP complexes and are likely to play distinct roles within the minor spliceosome.

## RBM41 interacts with DHX8 and localizes to Cajal bodies

One of the top BioID proximity-labeling hits of MAC-RBM41-WT was the DEAH-box RNA helicase DHX8 (Figure 5A, Supplementary Table S4). The interaction with DHX8 was dependent on the highly conserved N-terminal region of RBM41, while the RRM mutations and deletion of the RRM had no effect. In contrast, no interaction of U11/U12-65K with DHX8 was detected in BioID analysis. The DHX8:RBM41 interaction was also reported in a recently published reciprocal DHX8 BioID dataset (61). To validate these findings, we carried out co-immunoprecipitation experiments in Flp-In™ T-REx™ 293 cell lines expressing V5-RBM41 or V5-65K (Figure 5B). Consistent with BioID, V5-

RBM41 co-immunoprecipitated DHX8, while close to background levels of DHX8 were detected in V5-65K immunoprecipitates. Furthermore, both V5-RBM41 and V5-65K co-immunoprecipitated the U11/U12-31K protein at similar levels, consistent with U11/U12-31K associating with both the U12 mono-snRNP and the U11/U12 di-snRNP. Although not detected in BioID, Sm proteins were detected in both V5-RBM41 and V5-65K immunoprecipitates. In contrast, the U11 snRNP and U11/U12 di-snRNP-associated U11-48K and U11-59K proteins were only detected in V5-65K IPs.

DHX8 has not been shown to function in the minor spliceosome. In the major spliceosome, it is recruited before exon ligation, during the transition from C to C* complex (62), and drives the P-to-ILS1 transition presumably by pulling on the ligated exon, leading to its release and dissociation of at least nine proteins (63). We thus hypothesized that RBM41 could have a function in the late stages of minor splicing, possibly in the post-catalytic complexes. To test the association of RBM41 and DHX8 with late-stage minor spliceosomes, we
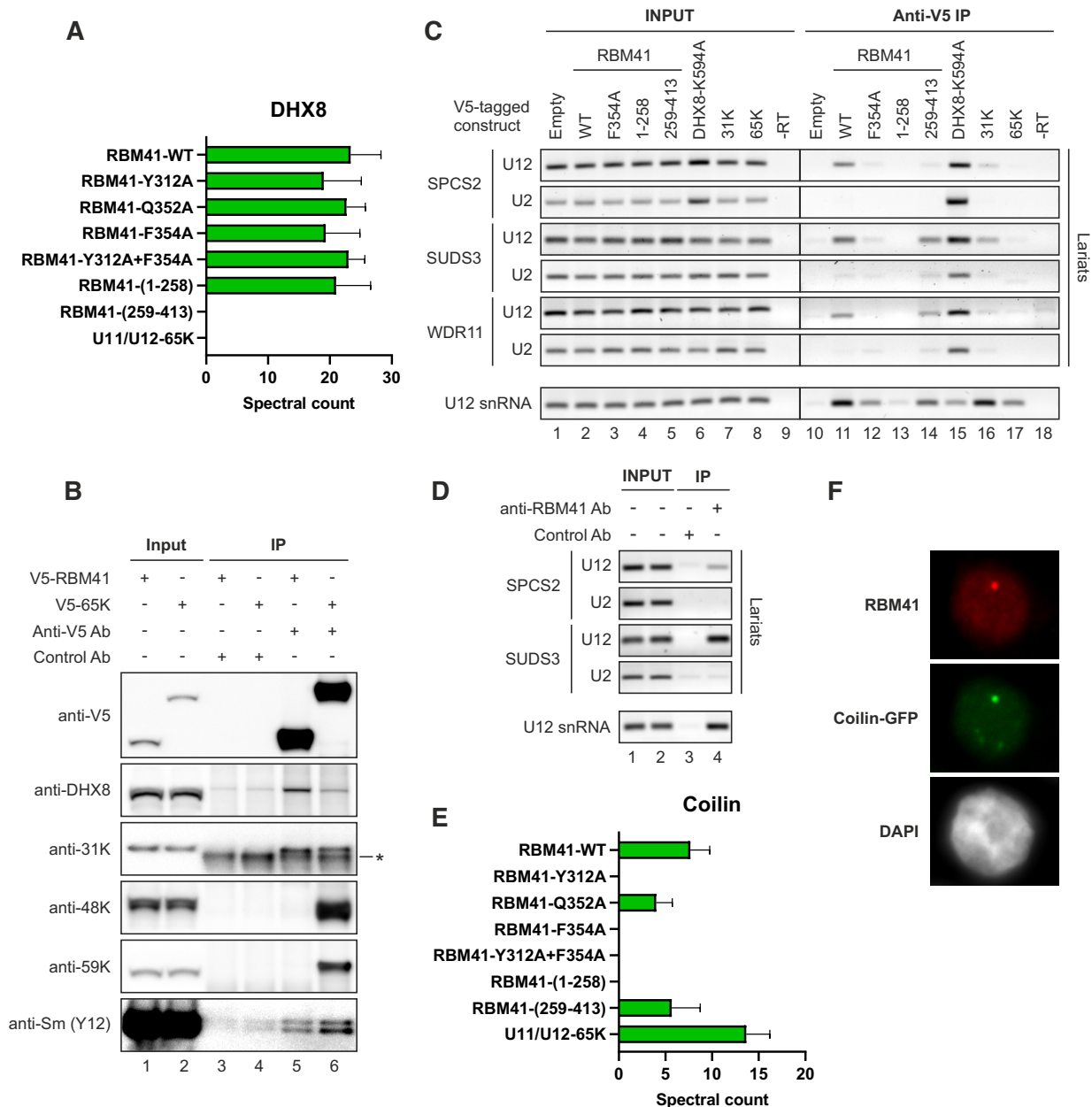
**Figure 5.** RBM41 interacts with DHX8 and localizes to Cajal bodies. (**A**) Spectral counts for DHX8 in RBM41 and U11/U12-65K BioID datasets. (**B**) Immunoprecipitation with anti-V5 or control antibody followed by western blot in Flp-In™ T-REx™ 293 cell lines expressing V5-RBM41 or V5-65K. The asterisk indicates a non-specific band detected in both control and anti-31K IPs and likely represents cross-reaction of the anti-rabbit secondary antibody with light chain from the IP antibody. (**C**) RNA immunoprecipitation with exogenously expressed V5-tagged proteins followed by RT-PCR. The indicated pCI-neo constructs for expressing V5-tagged proteins or empty pCI-neo vectors were transfected into HEK293 cells. 24 h later, RIP was carried out using anti-V5 antibody and RNA extracted from the beads analyzed by RT-PCR. Amplification across the branch junction was used to detect U2- and U12-type intron lariats and lariat intermediates from the following introns: SPCS2 introns 3–4 (U12) and 2–3 (U2), SUDS3 introns 7–8 (U12) and 9–10 (U2), WDR11 introns 28–29 (U12) and 27–28 (U2). (**D**) RNA immunoprecipitation with endogenous RBM41 in HEK293 cells followed by RT-PCR. (**E**) Spectral counts for coilin in RBM41 and U11/U12-65K BioID datasets. (**F**) Anti-RBM41 immunofluorescence in HEK293 cells transfected with a vector for expressing coilin-GFP.

carried out a RIP experiment in HEK293 cells transfected with V5-RBM41, V5-RBM41 mutant and truncation constructs (F354A, 1–258 and 259–413), or V5-tagged DHX8 carrying a helicase mutation (K594A) that stalls splicing in the P complex stage (14,64; Figure 5C). V5-65K and V5-31K were analyzed as controls for U11/U12 proteins. As DHX8 associates with post-branching spliceosomes, we used RT-PCR across branch sites (65,66) to detect U2-type and U12-type lariat intermediates and excised intron lariats in the immunoprecipitates.

V5-DHX8-K594A co-immunoprecipitated both U2-type and U12-type lariats from the same genes (Figure 5C, lane 15), suggesting that DHX8 functions in both spliceosomes. In contrast, RBM41 preferentially co-immunoprecipitated U12-type but not U2-type intron lariats (Figure 5C, lane 11). Similar results were obtained when RIP was carried out using an antibody against the endogenous RBM41 protein (Figure 5D). While U11/U12-65K did not co-IP any lariats (Figure 5C, lane 17), the U11/U12-31K showed variable and low levels of

lariat co-IP (Figure 5C, lane 16), suggesting that RBM41 BioID hits for DHX8 and U11/U12-31K may be originating from two separate complexes.

Another major interactor of MAC-RBM41-WT in our BioID data was coilin (Figure 5E, Supplementary Table S4), a key scaffolding protein and widely used marker for Cajal bodies. Consistently, we found that endogenous RBM41 localizes to Cajal bodies, as shown by colocalization with coilin-GFP in HEK293 cells (Figure 5F). While U11/U12-65K also interacted with coilin, the nuclear bodies labeled by the anti-RBM41 antibody did not colocalize with endogenous U11/U12-65K (Supplementary Figure S5B). This suggests that U11/U12-65K enters Cajal bodies only transiently. The coilin:RBM41 interaction was dependent on U12 snRNA binding, as mutating or deleting the RBM41 RRM reduced or completely eliminated the interaction, while MAC-RBM41-(259–413) was still able to interact with coilin (Figure 5E). Similarly, anti-HA immunofluorescence staining in BioID cell lines detected nuclear bodies in cells expressing MAC-RBM41-WT, but not in cells expressing any of the RBM41 RRM mutants or MAC-RBM41-(1–258) (Supplementary Figure S5A). This suggests that RBM41 localizes to Cajal bodies in an RNA binding-dependent manner.

### RBM41 is not essential for cell viability but affects the splicing of a subset of U12-type introns

To assess the effect of RBM41 loss on the splicing of the U12-type introns, we generated several independent RBM41 full knockout lines with HEK293 cells using CRISPR-Cas9 editing. The loss of functional RBM41 loci in each of the three X chromosomes in HEK293 cells was confirmed by Sanger sequencing and western blot analyses (Figure 6A; Supplementary Figure S6). The knockout cells were not only viable, but the loss of RBM41 did not lead to any noticeable growth phenotypes. This is consistent with the earlier investigations on human essentialomes that consistently indicated that RBM41 locus is not essential for the cell viability (67,68).

To analyze the effects of RBM41 knockout on splicing we carried out RNAseq analysis of three independent knockout cell lines and matching unedited lines. Subsequent bioinformatics analysis concentrated on U12-type intron retention and alternative/cryptic splice site activation with the U12-type intron containing genes. Intron retention analysis revealed weak splicing defects in the knockout cell lines for a small subset of genes (14 genes), such as the *NOL11* (Figure 6C; Supplementary Table S6), but also identified 15 genes that instead showed the opposite, that is, a reduction in the read levels mapping to the U12-type introns (Supplementary Table S6). Additionally, we investigated alternative splice donor (AD), splice acceptor (AA) and core exon (CE) usage. We further focused on the events within the U12-type introns and the surrounding exons and introns as these are the potential direct targets of RBM41 knockdown. We detected a total of 37 statistically significant alternative splicing events in 26 genes, as several genes showed multiple AS events being activated as a result of RBM41 knockout (Supplementary Table S7). Of these, the most notable were the ∼3-fold enrichment in alternative 3′ss (AA) usage (Figures 6B–D; $P = 3.4 \times 10^{-6}$, hypergeometric test) and the 1.7-fold reduction in core exon (CE) events (Figures 6B–D; $P = 1.2 \times 10^{-4}$, hypergeometric test) when compared to the alternative splicing events detected in genes containing only major introns. Notably, most of the identified alternative 3′ss events (13/18) affected the U12-type intron 3′ss choice (Supplementary Table S6), suggesting that the loss of RBM41 has a weak, but nevertheless statistically significant effect on the splicing of a subset of U12-type introns.

## Discussion

In this work, we have expanded the repertoire of unique protein components specific to minor spliceosome by providing evidence that RBM41 functions in post-splicing steps of the minor spliceosome assembly/disassembly cycle. RBM41 shows a similar phylogenetic co-evolution pattern as several other minor spliceosome components (Figure 1E) and it has earlier been annotated as a paralog of the U11/U12-65K protein, due to the highly similar C-terminal RRMs found in the two proteins. Here, we show that the C-terminal RRM of RBM41 binds to the 3′-terminal stem-loops of U12 and U6atac snRNAs both *in vitro* and *in vivo*. Compared to the U11/U12-65K C-terminal RRM, RBM41 has approximately 2× lower affinity to its RNA ligands. We further show that unlike U11/U12-65K, which is a component of the U11/U12 di-snRNP, RBM41 associates with a distinct U12 mono-snRNP. Both U12 mono-snRNP and U11/U12 di-snRNP complexes have been described previously (11,20,69), but the function or composition of the U12 mono-snRNP has not been studied further. Here, our ultracentrifugation and BioID analysis provides evidence that the U12 mono-snRNP is a distinct functional complex in the minor spliceosome and contains, in addition to RBM41, the U11/U12-31K (ZCRB1) protein as the specific protein components. Additionally, we show that RBM41 associates specifically with excised U12-type intron lariats and uses its unique N-terminal domain to interact with the DHX8 helicase, and likely cycles through the Cajal bodies. Together, our data suggests that the two paralogous proteins have distinct functions in U12-type intron splicing with U11/U12-65K functioning in the early steps of U12-type intron recognition and RBM41 in the post-splicing steps and during minor spliceosome disassembly (Figure 7).

Our results highlight the role of the 3′ stem-loop of U12 snRNA in the minor spliceosome assembly-disassembly cycle. The significance of the 5′ end of the U12 snRNA has long been recognized due to its function in the BPS recognition and the interactions with the U6atac snRNA in the catalytic core of the minor spliceosome (70). In contrast, the 3′-terminal stem-loop of the U12 snRNA has appeared as a static binding site for the U11/U12-65K protein, necessary for the formation of the U11/U12 di-snRNP. Our identification of RBM41 binding to the 3′-terminal stem–loop during minor spliceosome disassembly suggests more dynamic recognition events where the 3′-terminal stem-loop serves as a platform for the two paralogous proteins which guide the U12 snRNA though the minor spliceosome assembly and disassembly cycle. The previously characterized steps include the recognition of the 3′-terminal stem–loop of the U12 snRNA by the U11/U12-65K protein, which uses its N-terminus to interact with the U11-59K protein (22) to form the U11/U12 di-snRNP, which in turn is necessary for the U12-type intron recognition (12). Furthermore, during the formation of the catalytically active spliceosome ($B^{act}$ complex) the U11/U12-65K protein remain attached to the $B^{act}$ complex (presumably to the 3′-terminal stem-loop), while U11 snRNP and all the other specific protein components of the di-snRNP are released from the
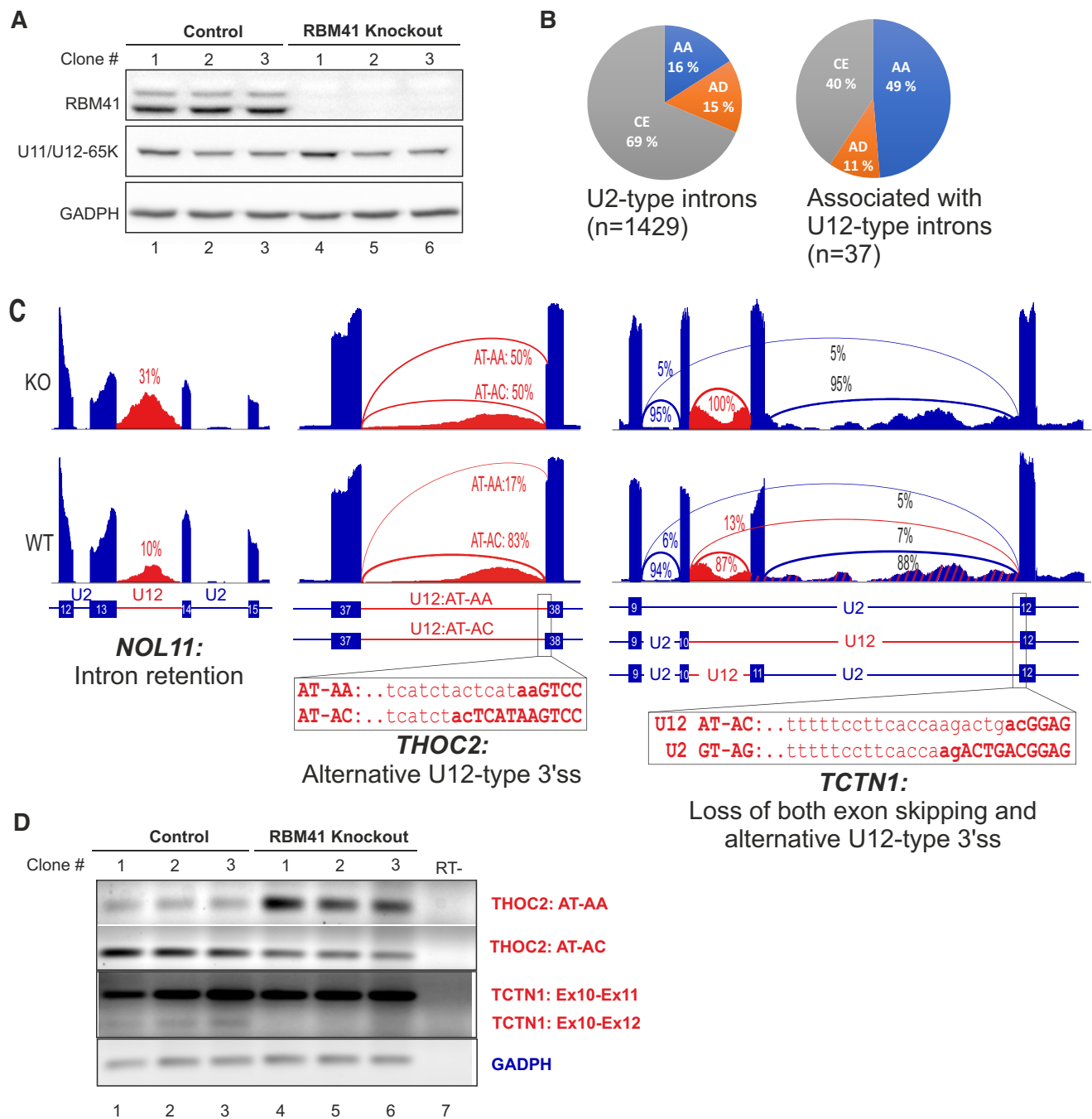
**Figure 6.** RBM41 knockout influences the splicing of U12-type introns. (**A**) Western blot analysis of RBM41 knockout and matching control cell lines used in the RNAseq analysis. (**B**) Comparison of the statistically significant (Whippet Probability > 0.9) alternative splicing events in the genes containing only U2-type introns and events either within or near proximity (immediate up- or downstream exons and introns) of the U12-type introns. AA - alternative acceptor, AD – alternative donor, CE – core exon. (**C**) Representative sashimi plots showing Intron retention (*NOL11*), Alternative U12-type 3'ss choice (*THOC2*) and loss of both exon skipping and alternative U12-type 3'ss usage in RBM41 knockout cells (*TCTN1*). The percentages refer to the intron retention levels (*NOL11*), the alternative 3' splice usage levels (*THOC2*) or exon skipping levels (*TCTN1*) as indicated by the arches in the Sashimi plot. (**D**) Validation of the *THOC2* and *TCTN1* alternative splicing changes using a set of three independent RBM41 knockout cell lines and their matching controls.

activated spliceosome (8). Our data indicate that later in the splicing process there is an exchange in the 3′-terminal stem-loop binding partner from U11/U12-65K to RBM41 which can be detected in post-splicing complexes containing excised minor intron lariats, and which is also in close proximity with the DHX8/hPrp22 helicase. Together these results suggest that RBM41 is present in the minor spliceosome post catalytic (P) and intron lariat spliceosome (ILS) complexes. Furthermore, proximity labeling of Cajal body marker coilin (Figure 5E; (61) and co-localization of RBM41 and coilin in Cajal bodies (Figure 5F, Supplementary Figure S5) suggests that RBM41 remains bound to the U12 snRNA during the entire spliceosome disassembly process and follows the snRNA to the Cajal body. This is the likely location for a new round of assembly of the U11/U12 di-snRNP complex. The association of U11/U12-31K with both the U12 mono-snRNP and
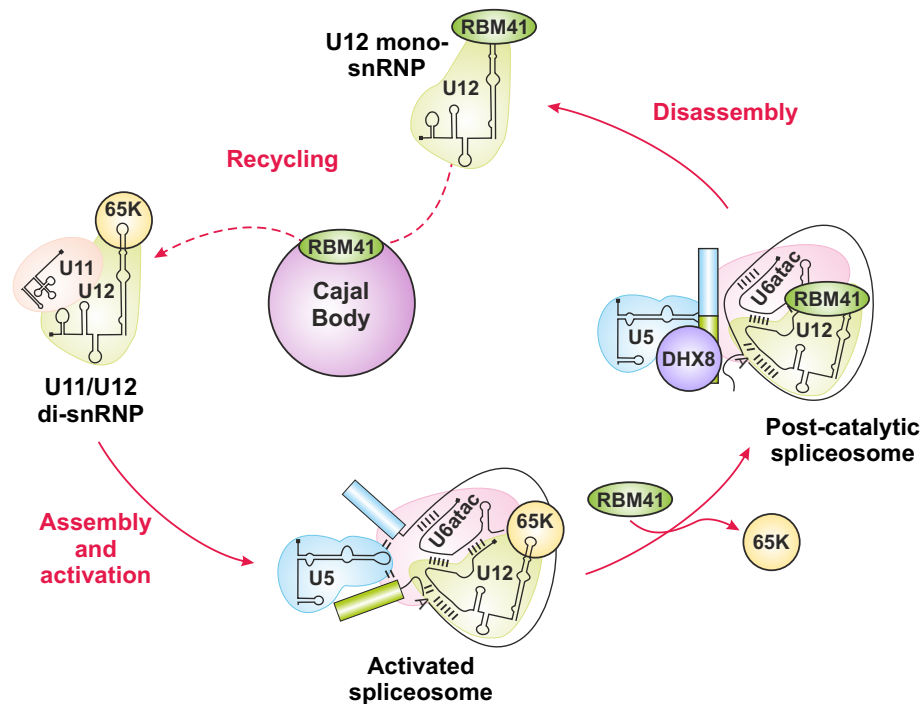
**Figure 7.** A model of the dynamic exchanges between RBM41 and 65K binding to U12 snRNA during the splicing cycle. In the U11/U12 di-snRNP, the 3′-terminal stem-loop is bound by U11/U12-65K, which mediates the connection between the U11 and U12 snRNPs. U11/U12-65K likely remains bound to the stem-loop throughout minor spliceosome assembly and activation but is exchanged for RBM41 during or after the catalytic steps of splicing. After spliceosome disassembly, RBM41 remains bound to the post-spliceosomal U12 mono-snRNP and accompanies it to Cajal bodies. During U11/U12 di-snRNP recycling, which likely takes place in Cajal bodies, RBM41 is again replaced by U11/U12-65K at the 3′-terminal stem-loop.

U11/U12 di-snRNP (Figures 4A, C, D) suggests that it remains bound with U12 snRNA (60) in the new round of U11/U12 di-snRNP assembly initiated by the exchange from RBM41 to U11/U12-65K.

RBM41 and U11/U12-65K proteins interact with both U12 and U6atac 3′ terminal stem-loops *in vitro* and *in vivo*. However, the U12 interactions appear more significant, given the sensitivity of U12-type intron splicing to mutations in the U12 single-stranded loop and the insensitivity to U6atac loop mutations (27,71). Furthermore, an 84C > U mutation that compromises the U12 3′-terminal stem-loop integrity leads to early onset cerebellar ataxia due to overtrimming of the 3′-terminal stem-loop which removes the binding site of the U11/U12-65K and RBM41 proteins (36,72). Similarly, the U11/U12-65K P474T mutation associated with isolated growth hormone deficiency has been shown to reduce the U11/U12 di-snRNP levels due to a folding defect of the U11/U12-65K C-RRM, which reduces its affinity to the 3′-terminal stem-loop (24,25). However, in that case the potential additional effects on U6atac binding *in vivo* or on the recycling of the U12 snRNA have not been ruled out.

Our data portrays a somewhat conflicting view on the significance of the RBM41 and the need for a specific protein factor(s) for the minor spliceosome disassembly process. The strong sequence conservation observed with the domains of RBM41 that interact either with the U12 snRNA or DHX8 (Supplementary Figure S2) suggests a strong selection pressure at the organismal level to maintain these interactions. This is reflected in phylogenetic co-occurrence in multiple evolutionary lineages (Figure 1E), though we note that the locus encoding RBM41 is more frequently absent in multiple evolutionary lineages than several other components of the minor spliceo-

some. On the other hand, both our knockout data (Figure 6A) and data from essentialomes (67,68) indicate that RBM41 is dispensable at least at the cellular level. Based on the weak, yet statistically significant effects of RBM41 knockout in human HEK293 cells specifically on 3′ss selection of U12-type introns (Figure 6B–D), we hypothesize that while RBM41 is dispensable at the cellular level, it may nevertheless be able to exert a weak kinetic effect on splicing in addition to later participating in the disassembly process. The effect on 3′ss choice is similar to that observed after major spliceosome catalytic step II factor knockdowns, which similarly influence the 3′ss choice, particularly with NAGNAG introns (73), further suggesting that the exchange from the U11/U12-65K to RBM41 may take place prior to step II. However, given that our BioID analysis did not provide supporting evidence for this possibility, it is also possible that the exchange from U11/U12-65K to RBM41 takes place at a later step and the effects on minor intron splicing are secondary effects of downstream processes being disturbed. Finally, while RBM41 protein is not absolutely needed for cell viability or splicing in the highly proliferative cell types used in essentialome and our knockout studies, it may provide selective advantage in specific cell types or in the context of whole organisms to account for the observed evolutionary conservation.

RBM41 may also play a role in substituting structures or interactions that are present in the major but not in the minor spliceosome. Specifically, the human minor B$^{act}$ complex lacks several key protein components that are present in the major B$^{act}$ complex. These include NTC complex proteins (PRPF19, SPF27 and SYF1), NTR complex proteins (BUD31 and RBM22), SF3a complex, and phosphoprotein isomerases (PPIL1 and CypE). Conversely, the minor B$^{act}$

complex contains four unique proteins RBM48, ARMC7, SCNM1 and CRIPT that are not present in the major B$^{act}$ complex (8,30,31). At least a subset of these differences in protein composition can be explained by the presence of distinct snRNA components in the two spliceosomes which influences the availability of the specific RNA elements for the binding of protein factors. For example, U6atac snRNA lacks the 5′ stem-loop structure which is present in U6 snRNA and recognized by the BUD31 and RBM22 proteins in major B$^{act}$ complex. Instead, in minor B$^{act}$ these proteins are replaced by RBM48 and ARMC7 proteins binding to the γ-monomethyl cap of U6atac snRNA (8). Similarly, as the U11/U12 di-snRNP lacks the SF3a complex (19) that is present both in the 17S U2 snRNP and the major B$^{act}$ complex (74,75), the functions of SF3a in minor B$^{act}$ has been taken over by the SCNM1, a unique protein component of the minor spliceosome (8). In this respect the absence of CypE and SYF1 from the minor B$^{act}$ (8) is intriguing as those are loaded to major B$^{act}$ and are later involved in post-spliceosomal complex transitions (63,64). From the available limited structural information, we hypothesize that there may be similar differences in the post-spliceosomal complex composition between the two spliceosomes such that RBM41 may substitute some of the major spliceosome structures and /or interactions that are missing from the minor spliceosome. Testing of this hypothesis would require similar high-resolution structures of the post-splicing minor spliceosome complexes that are available for the major spliceosome.

Unique protein components in the minor spliceosome can also offer potential opportunities for differential regulation of the two spliceosomes. Given the function of RBM41 in the post-splicing complexes, the regulation of minor intron splicing via RBM41 abundance or activity is unlikely. However, earlier work has nevertheless provided evidence of translational regulation of RBM41 via differential 3′ UTR isoform usage (76), particularly during neuronal development (77). As the abundance of U11/U12-65K is also regulated by a feedback/cross-regulation pathway, particularly during neuronal differentiation (78,79), it is possible that the cellular levels of RBM41 protein are similarly linked to the abundance of minor spliceosome components or to global regulation of mRNA processing pathways.

## Data availability

The RNAseq data generated in this study is available at NCBI Gene Expression Omnibus (GEO) with accession number GSE246283 and the mass spectrometry data at MassIVE with dataset ID MSV000092983.

## Supplementary data

Supplementary Data are available at NAR Online.

## Acknowledgements

## Funding

## Conflict of interest statement

None declared.

## References

1. Hall,S.L. and Padgett,R.A. (1996) Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science*, **271**, 1716–1718.
2. Tarn,W.-Y. and Steitz,J.A. (1996) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science*, **273**, 1824–1832.
3. Tarn,W.-Y. and Steitz,J.A. (1996) A novel spliceosome containing U11, U12 and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell*, **84**, 801–811.
4. Kolossova,I. and Padgett,R.A. (1997) U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA*, **3**, 227–233.
5. Yu,Y.-T. and Steitz,J.A. (1997) Site-specific crosslinking of mammalian U11 and U6atac to the 5' splice site of an AT-AC intron. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 6030–6035.
6. Hastings,M.L. and Krainer,A.R. (2001) Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *RNA*, **7**, 471–482.
7. Wilkinson,M.E., Charenton,C. and Nagai,K. (2020) RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.*, **89**, 359–388.
8. Bai,R., Wan,R., Wang,L., Xu,K., Zhang,Q., Lei,J. and Shi,Y. (2021) Structure of the activated human minor spliceosome. *Science*, **371**, eabg0879.
9. Zhao,J., Peter,D., Brandina,I., Liu,X. and Galej,W.P. (2023) Structure of the minor spliceosomal U11 snRNP. bioRxiv doi: https://doi.org/10.1101/2023.12.22.573053, 22 December 2023, preprint: not peer reviewed.
10. Frilander,M.J. and Steitz,J.A. (2001) Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. *Mol. Cell*, **7**, 217–226.
11. Wassarman,K.M. and Steitz,J.A. (1992) The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. *Mol. Cell Biol.*, **12**, 1276–1285.
12. Frilander,M.J. and Steitz,J.A. (1999) Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. *Genes Dev.*, **13**, 851–863.
13. Company,M., Arenas,J. and Abelson,J. (1991) Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature*, **349**, 487–493.
14. Ohno,M. and Shimura,Y. (1996) A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. *Genes Dev.*, **10**, 997–1007.
15. Arenas,J.E. and Abelson,J.N. (1997) Prp43: an RNA helicase-like factor involved in spliceosome disassembly. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 11798–11802.

16. Staněk,D. (2017) Cajal bodies and snRNPs - friends with benefits. *RNA Biol*, **14**, 671–679.

17. Damianov,A., Schreiner,S. and Bindereif,A. (2004) Recycling of the U12-type spliceosome requires p110, a component of the U6atac snRNP. *Mol. Cell. Biol*, **24**, 1700–1708.

18. Bell,M., Schreiner,S., Damianov,A., Reddy,R. and Bindereif,A. (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.*, **21**, 2724–2735.

19. Will,C.L., Schneider,C., Reed,R. and Lührmann,R. (1999) Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science*, **284**, 2003–2005.

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

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

22. Benecke,H., Lührmann,R. and Will,C.L. (2005) The U11/U12 snRNP 65K protein acts as a molecular bridge, binding the U12 snRNA and U11-59K protein. *EMBO J.*, **24**, 3057–3069.

23. Turunen,J.J., Will,C.L., Grote,M., Lührmann,R. and Frilander,M.J. (2008) The U11-48K protein contacts the 5' splice site of U12-type introns and the U11-59K protein. *Mol. Cell. Biol*, **28**, 3548–3560.

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

25. Norppa,A.J., Kauppala,T., Heikkinen,H.A., Verma,B., Iwai,H. and Frilander,M.J. (2018) Mutations in the U11/U12-65K protein associated with isolated growth hormone deficiency lead to structural destabilization and impaired binding of U12 snRNA. *RNA*, **24**, 396–409.

26. Netter,C., Weber,G., Benecke,H. and Wahl,M.C. (2009) Functional stabilization of an RNA recognition motif by a noncanonical N-terminal expansion. *RNA*, **15**, 1305–1313.

27. Singh,J., Sikand,K., Conrad,H., Will,C.L., Komar,A.A. and Shukla,G.C. (2016) U6atac snRNA stem-loop interacts with U12 p65 RNA binding protein and is functionally interchangeable with the U12 apical stem-loop III. *Sci. Rep.*, **6**, 31393.

28. Madan,V., Kanojia,D., Li,J., Okamoto,R., Sato-Otsubo,A., Kohlmann,A., Sanada,M., Grossmann,V., Sundaresan,J., Shiraishi,Y., *et al.* (2015) Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat. Commun.*, **6**, 6042.

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

30. Bai,F., Corll,J., Shodja,D.N., Davenport,R., Feng,G., Mudunkothge,J., Brigolin,C.J., Martin,F., Spielbauer,G., Tseung,C.-W., *et al.* (2019) RNA binding motif protein 48 is required for U12 splicing and maize endosperm differentiation. *Plant Cell*, **31**, 715.

31. Siebert,A.E., Corll,J., Paige Gronevelt,J., Levine,L., Hobbs,L.M., Kenney,C., Powell,C.L.E., Battistuzzi,F.U., Davenport,R., Mark Settles,A., *et al.* (2022) Genetic analysis of human RNA binding motif protein 48 (RBM48) reveals an essential role in U12-type intron splicing. *Genetics*, **222**, iyac129.

32. de Wolf,B., Oghabian,A., Akinyi,M.V., Hanks,S., Tromer,E.C., van Hooff,J.J.E., van Voorthuijsen,L., van Rooijen,L.E., Verbeeren,J., Uijttewaal,E.C.H., *et al.* (2021) Chromosomal instability by mutations in the novel minor spliceosome component CENATAC. *EMBO J.*, **40**, e106536.

33. Suzuki,T., Shinagawa,T., Niwa,T., Akeda,H., Hashimoto,S., Tanaka,H., Hiroaki,Y., Yamasaki,F., Mishima,H., Kawai,T., *et al.* (2022) The DROL1 subunit of U5 snRNP in the spliceosome is specifically required to splice AT-AC-type introns in Arabidopsis. *Plant J.*, **109**, 633–648.

34. Liu,X., Salokas,K., Weldatsadik,R.G., Gawriyski,L. and Varjosalo,M. (2020) Combined proximity labeling and affinity purification-mass spectrometry workflow for mapping and visualizing protein interaction networks. *Nat. Protoc*, **15**, 3182–3211.

35. Bloh,K., Kanchana,R., Bialk,P., Banas,K., Zhang,Z., Yoo,B.C. and Kmiec,E.B. (2021) Deconvolution of complex DNA repair (DECODR): establishing a novel deconvolution algorithm for comprehensive analysis of CRISPR-edited Sanger sequencing data. *CRISPR J*, **4**, 120–131.

36. Norppa,A.J. and Frilander,M.J. (2021) The integrity of the U12 snRNA 3' stem–loop is necessary for its overall stability. *Nucleic Acids Res.*, **49**, 2835–2847.

37. Tarn,W.-Y. and Steitz,J.A. (1994) SR proteins can compensate for the loss of U1 snRNP functions *in vitro*. *Genes Dev.*, **8**, 2704–2717.

38. Liu,X., Salokas,K., Tamene,F., Jiu,Y., Weldatsadik,R.G., Öhman,T. and Varjosalo,M. (2018) An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat. Commun.*, **9**, 1188.

39. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2012) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

40. Sterne-Weiler,T., Weatheritt,R.J., Best,A.J., Ha,K.C.H. and Blencowe,B.J. (2018) Efficient and accurate quantitative profiling of alternative splicing patterns of any complexity on a laptop. *Mol. Cell*, **72**, 187–200.

41. Middleton,R., Gao,D., Thomas,A., Singh,B., Au,A., Wong,J.J.-L., Bomane,A., Cosson,B., Eyras,E., Rasko,J.E.J., *et al.* (2017) IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol.*, **18**, 51.

42. Lorenzi,C., Barriere,S., Arnold,K., Luco,R.F., Oldfield,A.J. and Ritchie,W. (2021) IRFinder-S: a comprehensive suite to discover and explore intron retention. *Genome Biol.*, **22**, 307.

43. Oghabian,A., Greco,D. and Frilander,M.J. (2018) IntEREst: intron-exon retention estimator. *BMC Bioinform.*, **19**, 130.

44. Moyer,D.C., Larue,G.E., Hershberger,C.E., Roy,S.W. and Padgett,R.A. (2020) Comprehensive database and evolutionary dynamics of U12-type introns. *Nucleic Acids Res.*, **48**, 7066–7078.

45. Olthof,A.M., Hyatt,K.C. and Kanadia,R.N. (2019) Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. *BMC Genom.*, **20**, 686.

46. Deutekom,E.S., Snel,B. and van Dam,T.J.P. (2021) Benchmarking orthology methods using phylogenetic patterns defined at the base of Eukaryotes. *Brief. Bioinform.*, **22**, bbaa206.

47. Vosseberg,J., van Hooff,J.J.E., Marcet-Houben,M., van Vlimmeren,A., van Wijk,L.M., Gabaldón,T. and Snel,B. (2021) Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat. Ecol. Evol.*, **5**, 92–100.

48. Emms,D.M. and Kelly,S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.*, **20**, 238.

49. Huerta-Cepas,J., Szklarczyk,D., Heller,D., Hernández-Plaza,A., Forslund,S.K., Cook,H., Mende,D.R., Letunic,I., Rattei,T., Jensen,L.J., *et al.* (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.*, **47**, D309–D314.

50. Mistry,J., Finn,R.D., Eddy,S.R., Bateman,A. and Punta,M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.*, **41**, e121.

51. Katoh,K., Kuma,K.-i., Toh,H. and Miyata,T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.*, **33**, 511–518.

52. Nguyen,L.-T., Schmidt,H.A., von Haeseler,A. and Minh,B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for

estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*, **32**, 268–274.

53. Vosseberg,J., Stolker,D., von der Dunk,S.H.A. and Snel,B. (2023) Integrating phylogenetics with intron positions illuminates the origin of the complex spliceosome. *Mol. Biol. Evol.*, **40**, msad011.

54. Muto,Y. and Yokoyama,S. (2012) Structural insight into RNA recognition motifs: versatile molecular Lego building blocks for biological systems. *WIREs RNA*, **3**, 229–246.

55. DeKoster,G.T., Delaney,K.J. and Hall,K.B. (2014) A compare-and-contrast NMR dynamics study of two related RRMs: U1A and SNF. *Biophys. J.*, **107**, 208–219.

56. Weber,G., DeKoster,G.T., Holton,N., Hall,K.B. and Wahl,M.C. (2018) Molecular principles underlying dual RNA specificity in the Drosophila SNF protein. *Nat. Commun.*, **9**, 2220.

57. Varadi,M., Anyango,S., Deshpande,M., Nair,S., Natassia,C., Yordanova,G., Yuan,D., Stroe,O., Wood,G., Laydon,A., *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.

58. Ray,D., Kazan,H., Cook,K.B., Weirauch,M.T., Najafabadi,H.S., Li,X., Gueroussov,S., Albu,M., Zheng,H., Yang,A., *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.

59. Dominguez,D., Freese,P., Alexis,M.S., Su,A., Hochman,M., Palden,T., Bazile,C., Lambert,N.J., Van Nostrand,E.L., Pratt,G.A., *et al.* (2018) Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell*, **70**, 854–867.

60. Li,B., Liu,S., Zheng,W., Liu,A., Yu,P., Wu,D., Zhou,J., Zhang,P., Liu,C., Lin,Q., *et al.* (2024) RIP-PEN-seq identifies a class of kink-turn RNAs as splicing regulators. *Nat. Biotechnol.*, **42**, 119–131.

61. Go,C.D., Knight,J.D.R., Rajasekharan,A., Rathod,B., Hesketh,G.G., Abe,K.T., Youn,J.Y., Samavarchi-Tehrani,P., Zhang,H., Zhu,L.Y., *et al.* (2021) A proximity-dependent biotinylation map of a human cell. *Nature*, **595**, 120–124.

62. Zhan,X., Yan,C., Zhang,X., Lei,J. and Shi,Y. (2018) Structure of a human catalytic step I spliceosome. *Science*, **359**, 537–545.

63. Zhang,X., Zhan,X., Yan,C., Zhang,W., Liu,D., Lei,J. and Shi,Y. (2019) Structures of the human spliceosomes before and after release of the ligated exon. *Cell Res.*, **29**, 274–285.

64. Fica,S.M., Oubridge,C., Wilkinson,M.E., Newman,A.J. and Nagai,K. (2019) A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science*, **363**, 710–714.

65. Lorsch,J.R., Bartel,D.P. and Szostak,J.W. (1995) Reverse transcriptase reads through a 2′–5′ linkage and a 2′-thiphosphate in a template. *Nucleic Acids Res.*, **23**, 2811–2814.

66. Suzuki,H., Zuo,Y., Wang,J., Zhang,M.Q., Malhotra,A. and Mayeda,A. (2006) Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res.*, **34**, e63.

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

68. Yilmaz,A., Peretz,M., Aharony,A., Sagi,I. and Benvenisty,N. (2018) Defining essential genes for human pluripotent stem cells by CRISPR–Cas9 screening in haploid cells. *Nat. Cell Biol*, **20**, 610–619.

69. Montzka,K.A. and Steitz,J.A. (1988) Additional low-abundance human small nuclear ribonucleoproteins: U11, U12 etc. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 8885–8889.

70. Turunen,J.J., Niemelä,E.H., Verma,B. and Frilander,M.J. (2013) The significant other: splicing by the minor spliceosome. *WIREs RNA*, **4**, 61–76.

71. Sikand,K. and Shukla,G.C. (2011) Functionally important structural elements of U12 snRNA. *Nucleic Acids Res.*, **39**, 8531–8543.

72. Elsaid,M.F., Chalhoub,N., Ben-Omran,T., Kumar,P., Kamel,H., Ibrahim,K., Mohamoud,Y., Al-Dous,E., Al-Azwani,I., Malek,J.A., *et al.* (2017) Mutation in noncoding RNA RNU12 causes early onset cerebellar ataxia. *Ann. Neurol.*, **81**, 68–78.

73. Dybkov,O., Preußner,M., El Ayoubi,L., Feng,V.-Y., Harnisch,C., Merz,K., Leupold,P., Yudichev,P., Agafonov,D.E., Will,C.L., *et al.* (2023) Regulation of 3′ splice site selection after step 1 of splicing by spliceosomal C* proteins. *Sci. Adv.*, **9**, eadf1785.

74. Bessonov,S., Anokhina,M., Krasauskas,A., Golas,M.M., Sander,B., Will,C.L., Urlaub,H., Stark,H. and Lührmann,R. (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA*, **16**, 2384–2403.

75. Schmidt,C., Grønborg,M., Deckert,J., Bessonov,S., Conrad,T., Lührmann,R. and Urlaub,H. (2014) Mass spectrometry–based relative quantification of proteins in precatalytic and catalytically active spliceosomes by metabolic labeling (SILAC), chemical labeling (iTRAQ), and label-free spectral count. *RNA*, **20**, 406–420.

76. Sterne-Weiler,T., Martinez-Nunez,R.T., Howard,J.M., Cvitovik,I., Katzman,S., Tariq,M.A., Pourmand,N. and Sanford,J.R. (2013) Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res.*, **23**, 1615–1623.

77. Yap,K., Xiao,Y., Friedman,B.A., Je,H.S. and Makeyev,E.V. (2016) Polarizing the neuron through sustained co-expression of alternatively spliced isoforms. *Cell Rep.*, **15**, 1316–1328.

78. Verbeeren,J., Niemelä,E.H., Turunen,J.J., Will,C.L., Ravantti,J.J., Lührmann,R. and Frilander,M.J. (2010) An ancient mechanism for splicing control: U11 snRNP as an activator of alternative splicing. *Mol. Cell*, **37**, 821–833.

79. Verbeeren,J., Verma,B., Niemelä,E.H., Yap,K., Makeyev,E.V. and Frilander,M.J. (2017) Alternative exon definition events control the choice between nuclear retention and cytoplasmic export of U11/U12-65K mRNA. *PLoS Genet.*, **13**, e1006824.

80. Robert,X. and Gouet,P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, W320–W324.

81. Luo,Y., Hitz,B.C., Gabdank,I., Hilton,J.A., Kagda,M.S., Lam,B., Myers,Z., Sud,P., Jou,J., Lin,K., *et al.* (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.*, **48**, D882–D889.