Nucleic Acids Research, 2024, **52**, 3291–3309
https://doi.org/10.1093/nar/gkad1243
Advance access publication date: 2 January 2024
RNA and RNA-protein complexes

THUMPD2 catalyzes the N^2-methylation of U6 snRNA of the spliceosome catalytic center and regulates pre-mRNA splicing and retinal degeneration

Wen-Qing Yang[1,†], Jian-Yang Ge[1,†], Xiaofeng Zhang[2], Wen-Yu Zhu[1], Lin Lin[1], Yigong Shi[3], Beisi Xu[4,*] and Ru-Juan Liu[1,*]

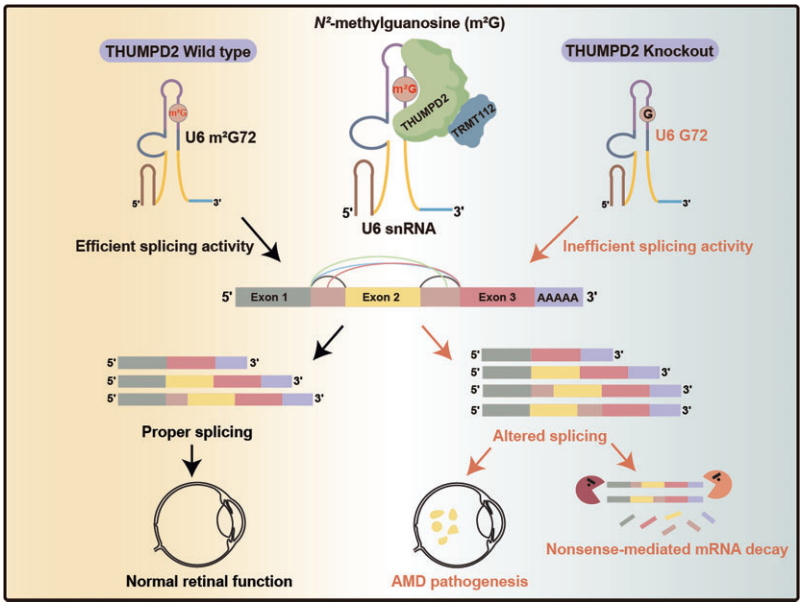
[1]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China
[2]Division of Reproduction and Genetics, The First Affiliated Hospital of USTC, MOE Key Laboratory for Membraneless Organelles and Cellular Dynamics, Hefei National Research Center for Interdisciplinary Sciences at the Microscale, Biomedical Sciences and Health Laboratory of Anhui Province, Division of Life Sciences and Medicine, University of Science and Technology of China, 230027 Hefei, China
[3]Institute of Biology, Westlake Institute for Advanced Study, Westlake University, Hangzhou 310064, Zhejiang Province, China
[4]Center for Applied Bioinformatics, St. Jude Children's Research Hospital, Memphis, TN, USA

[*]To whom correspondence should be addressed. Tel: +86 21 2068 4574; Email: liurj@shanghaitech.edu.cn
Correspondence may also be addressed to Beisi Xu. Tel: +1 901 595 8119; Email: Beisi.Xu@STJUDE.ORG
[†]The first two authors should be regarded as Joint First Authors.

Abstract

The mechanisms by which the relatively conserved spliceosome manages the enormously large number of splicing events that occur in humans (~200 000 versus ~300 in yeast) are poorly understood. Here, we show deposition of one RNA modification-N^2-methylguanosine (m^2G) on the G72 of U6 snRNA (the catalytic center of the spliceosome) promotes efficient pre-mRNA splicing activity in human cells. This modification was identified to be conserved among vertebrates. Further, THUMPD2 was demonstrated as the methyltransferase responsible for U6 m^2G72 by explicitly recognizing the U6-specific sequences and structural elements. The knock-out of *THUMPD2* eliminated U6 m^2G72 and impaired the pre-mRNA splicing activity, resulting in thousands of changed alternative splicing events of endogenous pre-mRNAs in human cells. Notably, the aberrantly spliced pre-mRNA population elicited the nonsense-mediated mRNA decay pathway. We further show that THUMPD2 was associated with age-related macular degeneration and retinal function. Our study thus demonstrates how an RNA epigenetic modification of the major spliceosome regulates global pre-mRNA splicing and impacts physiology and disease.

Graphical abstract



Received: August 26, 2023. **Revised:** December 13, 2023. **Editorial Decision:** December 14, 2023. **Accepted:** December 21, 2023

Introduction

Splicing of the precursor messenger RNA (pre-mRNA) into mature mRNA executed by a mega RNA-protein complex called the 'spliceosome' plays a pivotal role in eukaryotes (1–3). From yeast to mammals, the frequency of splicing events increases from ∼5% to ∼98% of all protein-coding genes (4,5). Approximately 300 genes containing a single intron occur in ∼6000 genes in *Saccharomyces cerevisiae* (6,7), while ∼20 000 genes, on average, contain eight introns in humans (8). Intriguingly, plenty of biochemical and structural studies show that the overall organization of the spliceosome and the configuration of the splicing catalytic core are incredibly conserved between humans and yeast (9–12). The enormously greater number of splicing events in higher eukaryotes are, therefore, a challenge to the operational capability of the spliceosome, but the precise underlying mechanism remains unclear. The spliceosome is a protein-orchestrated ribozyme dynamically formed anew on pre-mRNA involving five conserved small nuclear RNAs (snRNAs): U1, U2, U4, U5 and U6 with multitudinous proteins (12–14). SnRNAs catalyze pre-mRNA splicing, and U6 is located in the catalytic core of the spliceosome (15–18) (Figure 1A). The internal stem-loop (ISL) of U6 constitutes a major part of the splicing catalytic center (18–22). During pre-mRNA splicing, nucleotides of the catalytic core chelate two catalytic metal ions to fulfill branching and exon ligation (18,23–25).

RNA modifications play a critical regulatory role in various biological events (26). U6 is heavily modified post-transcriptionally in higher eukaryotes (27–29), with pseudouridine (ψ) and 2′-*O*-methylation (Nm) detected at multiple sites and two individual N^6-methyladenosine (m^6A, nucleoside 43 of human U6) and N^2-methylguanosine (m^2G, nucleoside 72 of human U6) (29–31). The ψ and Nm of U6 were predominantly introduced by Box C/D snoRNP (dyskerin) and Box H/ACA snoRNP (fibrillarin), respectively (32) (Figure 1B), and may increase the base-stacking and enhance base-pairing in U6 (33,34). Recently, U6 m^6A was found to be methylated by METTL16 and vital for recognizing 5′ splice site (5′ss) (35–37). However, the role of the only other base methylation (m^2G72) remains elusive.

G72 is a conserved catalytic active site in organisms varying from yeast (G78) to humans (18,38). The structures of the human major spliceosome reveal that G72 (G78 in yeast) of the ISL is involved in forming the activated catalytic center of the spliceosome and directly binds to the catalytic metal ion M1 (18,39,40) (Figure 1C). In the two-metal-ion model for splicing catalysis, human U6-G72, -U74 and pre-mRNA provide S_p phosphate oxygen to bind two metal ions, which bind with the phosphate of the guanosine of 5′ss to excise the intron (18). In a few sequenced U6s from mammals, an m^2G modification on G72 could be identified (Figure 2A). Although this modification was first identified four decades ago (30,31), its impact on pre-mRNA splicing and its physiological roles remain unclear.

This study shows the presence of m^2G72 in the U6 of zebrafish, mice and humans but absent in U6atac (the functional analog of U6 in the minor spliceosome). Using a reverse genetics approach coupled with RNA-mass spectrometry, we revealed that THUMP domain containing protein 2 (THUMPD2) is the methyltransferase responsible for the modification. THUMPD2 interacts with an auxiliary protein, a multifunctional methyltransferase subunit TRM112-like protein (TRMT112), to catalyze U6 m^2G72. *In vitro*, pre-mRNA splicing assay indicated that the loss of U6 m^2G72 decreased the splicing activity of the spliceosome. *THUMPD2* depletion causes thousands of changed alternative splicing (AS) events and upregulates the nonsense-mediated mRNA decay (NMD) pathway. Recent genome-wide association studies (GWAS) identified a genetic variant in *THUMPD2* associated with age-related macular degeneration (AMD). This study finds that the pre-mRNAs of about 50 reported AMD-associated genes and an additional 30 genes with retina-specific functions were the direct targets of THUMPD2-mediated regulation of splicing, indicating a role of THUMPD2 in AMD pathogenesis.

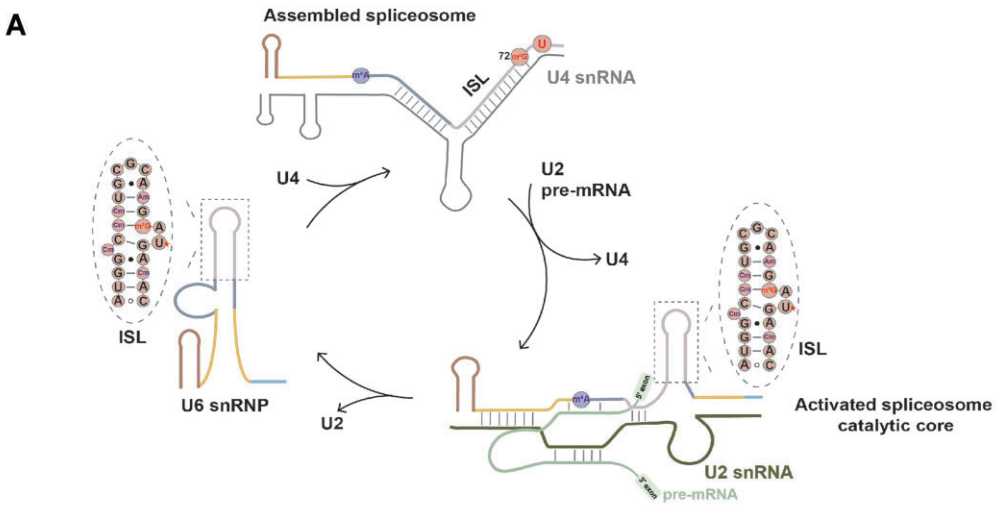
Materials and methods

Materials

Adenosine (A), guanosine (G), cytidine (C), uridine (U), N^2-methylguanosine (m^2G), N^6-methyladenosine (m^6A), 2′-*O*-methyladenosine (Am), 2′-*O*-methylguanosine (Gm), 2′-*O*-methylcytidine (Cm), pseudouridine (ψ), Ammonium acetate (NH_4OAc), Tris-base, β-mercaptoethanol (β-Me), Benzonase nuclease, Pyrophosphate, Phosphodiesterase I, sodium acetate (NaAc), Trypan Blue Stain 0.4%, and RIPA lysis Buffer (10×) were purchased from Sigma Aldrich. *S*-Adenosylmethionine (SAM) was purchased from New England Biolabs. CCK-8 Cell Counting Kit and ChamQ Universal SYBR QPCR Master Mix universal were purchased from Vazyme Biotech. RNase A, Ribolock RNase inhibitor, T4 DNA ligase, Streptavidin-conjugated agarose beads, 4′,6-diamidino-2-phenylindole (DAPI), restriction endonucleases, and Polyvinylidene fluoride (PVDF) membranes were purchased from Thermo Fisher Scientific. TRNzol Universal reagent was purchased from TIANGEN Biotech. Ni-NTA Superflow resin was purchased from Qiagen. Alkaline phosphatase (Calf intestine) was purchased from Takara Biomedical Technology. KOD-Plus-Neo Kit was purchased from TOYOBO Biotech. 1 M Tris–HCl Solution (Sterile), NaCl, magnesium chloride ($MgCl_2$), ATP, GTP, CTP, UTP and isopropyl-D-thiogalactoside (IPTG) were purchased from Sangon Biotech. Anti-HA magnetic beads, anti-Flag magnetic beads, Proteinase Inhibitor cocktail, and puromycin were purchased from MedChemExpress (MCE). Peroxidase-AffiniPure goat anti-rabbit/mouse IgG (H + L) were purchased from Yeasen. 5′-Biotin-DNA and 5′-Digoxigenin–DNA probes were purchased from BioSune. PCR or RT-qPCR primers were purchased from Tsingke Biological Technology. The antibodies used in this study were purchased from different companies and listed as follows: anti-HA antibody (C29F4 and 6E2, Cell signaling Technology); anti-Flag antibody (D6W5B, Cell signaling Technology); anti-TRMT112 antibody (A14310, ABclonal); anti-β-Tubulin (A12289, ABclonal); anti-GAPDH (AC002, ABclonal); anti-UPF1 (A5071, ABclonal); anti-phospho-UPF1 (Ser1127) (07-1016, Sigma-Aldrich). Alexa Fluor 647 AffiniPure Donkey Anti-Mouse IgG(H + L) (34113ES60, Yeasen).

Cell culture

HEK293T, SH-SY5Y, HepG2, HeLa, SK-BR-3, MDA-MB-231, MCF-7, NCl-H1299, A549, HCC38 and BT-549 cells were purchased from the cell resource center of the Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. The U-2 OS and hTERT RPE-1 cell



B

Modification	Positions	Chemical structure	Enzymes
N^2-methylguanosine (m^2G)	72		?
N^6-methyladenosine (m^6A)	43		METTL16
2'-*O*-methylation (Nm)	Am (47,53,70) Cm (60,62,63,77) Gm (54)		Box H/ACA snoRNP (Fibrillarin)
Pseudouridine (Ψ)	31,40,86		Box C/D snoRNP (Dyskerin)

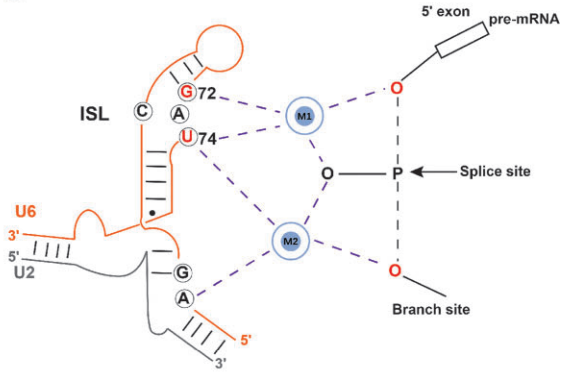


Figure 1. The characteristics of U6 snRNA. (**A**) Conformational changes of U6 snRNA during the pre-mRNA splicing cycles. The secondary structure of U6 in the U6 snRNP, the assembled spliceosome and the activated spliceosome are represented (21). (**B**) Post-transcriptional modifications of human U6. The identified sites, chemical structure and enzymes involved are listed. (**C**) The two metal-ion model of the splicing reaction. G72 and other key residues in the splicing catalytic center of U6 are labeled. M1 and M2, stand for the two metal ions; pre-mRNA is on the right (25).

lines were from Han-Hui Ma's Lab. The HCC38 cell line was from Jing-Yi Hui's Lab. The HeLa S3 cell line was from Yi-Gong Shi's Lab. Except for hTERT RPE-1 and SH-SY5Y, all of them were cultured at 37°C with 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM, Corning) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin. The medium for hTERT RPE-1 is Dulbecco's modified Eagle's medium/nutrient mixture F-12 (DMEM/F-12, Gibco). The medium for SH-SY5Y is a half-and-half mixture of DMEM/F-12 and DMEM. The viable cell numbers were counted by 0.4% trypan blue staining assay. Insect cells, *Spodoptera frugiperda* Sf9 and High Five, were cultured on a shaking incubator at 27°C and 120 rpm in ESF921 medium (Expression Systems).

Construction of knockout cell lines

Sense and anti-sense oligonucleotides for a guide RNA (sgRNA) were computationally designed for the selected genomic targets (http://crispor.tefor.net) and were cloned into pX330-mcherry vector (Addgene, 98750) which expresses red fluorescence protein (41). Three sgRNA sets were designed for *THUMPD2*. The sgRNA sequences (5′ to 3′) for relevant genes and targeting sites in HEK293T, HeLa, and HeLa S3 cells are shown in Supplementary Figure S4A-C respectively. For generating KO cell lines, sgRNA plasmids were transfected into cells using EZ Cell transfection reagent (Life-iLab, Shanghai, China). After transfection for 36 h, cells expressing red fluorescent protein were enriched by FACS Aria Fusion SORP (BD Bioscience) and plated into 96-well plates. The genotype of the stable cell lines was identified by confirming the frameshift mutations in the target region of the genome.

Western blotting

Cell lysates and immunoprecipitation complexes were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and the protein bands were transferred to 0.2 μm or 0.45 μm PVDF membranes. After blocking with 5% (w/v) non-fat dried milk, the membranes with targeted proteins were incubated with the corresponding primary antibodies overnight at 4°C. Membranes were then washed three times with PBST (phosphate-buffered saline with Tween-20) and incubated with HRP-conjugated secondary antibody at room temperature for 1 h. After washing three times with PBST, the membranes were treated with the chemiluminescent

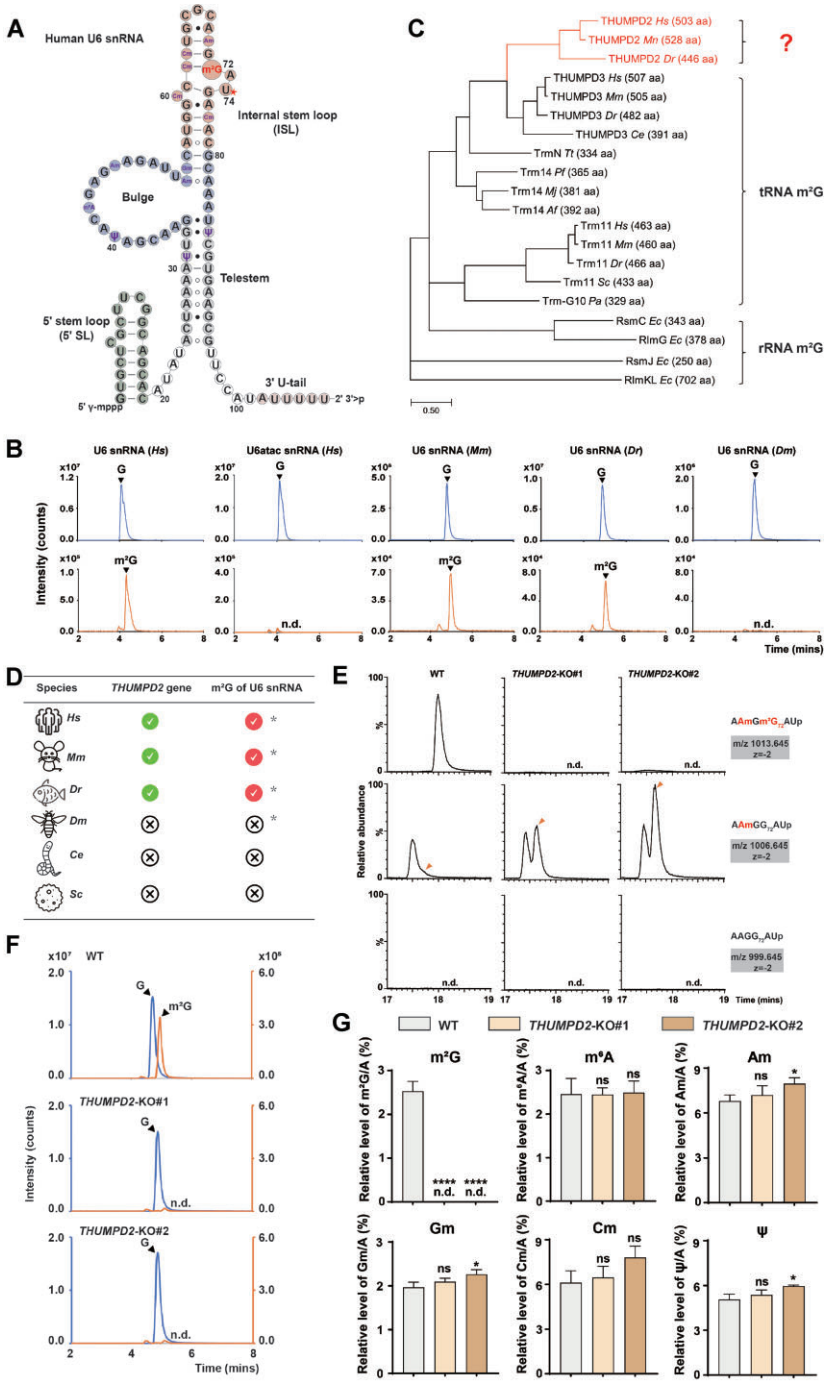


Figure 2. Analysis of U6 m^2G72 in the model eukaryotes and identification of THUMPD2 as the catalytic methyltransferase. (**A**) Secondary structures of the human U6 marked with the following elements: 5′γ-monomethyl cap, 5′ stem loop (5′SL), Bulge, Internal stem loop (ISL), Telestem and 3′U-tail (32). All the identified RNA modifications are labeled. G, guanosine; m^2G, N^2-methylguanosine. (**B**) Mass chromatograms of the nucleosides, G (upper panel) and m^2G (lower panel) of U6 extracted from the human HEK293T cells, brains of 11-week-old mice, tails of zebrafish and whole drosophila, respectively. U6atac snRNA extracted from HEK293T cells was also analyzed. Black asterisks indicate target peaks. The relative abundance of G was used as the control. (**C**) Phylogenetic analysis of the RNA m^2G-associated enzymes indicated that THUMPD2 is likely to be an evolutionarily conserved, N^2-methylation-related enzyme in vertebrates. aa, amino acids. (**D**) The evolution of *THUMPD2* and m^2G of U6 from yeast to humans. *m^2G on U6 verified in this study. (**E**) RNA-fragment-MS analysis of m^2G-containing fragments of U6 extracted from the WT and *THUMPD2*-KO HEK293T cells, digested with RNase A with (upper panels) or without (lower two panel) m^2G. The fragment sequences are indicated. The m/z value and charge state of each fragment are shown on the right. Orange asterisks indicate target peaks. n.d., not detected. (**F**) Mass chromatograms of the nucleosides, G and m^2G of the U6 extracted from the WT and *THUMPD2*-KO HEK293T cells. Black asterisks indicate target peaks. The relative abundance of G was used as the control. The value of the left vertical axis indicates the intensity of G, and that of the right vertical axis indicates the intensity of m^2G. (**G**) Quantification of the nucleosides, m^2G, m^6A, Am, Gm, Cm and ψ levels of U6 between the WT and *THUMPD2*-KO HEK293T cells. Data were normalized to the relative abundance of A. *Hs, Homo sapiens*; *Mm, Mus musculus*; *Dr, Danio rerio*; *Dm, Drosophila melanogaster*; *Ce, Caenorhabditis elegans*; *Tt, Thermus thermophilus*; *Pf, Pyrococcus furiosus*; *Mj, Methanocaldococcus jannaschii*; *Af, Archaeoglobus fulgidus*; *Sc, Saccharomyces cerevisiae*; *Pa, Pyrococcus abyssi*; and *Ec, Escherichia coli*. In Figure G, statistical analysis was performed using *t*-tests, and the error bars indicate the mean ± SD for three independent experiments. n.d., not detected; ns, not statistically significant; $^*P < 0.05$; $^{****}P < 0.0001$.

substrates (EpiZyme), and imaging was performed using the Amersham™ ImageQuant™ 800 (GE Healthcare).

Confocal immunofluorescence microscopy

HEK293T cells were transfected with pcDNA3.1(+)-THUMPD2-HA plasmid. After transfection for 36 h, the cells were fixed in 4% paraformaldehyde for 10 min and then permeated in 0.2% Triton X-100 for 5 min on ice. After washing with phosphate-buffered saline (PBS), the fixed cells were blocked in PBS containing 5% bovine serum albumin (BSA) and then incubated with mouse anti-HA antibody (6E2) with 1:800 dilution overnight at 4°C. The cells were then immunolabeled with Alexa Fluor 647-conjugated donkey anti-mouse IgG in PBS with 1:400 dilution for 1 h and the nuclear counterstain DAPI for 5 min at room temperature. Fluorescent images were taken and analyzed using a LSM980 Airyscan2 confocal microscope (Zeiss).

Immunoprecipitation

The cDNA of *HsTHUMPD2* (NM_025264.5) (along with a DNA sequence encoding a C-terminal HA-tag) and the cDNA of *HsTRMT112* (NM_016404.3) (along with a DNA sequence encoding a C-terminal Flag-tag) was inserted in pcDNA3.1(+), respectively. For immunoprecipitation, HEK293T cells were transfected with pcDNA3.1(+)-THUMPD2-HA or pcDNA3.1(+)-TRMT112-Flag. EZ Cell transfection reagent was used for transfection according to the manufacturer's protocol. After transfection for 36 h, the cells were washed three times with ice-cold PBS and then lysed with ice-cold 1 × RIPA lysis buffer supplemented with a Proteinase Inhibitor Cocktail for 20 min. The supernatant was collected by centrifugation at 12000 rpm for 10 min. Subsequently, the supernatant was incubated with the anti-HA magnetic beads or anti-Flag magnetic beads with gentle agitation for appropriate time. Recovered immunoprecipitation complexes were washed three times with ice-cold 1 × RIPA lysis buffer. All procedures were performed at 4°C. The immunoprecipitation complex was then eluted by buffer [20 mM Tris–HCl (pH 7.5), 150 mM NaCl, and 1% SDS]. The eluted IP samples were loaded on SDS-PAGE followed by western blotting analysis. To identify the protein interactome of *Hs*THUMPD2, the eluted samples of THUMPD2-HA were digested into peptides and then subjected to LC–MS/MS for protein identification.

Isolation of U6 or U6atac snRNAs by biotinylated DNA probes

Total RNA was extracted from the human cells, mouse tissues, fruit fly, and tails of zebrafish using TRNzol. The endogenous specific snRNAs used in this study were isolated from total cellular RNAs by their own biotinylated DNA probes using streptavidin agarose resin as described before (42). 30 μl of high-capacity streptavidin-conjugated agarose beads were washed with 10 mM Tris–HCl (pH 7.5) and then resuspended in 100 mM Tris–HCl (pH 7.5). Subsequently, 200 μM biotinylated oligonucleotides probes were mixed with beads and incubated at room temperature for 90 min. After the incubation, the oligonucleotide-conjugated beads were then washed three times with 10 mM Tris–HCl (pH 7.5). The total RNAs were dissolved in 6 × NTE solution [1 × NTE solution is 200 mM NaCl, 5 mM Tris–HCl (pH 7.5), and 2.5 mM EDTA]. The oligonucleotide-conjugated beads and total RNAs were mixed and heated at 70°C for 15 min and then cooled down slowly to 30°C. Then, the beads were washed with 3 × NTE for twice and 1 × NTE for twice. The specific snRNA retained on the beads was eluted with 0.1 × NTE at 70 °C and precipitated using 75% (v/v) ethanol. The probes for U6 snRNAs or U6atac snRNAs isolation used in this study are shown in Supplementary Table S1.

UPLC–MS/MS analysis of modified RNA nucleosides

About 200 ng of specific endogenous snRNAs purified by the biotinylated DNA probes or snRNA transcripts were digested with 0.2 μl benzonase nuclease, 0.25 μl phosphodiesterase I, and 0.25 μl bacterial alkaline phosphatase in a 20 μl solution including 4 mM NH_4OAc at 37°C overnight. After complete hydrolysis, the products were dissolved in acetonitrile and then applied to ultra-performance liquid chromatography-mass spectrometry/mass spectrometry (UPLC-MS/MS). The nucleosides were separated on an Atlantis® HILIC Silica column (3 μm, 2.1 × 150 mm) and then detected by a triple quadrupole mass spectrometer (AB Sciex QTRAP 6500+) in the positive ion multiple reaction monitoring (MRM) mode. Quantification was performed by comparison with the standard curve obtained from pure nucleoside standards running in the same batch. The mass of the precursor ion and the mass of the product ion (Q1/Q3) of the detected nucleosides are A (Q1/Q3 = 268.1/136.2), G (Q1/Q3 = 284.1/152.2), m^2G (Q1/Q3 = 298.1/166.1), m^6A (Q1/Q3 = 282.1/150.1), Am (Q1/Q3 = 282.1/136.2), Cm (Q1/Q3 = 258.1/112.1), Gm (Q1/Q3 = 298.1/152.1) and ψ (Q1/Q3 = 269.1/136.9).

Analysis of RNA fragment by mass spectrometry

About 500 ng U6 snRNA were digested with RNase A (5 U/μl) in a solution containing 10 mM NH_4OAc at 37°C for 1 h, as described (43). The digested products were applied to LC–ESI-MS on an ACQUITY UPLC M-CLASS/XevoG2-XS QTOF mass spectrometer (Waters Corporation). Resulting U6 snRNA fragments were separated at a 0.3 ml/min flow rate by an ACQUITY Premier BEH C18 column (Waters, 130 Å, 1.7 μm, 2.1 × 50 mm). The mobile phase consisted of ultra-pure water containing 0.1 M 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) (pH 7.0) & 15 mM triethylamine (TEA) (solvent A) and methanol containing 0.1 mM HFIP and 15 mM TEA (solvent B). The column was equilibrated in 1% B, and the gradient program was as follows: 0–2 min, 0% B; 2–20 min, 0–25% B; 20–25 min, 25–95% B; 25–30 min, 95% B; 30–35 min, 95–0% B. The eluent was ionized by an ESI source in negative polarity mode and scanned over an m/z range of 500–10 000. Ion source parameters were optimized by oligonucleotide standards ranging from 15 to 30 nt. MassLynx (Waters Corporation) extracts signals of modified or unmodified snRNA fragments. The sequence and modified sites of each fragment were confirmed by MS spectra. The extracted ion chromatography and the m/z of targeted modified or unmodified RNA fragments were analyzed using the embedded 'chromatogram' interface. And then the peak intensity of the RNA fragments was calculated by peak integration.

Preparation of snRNA transcripts

The DNA sequences of the T7 promoter and the U6s (of *Homo sapiens*, *Mus musculus* and *Danio rerio*) or *Hs*U6atac

were obtained from NCBI and cloned into pTrc99b vector (two restriction enzyme cutting sites for the vector were EcoRI and BamHI) to construct pTrc99b-T7-snRNA plasmids. The sequences of the mutated human U6s including U6 (G62-C72), U6 (G72A), U6-apical loop (atac), U6-Δ5'SL, U6-Δ3'U-tail, U6-Δ5'SL/3'U-tail, U6 ISL-apical loop-m.A (G65U), U6 ISL-apical loop-m.B (C66G), U6 ISL-apical loop-m.C (G67A), U6 ISL-apical loop-m.D (C68A), U6 ISL-apical loop-m.E (A69C), U6 ISL-apical loop-m.F (C61U), U6 ISL-apical loop-m.G (A73G), and U6 ISL-apical loop-m.H (ΔU) were constructed into pTrc99b vector as well as wildtype U6. All snRNA transcripts were generated via *in vitro* transcription using T7 RNA polymerase as described before (44). The transcribed snRNAs were denatured and annealed to form the optimum conformation in 5 mM $MgCl_2$. Subsequently, the concentration of snRNAs was determined by UV absorbance at 260 nm, and the molar absorption coefficient was calculated based on the sequence of each snRNA (45).

Cell counting Kit-8 (CCK-8) assay

Cell proliferation was determined using the Cell Counting Kit-8. Briefly, 1.2 × 10^3 WT or *THUMPD2* knockout HEK293T and HeLa cells were seeded in a 96-well flat-bottomed plate under normal culture. At days 1, 2, 3, 4 and 5, 10 μl of CCK-8 reagent was added into each well and the cells were incubated for 2 h at 37°C with 5% CO_2. The optical density at 450 nm (OD450) was measured using a Spark multimode plate reader (Tecan) to assay the growth curve. The experiments were repeated for three times.

RT-PCR and RT-qPCR

For quantification of specific mRNA level, total RNAs were enriched from cells using TRNzol and were reverse transcribed into cDNA using 4 × EZscript Reverse Transcription Mix II (EZBioscience). RT-PCR primers were designed to amplify two or multiple isoforms with different sizes. PCR products were quantified using Gel Dox XR + imaging system (Bio-Rad). RT-qPCR was performed using the ChamQ Universal SYBR QPCR Master Mix. The primers used for the corresponding mRNA in RT-PCR or qRT-PCR assays are listed Supplementary Table S1.

Absolute quantitative real-time PCR (RT-qPCR)

Total RNAs from mouse tissues were extracted using TRNzol according to the manufacturer's instructions. The first-strand DNA synthesis using total RNAs from mouse tissues as the templates was performed with 4 × EZscript Reverse Transcription Mix II. The standard material for calibration curves of coding sequence of mouse *Thumpd2* (NM_028138.1) was constructed into plasmids pcDNA3.1(+)-Thumpd2. The copy number of standards could range from 10^1 to 10^{10}, which based on the known concentrations of DNA standard molecules. RT-qPCR was performed using the standard curve method in QuantStudio 7 (Life Technology) with ChamQ Universal SYBR QPCR Master Mix as the dsDNA fluorescence dye. The amplification efficiency (*E*%) of standard material must be in the range of 90–110%. Thus, we designed several primers for each gene and pick up one that meets the amplification efficiency. The primers finally used for the *Thumpd2* in the RT-qPCR are listed in Supplementary Table S1.

Gene expression and protein purification

The genes encoding *Hs*THUMPD2 (along with a DNA sequence encoding a N-terminal 6 × His tag) and *Hs*TRMT112 (along with a DNA sequence encoding a C-terminal 6 × His tag) were inserted into pET-22b vector. The genes encoding *Mm*THUMPD2 (NM_028138.1) (along with a DNA sequence encoding a N-terminal 10 × His tag) and *Mm*TRMT112 (NM_001166370.1) were inserted after T7-Promoter-1 and T7-Promoter-2 site of pET-Duet-1 vector, respectively. The genes encoding *Dr*THUMPD2 (XM_686171.7) (along with a DNA sequence encoding a N-terminal 6 × His tag) and *Dr*TRMT112 (NM_001020752.1) were inserted after T7-Promoter-1 and T7-Promoter-2 site of pET-Duet-1 vector, respectively. All above constructs were expressed in *Escherichia coli* Rosetta.

The genes encoding *Hs*THUMPD2 fused with a C-terminal 10 × His-tag (THUMPD2-His) and *Hs*TRMT112 were inserted into pKL vector and then co-expressed in baculovirus-mediated transduction of High Five insect cells (46,47). Upon all the recombinant standalone proteins or protein complexes were purified by affinity chromatography on Ni-NTA Superflow resin, followed by gel filtration chromatography on a Superdex 200 Increase 10/300 GL column (GE Healthcare).

snRNA methyl-transfer assays

To confirm whether the modification introduced on snRNAs by purified *Hs*THUMPD2, *Hs*TRMT112, *Hs*THUMPD2-TRMT112, *Mm*THUMPD2-TRMT112, and *Dr*THUMPD2-TRMT112 is indeed m^2G on the G72, the reactions were carried out at 37°C for 2 h in a 50 μl reaction mixture containing 50 mM Tris–HCl (pH 7.5), 100 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT, 200 μM SAM, 5 μM snRNA and 1 μM THUMPD2-TRMT112 complex, respectively. After the reaction, the snRNAs were extracted with phenol/chloroform and precipitated using a two-fold volume of ethanol. Subsequently, the modified RNAs were digested into nucleosides or RNA fragments for following mass spectrometry analysis.

Northern blotting

6 μg total RNAs were electrophoresed through a 12% polyacrylamide 8 M urea gel and then transferred to a nylon membrane (MILLIPORE). The membrane was cross-linked twice at 254 nM, with 150 mJ/cm^2 using UV Stratalinker 1800. The membrane was then pretreated with ULTRAhyb Ultrasensitive Hybridization Buffer (Thermo Fisher Scientific) at 55°C for 2 h. Digoxigenin (DIG)-labeled DNA probes were denatured and subsequently added into the hybridization buffer to hybridize at 55°C overnight. The membrane was washed at 55°C and then incubated in blocking buffer (Roche) for 30 min at room temperature. Anti-Digoxigenin-AP (Fab fragments) antibody (Roche) was added and incubated for 30 min at room temperature. The membrane was washed with wash buffer [100 mM Maleic acid, 150 mM NaCl, 0.3%(v/v) Tween-20, pH 7.5] and incubated in detection buffer (100 mM NaCl, 100 mM Tris–HCl, pH 9.5). The membrane was incubated with CDP-Star (Roche) and signal intensities were detected by Amersham™ ImageQuant™ 800. The DNA probes used in this study are shown in Supplementary Table S1.

Construction of lentivirus mediated *THUMPD2* expression cell lines

Lentiviruses were generated in HEK293T cells by EZ Cell transfection reagent co-transfection of lentiviral-based pLVX-IRES-PURO plasmids, psPAX2 (lentiviral packaging plasmid), and pMD2.G (VSV-G envelope expressing plasmid) for 48 h. THUMPD2 knockout and WT HEK293T or HeLa S3 cells were incubated with viruses of *Hs*THUMPD2-HA and *Hs*THUMPD2 (D329A)-HA, for 24 h respectively. Followed by culturing in fresh media for 24 h, cells were then cultured in media supplemented with 2 μg/ml puromycin for another 96 h.

In vitro pre-mRNA splicing assay

Nuclear extracts from HeLa S3 cell lines were prepared for the *in vitro* pre-mRNA splicing as described before (48). The *in vitro* splicing reaction, assembled in a ∼25 μl volume, was performed in the presence of 10 ng MINX pre-mRNA substrate and 60% nuclear extracts, in the buffer containing 20 mM HEPES–KOH (pH 7.9), 55 mM KCl, 3 mM $MgCl_2$, 2 mM adenosine 5′-triphosphate (ATP), and 20 mM creatine phosphate. The splicing reaction mixture was incubated at 30°C at varying time points from 10 to 90 min, followed by proteinase K (20 mg/ml) digestion. Total RNAs from the *in vitro* splicing assay was extracted using phenol:chloroform:isopentanol at a volume ratio of 25:24:1. Reverse transcription was performed using RevertAid™ First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) with MINX RNA specific reverse primer. The RT-PCR products were resolved on 2% (w/v) agarose gel and stained by YeaRed Nucleic Acid Gel Stain (10 000× in Water) (Yeasen, Shanghai). The primers used for the *in vitro* pre-mRNA splicing assay in the RT-PCR are listed in Supplementary Table S1.

Transcriptome-seq and bioinformatic analyses

Transcriptomes of the WT and both two *THUMPD2*(KO#1 and KO#2) HEK293T cells were sequenced on Illumina NovaSeq sequencers, per the manufacturer's directions (Illumina). Each cell type has two biological replicates and each replicate got > 240 M paired-end reads to assure the confidence of detecting alternative splicing events. Paired-end 100-cycle sequencing reads were first trimmed for adapters using trim_galore (parameters: '–paired –retain_unpaired'). Then alignment to the reference genome hg38 (GRCh38.p6) by STAR (version 2.7.1a, parameters: ' –outFilterType BySJout –outFilterMultimapNmax 20 –alignSJoverhangMin 8 –alignSJDBoverhangMin 1 –outFilterMismatchNmax 999 –alignIntronMin 20 –alignIntronMax 1 000 000 –alignMatesGapMax 1 000 000 –outSAMstrandField intronMotif –outSAMattributes NH HI NM MD AS nM XS –outSAMunmapped Within KeepPairs –twopassMode Basic –outFilterScoreMinOverLread 0.66 –outFilterMatchNminOverLread 0.66') (49). For differential gene expression analysis, HTSeq (version 0.6.1p1) were used to estimate the counts based on GENCODE (hg38 version 31) (50,51). After normalization by TMM, Voom was used to identify differential expressed genes (52).

For splicing analysis, rMATS (Docker version v4.1.1, parameters: '–readLength 100 –variable-read-length –novelSS –allow-clipping') were used to call alternative splicing events (53). We performed three analyses as KO#1 (2 replicates) versus WT (2 replicates), KO#2 (2 replicates) versus WT and KO (4 replicates) versus WT. To assure the confidence, we take the results of four KO samples versus two WT samples and require they also significant in results for KO#1 vs WT and KO#2 vs WT, then we further required at least 10 reads and 0.5 counts per million. rmats2sashimiplot (https://github.com/Xinglab/rmats2sashimiplot v2.0.4 '–exon_s 1 –intron_s 5 –no-text-background') were used to make sashimi plots. The scores of the splice sites were calculated using MaxEnt software (54). Polypyrimidine tract length was computed in an interval of 50 nt upstream of 3′ splice sites. The distance in nt from the nucleotide position where the PY-fraction (C or U nucleotides) increased to 0.5 (sliding 5 nt windows across the 50 nt region) or higher and stayed higher to the end of the intron was taken as PY-tract length (code provided at https://doi.org/10.6084/m9.figshare.13369595).

Building and presentation of structural models

The model of THUMPD2 was built based on the structures of the *Tt*TrmN and *Pf*Trm14 (PDB codes: 3TMA and 3TM4) using the automated protein structure homology-modeling server, I-TASSER (55–59). The model of the human THUMPD2-TRMT112 complex was generated by manually docking TRMT112 from the human METTL5-TRMT112 complex (PDB code: 6H2V) onto the THUMPD2 through superimposition (60). All the structural models were presented in cartoon or stick using Pymol (https://www.pymol.org/).

Statistical analysis

Student's *t*-test was performed to compare the differences between knockout groups relative to their controls using GraphPad Prism v8. All computational results were presented as the mean ± SD *P*-values (*P*) are indicated in the figures above the two groups.

Results

Analysis of U6 m^2G72 and identification of THUMPD2 as the catalytic methyltransferase

An m^2G modification was identified on the spliceosomal active site (G72) of U6 in some mammals and G72 (G78 in yeast), which directly participated in pre-mRNA splicing via binding the catalytic metal ion M1 (18). However, the function of U6 m^2G72 remains unclear. To verify U6 m^2G72 in some representative eukaryotes, U6s were purified from humans, mice, zebrafish and drosophila, and then the RNA modifications were profiled using RNA mass spectrometry (RNA-MS) (Supplementary Figure S1A). An m^2G modification of U6 was evident in the three examined vertebrates (human, mouse and zebrafish) but not in drosophila (Figures 2B and D, Supplementary Figure S1B). The m^2G was localized at position 72 of the U6 isolated from human HEK293T cells, further confirmed by the mass spectrometry of the RNA fragments (RNA-fragment-MS). The fragments containing m^2G72, i.e. AAmG$\underline{m^2G}_{72}$AU were detected in the majority (70%–100%) when compared to G72 unmodified fragments: AAmG$\underline{G}_{72}$AU and AAG$\underline{G}_{72}$AU (Figure 2E left panel), suggesting a high level of m^2G72 modification of U6 *in vivo*. U6atac folded into a similar secondary structure as U6, and the equivalent nucleoside G72 (G44 of human U6atac) was also present in the splicing catalytic center (61) (Supplementary Figure S1C). U6atac was not modified with m^2G in humans, mice, or zebrafish (Figure 2B, Supplementary Figure S1D). Despite many efforts,

U6atac from zebrafish could not be obtained, which might be attributed to the infinitesimal U6atac in zebrafish. These findings suggested that U6 m^2G is prevalent in the examined vertebrates. Furthermore, we found that the m^2G level of U6 snRNA is remarkedly decreased under FBS (Fetal bovine serum) starvation in HeLa and HEK293T cells, suggesting that m^2G is a dynamic modification in response to external cues (Supplementary Figure S1E).

A reverse genetics approach coupled with RNA-MS was used to identify the enzyme involved in U6 m^2G72. First, a phylogenetic analysis of the known and predicted RNA m^2G methyltransferases revealed three candidates in vertebrates (Figure 2C). Among them, the THUMPD2-a predicted RNA m^2G methyltransferase was focused on, as it was the only one conserved in vertebrates but not in drosophila or lower eukaryotes as revealed by multiple protein sequence alignment (Figure 2D, Supplementary Figure S2). Using immunofluorescence assays, THUMPD2 was predominantly located in the nucleus, also the organelle to which U6 is localized (Supplementary Figure S3A). The *THUMPD2* in the human HEK293T, HeLa and HeLa S3 cells were knocked out (KO) using the CRISPR-Cas9 system (Supplementary Figures S4A-D). The U6 was then isolated from the *THUMPD2*-KO cells and subjected to RNA-MS and RNA-fragment-MS. Notably, m^2G72 on U6 was lost entirely in *THUMPD2*-KO HEK293T cells (Figures 2E right two panel and F), unlike the rest of the modifications, including m^6A, Nm (Am, Gm and Cm) and ψ (Figure 2G, Supplementary Figure S5). Further, *THUMPD2* depletion did not affect the m^2G modification levels of the other RNA species, such as poly(A)$^+$ RNA or tRNAs (62) (Supplementary Figure S1F). *THUMPD2*-KO HEK293T cells were rescued by plasmid-encoded THUMPD2, and the U6 m^2G72 was efficiently restored, demonstrating that THUMPD2 was responsible for this modification (Supplementary Figure S1G middle panel).

The function of THUMPD2 has been barely characterized; thus, its expression patterns in human cell lines and mouse tissues were measured. THUMPD2 is widely expressed in all the human cell lines derived from embryonic kidneys, neuroblastoma, liver, cervix, uterus, tibia, breasts, lungs and retina, tested in this study (Supplementary Figure S6A). *Thumpd2* mRNA was expressed in all the tissues examined from 11-week-old mice, with a relatively higher level in the thymus, brain and eyes (Supplementary Figure S6B).

In summary, U6 m^2G72 is conserved in vertebrates, and its methyltransferase THUMPD2 is widely expressed in various cells and tissues.

Molecular basis of the THUMPD2-mediated U6 m^2G72 modification

The *in vitro* methylation assays with the purified THUMPD2 protein to validate its methyltransferase activity indicated that THUMPD2 alone showed no detectable activity (Figure 3A). Many RNA methyltransferases functioning as complexes prompted the identification of THUMPD2-binding proteins. Thus, immunoprecipitation was performed coupled with proteomic analysis using THUMPD2-HA, which identified TRMT112 (Supplementary Figures S3B-E, Supplementary Table S2). TRMT112 is conserved in higher eukaryotes (Supplementary Figure S7) and is known to bind and regulate other methyltransferases involved in the methylation of rRNAs, tRNAs and proteins (62–64). A direct interaction between THUMPD2 and TRMT112 from human, mouse and zebrafish was confirmed by gel filtration analysis (Supplementary Figure S8A). A structural model of the human THUMPD2-TRMT112 was generated, in which THUMPD2 contains a classical 'Rossman-Fold' methyltransferase domain linked to a THUMP-containing domain, while TRMT112 possesses a small single domain (Figures 3B and C). *In vitro* methylation assays showed that the THUMPD2-TRMT112 complex of all three tested species could catalyze m^2G methylation of their cognate U6 (65) (Figure 3A, Supplementary Figures S8B–D). The specific methylation site 'G72' was confirmed by RNA-fragment-MS and mutagenesis studies *in vitro* (Supplementary Figures S3F and G). These results indicate that THUMPD2 formed a stable complex with TRMT112 to catalyze U6 m^2G72.

THUMPD2-TRMT112 catalyzed the m^2G modification of U6 but not U6atac (Figure 2B, Supplementary Figures S1B and D). To elucidate the molecular mechanism behind the substrate specificity of THUMPD2-TRMT112, mutagenic analysis based on differences between U6 and U6atac was performed, followed by *in vitro* methylation assays. U6 and U6atac are different in the 5′ stem loop (5′SL), 3′ U-tail and apical loop of ISL (32) (Figure 2A, Supplementary Figure S1C). The truncated 5′SL and/or 3′ U-tail or the ISL-apical loop on U6 substituted with the U6atac counterpart generated four U6 variants. The *in vitro* methylation assays using THUMPD2-TRMT112 indicated that m^2G-modification of U6 was unaffected by the truncation of 5′SL or 3′ U-tail but was completely inhibited when the ISL-apical loop was substituted with that of U6atac (Figure 3D). Subsequently, each nucleoside on the U6 ISL-apical loop ($_{65}$GCGCA$_{69}$) was mutated, with the *in vitro* methylation assays revealing that the m^2G levels of all the mutants decreased to ∼15–50% in comparison to the control (Figure 3E). These results demonstrated that the $_{65}$GCGCA$_{69}$ motif of the U6 ISL-apical loop was required for recognition by THUMPD2.

G72 is involved in the formation of the spliceosome catalytic center. The nucleosides A73 and U74 near G72 formed a specific bulged structure (21,66–68), in which G72 base paired with C62 (Figure 2A, Supplementary Figure S9). Since the methyl group of m^2G was on the purine base, it was assumed that the purine base of G72 must have flipped out to be modified by THUMPD2. Pursuing the involvement of this bulged structure during m^2G72 deposition, C61 was mutated to U, which can base pair with A73, or A73 mutated to G that might form a G73-C61 base pair, or the U74 deleted to disrupt the bulged structure. The *in vitro* methylation assays indicated that the m^2G levels in all these mutants decreased to ∼7–25% in comparison to the control (Figure 3F), indicating that the bulged structure around G72 was essential for the recognition of U6 by THUMPD2.

In summary, the THUMPD2-TRMT112 complex catalyzed U6 m^2G72 by recognizing the $_{65}$GCGCA$_{69}$ motif of the U6 ISL-apical loop and the U6-specific bulged structure around G72.

Loss of U6 m^2G72 decreased the *in vitro* pre-mRNA splicing activity of the spliceosome

U6 G72 is involved in forming the activated catalytic center of the human major spliceosome and functioned in pre-mRNA splicing by chelating a catalytic metal ion 'M1' that first attacked 5′ss for cleavage and next bound to 3′ ss for exon

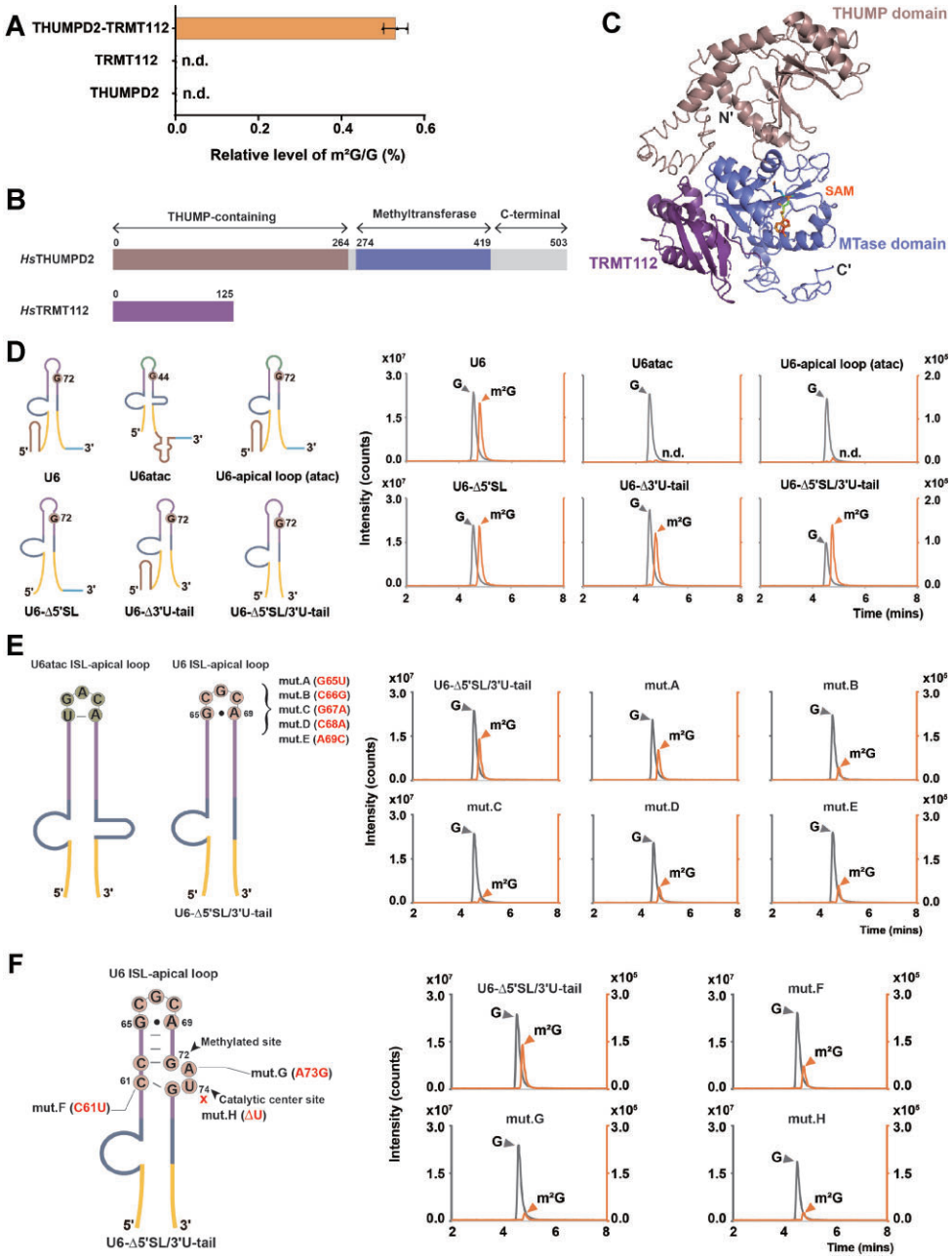


Figure 3. THUMPD2 interacts with TRMT112 to catalyze m^2G72 by recognizing U6-specific sequences and structural motifs. (**A**) Quantifying the methylation activity of purified *Hs*THUMPD2, *Hs*TRMT112 and *Hs*THUMPD2-TRMT112 complex on U6. Data were normalized to the relative abundance of G. (**B**) Domain composition of human THUMPD2 and TRMT112. *Hs*THUMPD2 was 503 aa long, and *Hs*TRMT112 was 125 aa long. THUMPD2 contains two main domains: an ancient RNA-binding THUMP domain and a S-adenosylmethionine (SAM)-dependent Methyltransferase (MTase) domain. (**C**) The structural model of the *Hs*THUMPD2-TRMT112 complex was generated using a protein structure homology-modeling server combined with manual structural superimposition and docking. The model is shown as a cartoon, and the THUMP domain (slate) and MTase domain (violet) of THUMPD2 with SAM (orange) plus TRMT112 (purple) are indicated. (**D**) Elements of U6 that are recognized by THUMPD2-TRMT112. Diagrams of the secondary structures of the different truncated or substitutive forms of U6 are shown on the left. Mass chromatograms of the G and m^2G from human U6, U6atac and U6 mutants after incubation with THUMPD2-TRMT112. (**E**) Mutations on the U6 ISL-apical loop. Diagrams of ISL and the sequences of the ISL-apical loop from human U6atac and U6 are shown on the left. The distinct nucleosides (G65, C66 and G67) in the ISL-apical loop of U6 were mutated to the corresponding bases of U6atac. The two identical nucleosides C68 and A69 were mutated to A68 and C69, respectively, and named m.A, m.B, m.C, m.D and m.E from 5′ to 3′. Mass chromatograms of the G and m^2G of U6 and the different mutants after incubation with THUMPD2-TRMT112. (**F**) A detailed diagram of the primary sequence around G72 of U6 is shown on the left. m.F, in which C61 was mutated to U61; m.G, in which A73 was mutated to G73; m.H, in which the catalytic site, U74 was deleted. Mass chromatograms of the G and m^2G of U6 and the various mutants after incubation with THUMPD2-TRMT112. In Figures D, E, and F, the target peaks are indicated by grey (**G**) and orange (m^2G) triangles. The value of the left vertical axis indicates the intensity of G, and that of the right vertical axis indicates the intensity of m^2G. In Figure A, statistical analysis was performed using *t*-tests, and the error bars indicate the mean ± SD for three independent experiments. n.d., not detected.

ligation (18) (Figure 1C). Since G72 participated in pre-mRNA splicing (18), the possible impacts of the m^2G modification of G72 on the pre-mRNA splicing activity of the spliceosome were investigated using the *in vitro* splicing assays.

It was first confirmed that *THUMPD2*-KO does not alter the U6 levels in all the HEK293T, HeLa and HeLa S3 cells tested (Figure 4A). Subsequently, *in vitro* pre-mRNA splicing assays were conducted using HeLa S3 cell nuclear extracts. An intron-containing MINX pre-mRNA (with two exons interrupted by an intron with the canonical splice sites of the major spliceosome with the 'GU-AG' rule) was incubated with the extracts. To assess the pre-mRNA splicing activity, unspliced and spliced MINX RNA were analyzed with RT-PCR at specific time points from 0 to 240 min (Figure 4B). The amounts of spliced MINX RNA in the *THUMPD2*-KO cells were conspicuously reduced compared to the WT cells starting at the 30 min time point, and the extent of reduction was proportional to the reaction time (Figure 4C). Notably, the pre-mRNA splicing activity could be rescued by nuclear extracts supplemented with plasmid-encoded WT THUMPD2 but not with the enzymatically inactive THUMPD2 mutant (D329A) (Figure 4D, Supplementary Figure S1G right panel), suggesting that the regulation of splicing activity was dependent on the U6 m^2G72 methylation capability of THUMPD2. Furthermore, overexpression of U6 but not the U6 (G72A) mutant could elevate the pre-mRNA splicing activity in the *THUMPD2*-KO cells (Supplementary Figures S10A and B), suggesting that the decreased splicing activity caused by *THUMPD2*-KO depended on the behavior of U6 to a certain extent. These results show that pre-mRNA splicing activity was impaired upon *THUMPD2*-KO, and the direct target may be U6.

THUMPD2-KO affects the splicing of endogenous pre-mRNAs

The probable impact of the *THUMPD2*-KO on the endogenous pre-mRNA substrates across the transcriptome was ascertained. To this end, RNA-seq analysis of the WT and *THUMPD2*-KO HEK293T cells was performed to assess the global pre-mRNA splicing changes using a junction read and exon expression approach rMATS (53). First, all KO samples (KO#1 and KO#2, four replicates) versus WT (two replicates) were called, in which each replicate possessed >240 M (million) paired-end reads. To assure the reproducibility of the results, KO#1 or KO#2 (each with two replicates) versus WT were also called independently, and the results were filtered by excluding those insignificant ($P < 0.05$) in either KO#1 or KO#2. Last, low-count events (>10 junction reads, >0.5 counts per million) were excluded to increase the confidence. A total of 4113 significant AS events were evident in the *THUMPD2*-KO cells, including 1907 skipped exons (SE), 906 alternative 3′ splice sites (A3SS), 604 retained introns (RI), 407 mutually exclusive exons (MXE), and 289 alternative 5′ splice sites (A5SS) (Figure 5A), affecting a total of 1923 genes [$|\psi| \geq 5\%$, false discovery rate (FDR) ≤0.05, >10 junction reads and >0.5 counts per million] (Supplementary Table S3).

Notably, SE events were predominant (Figure 5A, Supplementary Table S3). To assess the effects of *THUMPD2* depletion on SE events, donor or acceptor splice sites were analyzed to enrich unique motifs. The results revealed that the canonical sequences of 5′ ss and 3′ ss of the major spliceosome, GU and AG, were enriched in the donor and acceptor splice sites, respectively. No significant differences were observed in the sequences of these splice sites between the control and altered SE events, except that the reduced SE donor sites were enriched less with G at +5 while the increased SE acceptor sites enriched less for C at -3 (Supplementary Figure S11).

The pre-mRNA targets of the changed SE events comprised 956 protein-coding genes [FDR corrected $P < 0.05$] (Supplementary Table S3) and contained decreased or increased exon-skipping genes. These genes were involved in many bioprocesses (Figure 5B), such as *CDK2*, *CEP131*, *ENSA*, *CENPK* and *CDK20*, involved in the regulation of cell cycle and proliferation; *ALCAM*, *MAPK8*, *LLGL1*, *FLNA* and *CD46* in cell signaling pathways; and *NSMF*, *FKBP15*, *SLIT2*, *SETD5* and *ROBO3* in neuronal development. Notably, a moderate decrease in cell proliferation was consistently observed in *THUMPD2*-KO HEK293T and Hela cells compared with the WT cells (Supplementary Figures S12A and B). *THUMPD2*-KO increased the G0/G1 phase cell ratio but decreased those of the G2/M phase in both HEK293T and Hela cells (Supplementary Figures S12C and D).

A few selected representative AS events were validated by RT-PCR. For example, Sashimi plots indicated that SEs in *FLNA*, *CEP131*, and *FKBP15* while RIs in *SPOUT1* (Figure 5C) were higher in the *THUMPD2*-KO HEK293T cells compared to the WT. The increased exon inclusion or intron retention in these mRNAs was confirmed by RT-PCR (Figure 5D).

Thus, THUMPD2 affected the splicing of many endogenous pre-mRNAs that participate in many bioprocesses.

The expression of nonsense-mediated mRNA decay pathway-associated genes is upregulated in *THUMPD2*-KO cells

RNA-seq identified the global gene expression profile changes upon *THUMPD2*-KO. This study detected 511 and 592 genes to be significantly up- and down-regulated, respectively [$P < 0.05$, >2-fold] (Figure 6A, Supplementary Table S4). Pathway enrichment analysis could not identify the pathways enriched for the down-regulated genes. In contrast, the upregulated genes revealed several distinct enriched pathways (Figure 6B), including translation initiation, transcriptional activation and the NMD pathway. The up-regulation of several transcriptional activation- and translation initiation-related genes in the *THUMPD2*-KO cells was confirmed by RT-qPCR (Figure 6C).

The NMD pathway is an RNA surveillance mechanism for the selective degradation of transcripts with premature termination codons (PTCs) upstream of the exon junction complex (EJC) (69,70). It has been reported that approximately one-third of AS events produced transcripts containing PTCs that trigger NMD (71,72). Recalling the abundant AS events detected in the *THUMPD2*-KO cells (Figure 5A), the expression of the genes related to the NMD pathway was next examined. The central components of the NMD pathway are UPF1 with the stimulated factor SMG1, UPF3/UPF3X and EJC (the core of the EJC interacted with other proteins including RNPS1), the downstream key effectors included SMG5/SMG7 or SMG6, and the 5′-3′ exoribonuclease XRN1 (73) (Figure 6D). The key events in NMD are the phosphorylation and activation of UPF1 by SMG1,

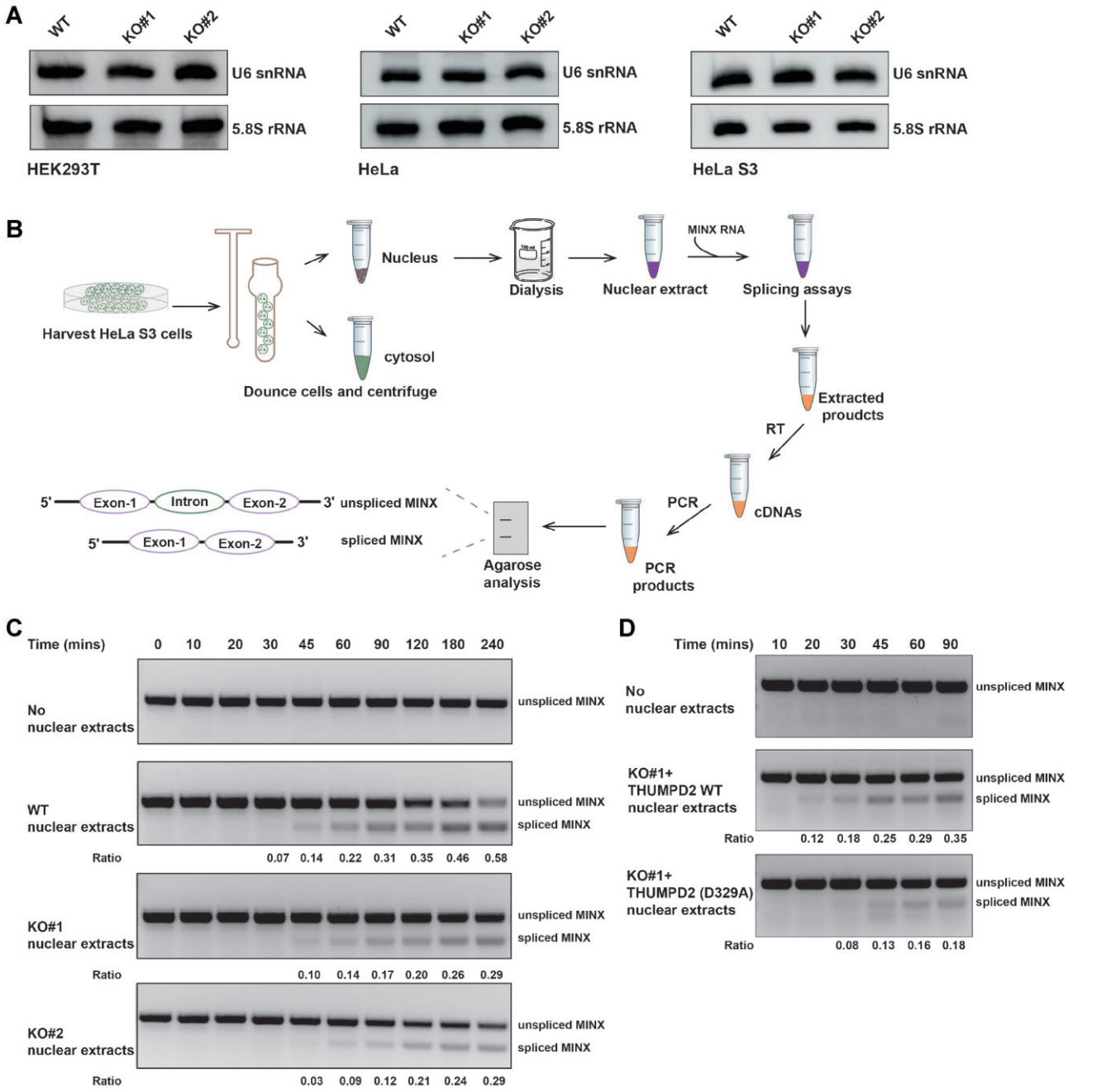


Figure 4. The *in vitro* assay of the splicing activity of the major spliceosome. (**A**) Northern blotting of the steady state level of U6 (upper lane) between the WT and *THUMPD2*-KO HEK293T, HeLa and HeLa S3 cells. 5.8S rRNA (lower lane) was used as a control. Northern blotting data shown are representative of three independent experiments. (**B**) Schematic of the *in vitro* splicing assays using the HeLa S3 nuclear extracts. (**C**) Detection of the pre-mRNA splicing activity using MINX pre-mRNA reporter in the WT and *THUMPD2*-KO HeLa S3 cells using RT-PCR. The unspliced (upper lane) and spliced (lower lane) MINX RNA were detected at varying time points from 0 to 240 min. *In vitro* splicing assay shown are representative of three independent experiments. (**D**) Detection of the pre-mRNA splicing activity using MINX pre-mRNA reporter in the *THUMPD2*-KO#1 HeLa S3 cells rescued by the plasmid-encoded human WT THUMPD2 and enzymatically inactive THUMPD2 mutant (D329A) using RT-PCR. The unspliced (upper lane) and spliced (lower lane) MINX RNA were detected at varying time points from 10 to 90 min. *In vitro* splicing assay shown are representative of four independent experiments. The ratios in Figures C and D represent the proportion of spliced MINX in the whole MINX RNA (unspliced and spliced MINX). The RT-PCR products were visualized using Gelred-stained agarose gels.

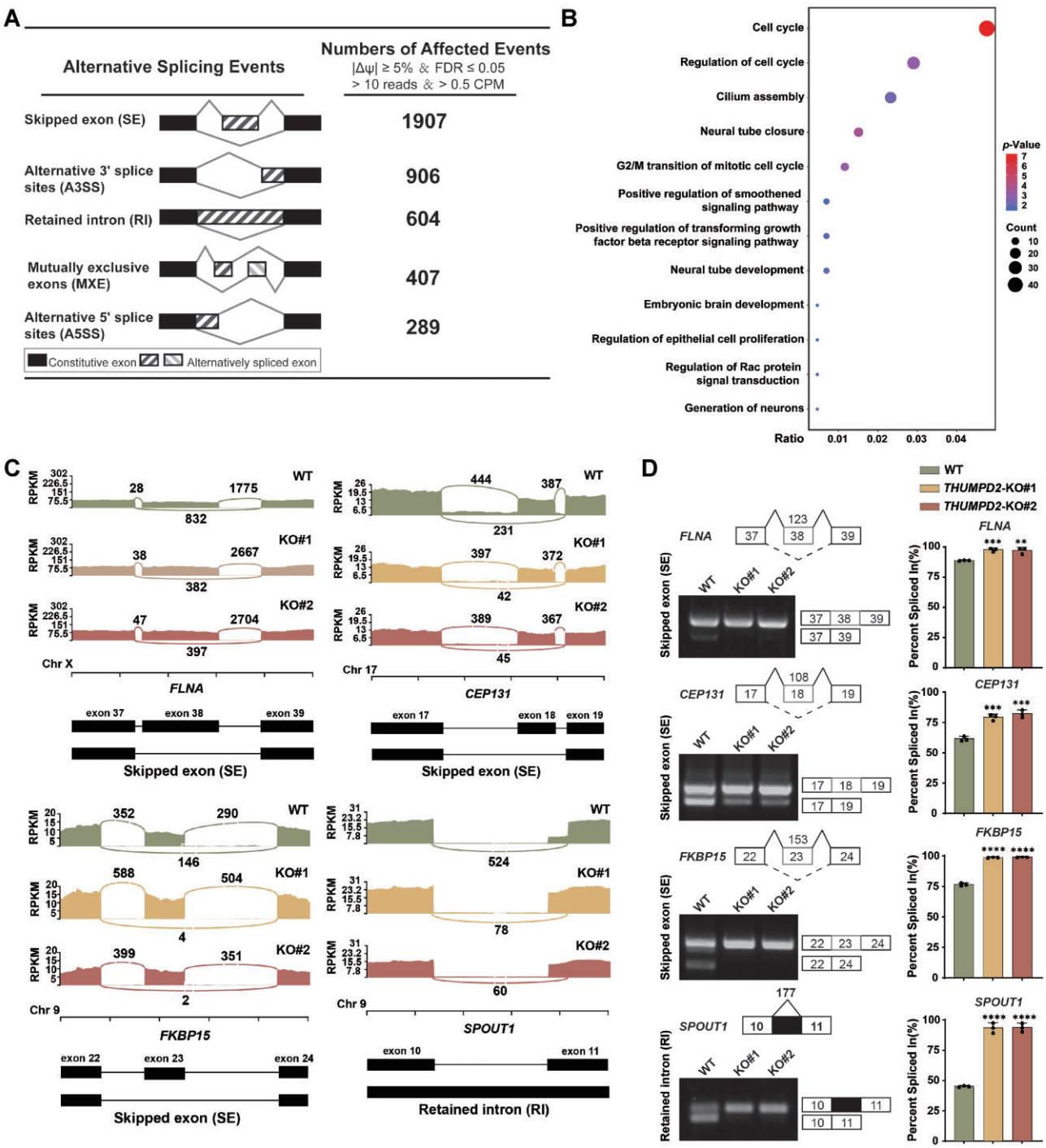


Figure 5. Analysis of endogenous pre-mRNA splicing in *THUMPD2*-KO cells. (**A**) Summary of the different alternative splicing (AS) events identified in *THUMPD2*-KO HEK293T cells and the number of splicing events affected by *THUMPD2* depletion. (**B**) Bubble chart indicating the enriched GO terms in the biological processes apart from the protein-coding genes with significantly altered skipped exon (SE) events in the *THUMPD2*-KO HEK293T cells. (**C**) Sashimi plots of the RNA-seq junction read counts for the representative SE events in *FLNA*, *CEP131* and *FKBP15* and retained intron (RI) events in *SPOUT1* in the *THUMPD2*-KO HEK293T cells. RPKM, reads per kilobase per million mapped reads. (**D**) Validation of the SE events in *FLNA*, *CEP131* and *FKBP15* and RI events in *SPOUT1* in the *THUMPD2*-KO HEK293T cells using RT-PCR with *β-actin* as the loading control. Left: visualization of the RT-PCR products using GelRed-stained agarose gels. Right: quantification of percent spliced in (PSI); $n = 3$. In Figure D, statistical analysis was performed using *t*-tests, and the error bars indicate the mean ± SD for three independent experiments. $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

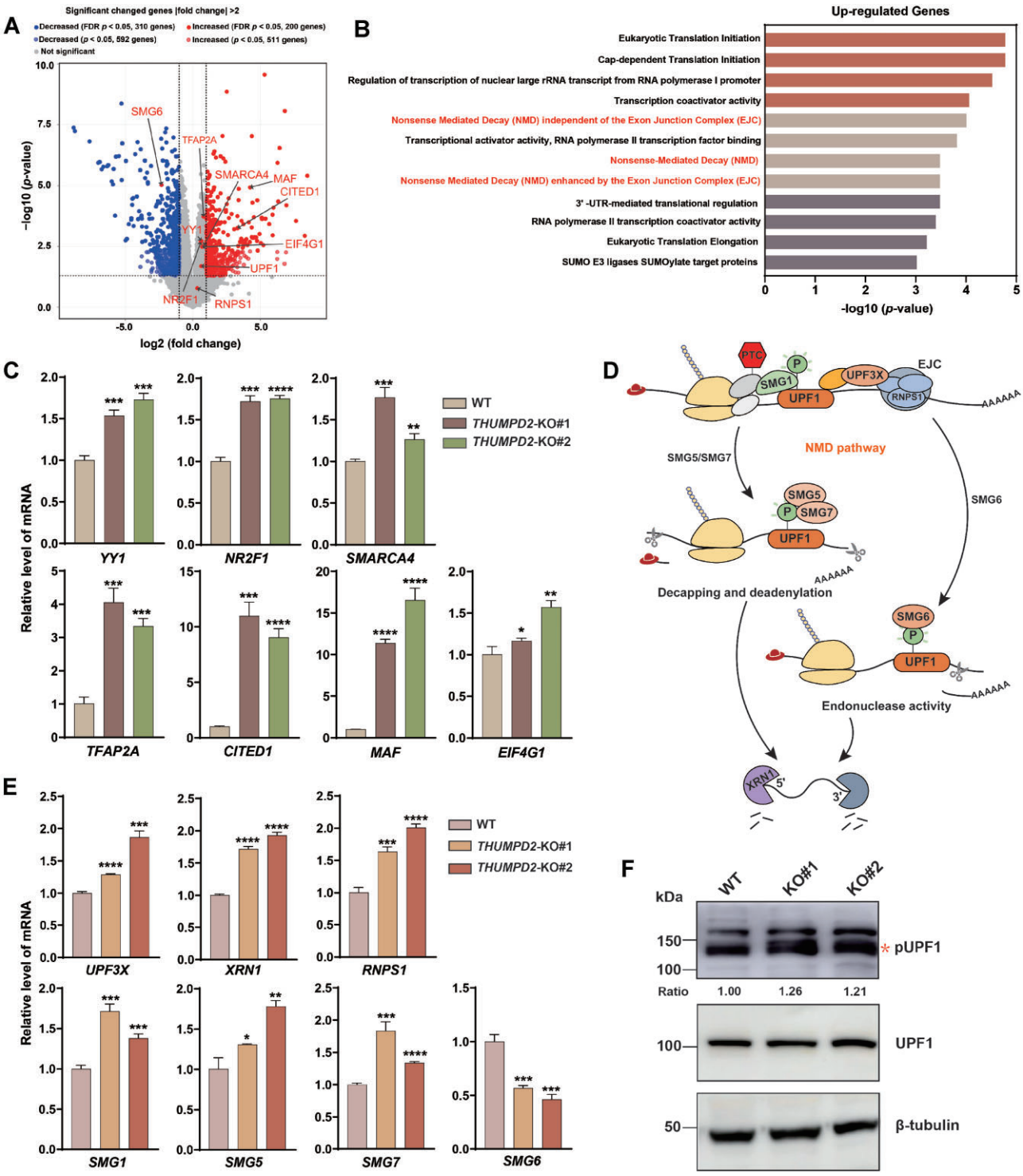


Figure 6. *THUMPD2*-KO upregulates the expression of key factors involved in the NMD pathway. (**A**) Volcano plot of the differentially expressed genes analyzed using the RNA-seq data comparing the *THUMPD2*-KO HEK293T ($n = 4$) to the WT cells ($n = 2$). The dots for the markedly regulated genes (|fold change| > 2; $P < 0.05$) were colored in red (increase) and blue (decrease). The significant genes that did not pass the false discovery rate (FDR) of 5% were transparent. The dots and labels of the genes of interest with transcriptional activation and those related to the nonsense-mediated RNA decay (NMD) pathway are colored in red. (**B**) Functional categories of the upregulated genes in the *THUMPD2*-KO HEK293T cells indicating the *P*-value for the enrichment of biological process-related GO terms. (**C**) The mRNAs of the key genes involved in transcriptional activation- and translation initiation-related genes such as *YY1*, *NR2F1*, *SMARCA4*, *TFAP2A*, *CITED1*, *MAF* and *EIF4G1* were measured using RT-qPCR. Data were normalized to *β-actin* expression. (**D**) The schematic diagram of the NMD pathway (73). The key factors involved were marked. (**E**) The mRNA levels of the key genes involved in the NMD pathway, such as *UPF3X*, *XRN1*, *RNPS1*, *SMG1*, *SMG5*, *SMG7* and *SMG6*, were measured by RT-qPCR. Data were normalized to *β-actin* expression. (**F**) Determination of the phosphorylation levels of UPF1 (pUPF1) by western blotting. The ratio represents the relative expression levels of pUPF1 normalized to *β-tubulin* expression. Western blotting data shown are representative of three independent experiments. In Figures C and E, statistical analysis was performed using *t*-tests, and the error bars indicate the mean ± SD for three independent experiments. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

inhibiting translation initiation and recruiting the factors mediating RNA decay (74). Therefore, the expression levels of these genes and the phosphorylation level of UPF1 were ascertained. Except for SMG6, the mRNAs of the other key NMD factors increased remarkably in the *THUMPD2*-KO cells than the WT with ∼1.3–2-fold change (Figure 6E). In consistency, the phosphorylation levels of UPF1 were ∼1.26 and 1.21 higher in the two *THUMPD2*-KO cells than in the WT (Figure 6F). These data support that NMD pathway-associated genes were conspicuously upregulated due to the increased AS events upon *THUMPD2*-KO, suggesting an elicitation of the NMD pathway.

The splicing-related regulatory function of THUMPD2 is associated with AMD

Splicing is a highly regulated, integral component of gene expression; thus, dysregulation in splicing can lead to altered gene expression and function, which may cause many diseases in humans (75). Recently, GWAS identified a genetic variant (rs557998486) in the enhancer of *THUMPD2* associated with AMD, one of the leading causes of blindness in the elderly, for which only limited therapeutic options are available (76). This variant was linked with the reduced expression of *THUMPD2*.

The regulation of pre-mRNA splicing played a role in the pathogenesis of retinal degeneration (77). Although the spliceosome is essential for accurately processing the pre-mRNA in all cell and tissue types, spliceosomopathies typically exhibit specificity towards specific cells or tissues (78). One of the major diseases in humans caused by mutations in genes encoding components of the spliceosome is retinitis pigmentosa (RP), a group of inherited retinal degeneration diseases that share similarities with AMD (79). AMD demonstrates a multifactor pathogenesis with the heritability estimated at 45–70% (80). Evaluating the contribution of one or more candidate pathogenic gene variants would more easily predict the development of AMD. To investigate the splicing-related regulatory role of THUMPD2 in the pathogenesis of AMD, the AS transcripts detected in the *THUMPD2*-KO cells linked with AMD in the clinical literature were classified.

First, 585 AMD-associated genes listed in a previous report (81) were collected with $r^2 > 0.5$, ± 500 kb surrounding 34 AMD-related loci, and their correlation with the AS events in the *THUMPD2*-KO cells was analyzed. SE and RI events were relatively more abundant, while SE overlapped even more (Figure 7A, Supplementary Table S5). Next, SE events were detected in 44 of the 585 genes (7.5%), which were candidate THUMPD2 SE targets, among which 37 genes (84%) showed a decreased inclusion level upon *THUMPD2*-KO (Supplementary Table S6). Further, 249 AMD-associated genes were collected from the GWAS catalog (EFO_0001365, 2022-07-16) (82), and five genes: *EIF3A, CD46, ARHGAP21, SKIV2L* and *DDR1* were also among the candidate THUMPD2 SE targets (Figures 7B and C). Interestingly, the THUMPD2 SE targets overlapped with differentially expressed genes (DEGs) in the choriocapillaris of the neovascular AMD and non-AMD donors (83). For instance, 26 THUMPD2 SE targets (Figure 7B, Supplementary Table S6) were shared with the 'arterial-derived endothelial' class of the total 525 AMD-patient-altered genes. To validate the SE events in these AMD-associated pre-mRNA upon *THUMPD2*-KO, exons in *LTBP3*, *LTBP4*, *SLTM*, *SKIV2L* and *EGFL7* with significantly higher exon-inclusion levels in *THUMPD2*-KO cells relative to the WT cells were chosen and confirmed by RT-PCR (Figure 7D). These results supported the idea that the splicing targets of THUMPD2 are involved in the pathogenesis of AMD.

The genes annotated with retina-specific functions were further focused on using the 445 Rod/Cone cell genes list from a previous report (84), and 30 genes overlapping with different AS events in the *THUMPD2*-KO cells were recognized (Supplementary Table S7). Among these, *STX3* is an important photoreceptor-encoding gene (85,86). Consistent with the results of this report, the SE of exon 10 regulated the *STX3* soluble isoform and nuclear localization, which could control the transcription of downstream genes (87). Additionally, several well-characterized retinal genes besides these 30 genes, including *STT3B*, *KIAA1549*, *POC1B*, and *MPP5*, were splicing targets of THUMPD2 (Supplementary Table S4). *STT3B* was discovered in GWAS for diabetic retinopathy (88,89). Dysregulation of *KIAA1549* causes retinal degeneration in murine models (90). The *poc1b* KO disrupted the development of the retina and impaired vision in zebrafish (91). Humans with a missense mutation of *POC1B* suffered from cone-rod dystrophy and reduced visual acuity since childhood (92). *MPP5* was involved in maintaining photoreceptor polarity in humans (93), and its loss of function mutations causes severe defects in the retina of zebrafish (94). The SE events in *KIAA1549*, *POC1B* and *MPP5* were then validated by RT-PCR (Figure 7E), identifying marked AS in all these pre-mRNAs in the *THUMPD2*-KO cells relative to the WT cells.

Together, the above data indicated that numerous AS targets of THUMPD2 are involved in AMD pathogenesis and retina-specific function, suggesting that THUMPD2 is related to retinal degeneration by regulating pre-mRNA splicing.

Discussion

From yeast to humans, the increased number of splicing events was coupled with increased modifications of snRNAs and regulatory splicing factors (11,12). SnRNA modifications could enhance the efficiency and fidelity of pre-mRNA splicing (37,95). Still, the specific underlying mechanisms remain largely unclear. Especially, m^2G72 on the U6 of the spliceosomal catalytic center was absent in lower eukaryotes but present in vertebrates. This data reveals that U6 m^2G72 may enhance the pre-mRNA splicing efficiency of the major spliceosome. The bulged secondary structure specifies the U6 ISL, which is the central part of the spliceosomal catalytic center. The conformation of this RNA bulge is essential for the coordination of the two catalytic metal ions, M1 and M2, and for both the steps of the splicing reaction (18,39). Since G72 (G78 in yeast) directly coordinates with M1, a subtle change by m^2G could plausibly impact splicing catalysis. Structurally, G72 base-paring with C62 (flanked by bulged A73 and U74) was positioned within a continuous stack of base-paired nucleotides(18,66–68). A recent study on $tRNA^{Trp}$ (96) reported that m^2G modification enhanced hydrophobicity and strengthened base-stacking interactions (97). It is also conceivable that the m^2G72 on U6 may stabilize the conformation of the spliceosomal catalytic center by chelating M1 to facilitate the splicing activity. The exact mechanism, however, awaits biochemical and structural investigations. Indeed, the overexpression of U6 appeared to suppress the slower splicing activity in the *THUMPD2*-KO cells, suggesting that U6 mainly caused the defects. However, it could

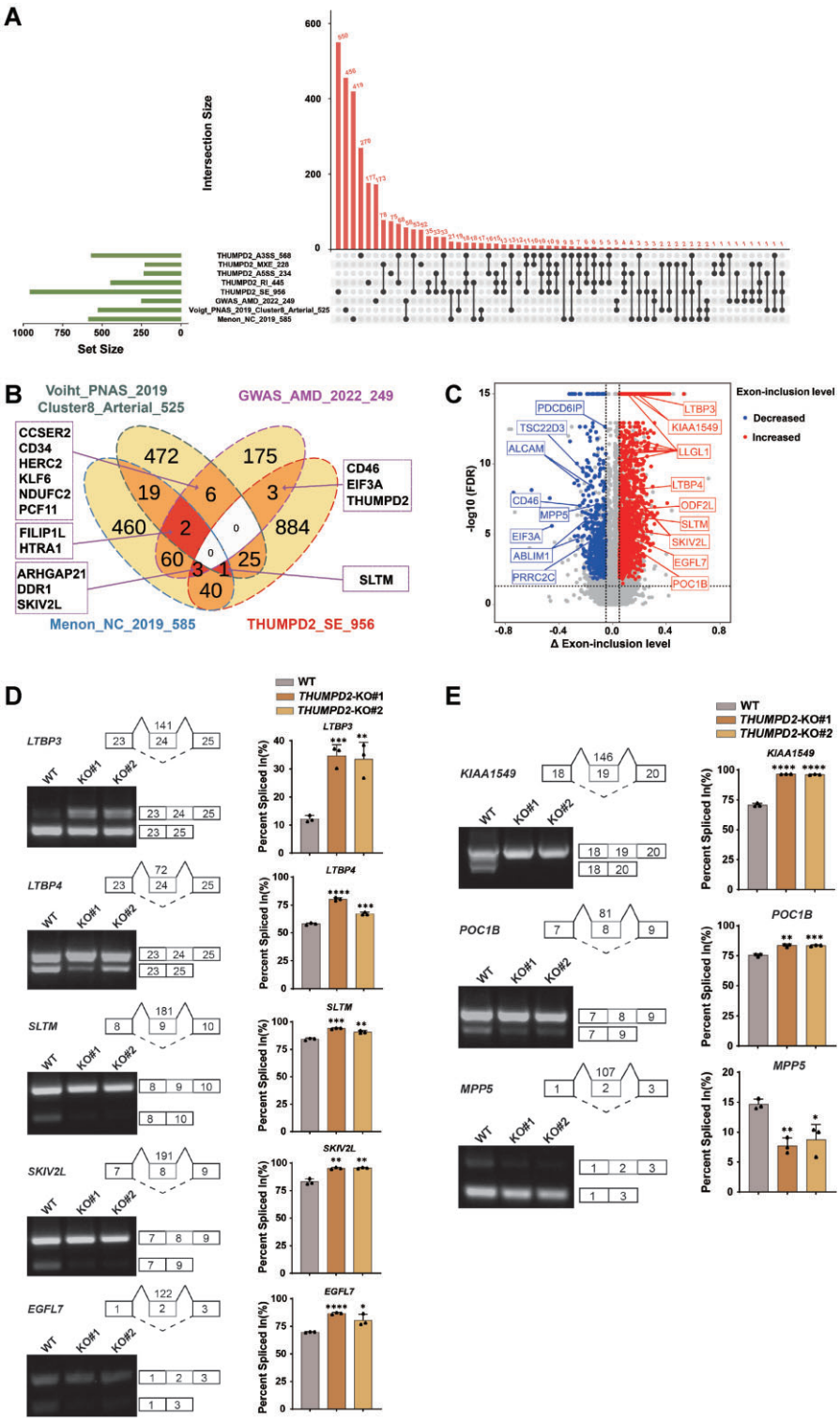


Figure 7. The function in splicing regulation of THUMPD2 is associated with AMD pathogenesis. (**A**) Upset plot showing the intersection of genes detected in the AS events of THUMPD2. The AMD-associated gene list was accessed from the GWAS Catalog EFO_0001365 (81,83). The Matrix layout for all intersections. Dark circles in the matrix indicate sets that are a part of the intersection. (**B**) Venn diagram of the THUMPD2 SE targets overlapping with curated, AMD-associated gene list. The gene list name suffixes indicate the number of genes in the corresponding lists. Groups with < ten genes are listed. (**C**) Volcano plots of the confidence level of up- or down-regulated SEs. $1e^{-15}$ were assigned to the FDR events with corrected *P*-value too negligible to be accurately estimated. Several representative genes associated with AMD are labeled. Blue and red stand for the decreased and increased exon-inclusion levels, respectively, known as PSI. (**D**) Validation of the SE events in *LTBP3*, *LTBP4*, *SLTM*, *SKIV2L* and *EGFL7* in the *THUMPD2*-KO HEK293T cells using RT-PCR with *β-actin* as a loading control. Left: visualization of the RT-PCR products using Gelred-stained agarose gels. Right: quantification of PSI; $n = 3$. (**E**) Validation of SE events in the retina-specific-functional genes, *KIAA1549*, *POC1B* and *MPP5* in the *THUMPD2*-KO HEK293T cells using RT-PCR with *β-actin* as a loading control. Left: visualization of the RT-PCR products using Gelred-stained agarose gels. Right: quantification of PSI; n = 3. In Figures D and E, statistical analysis was performed using *t*-tests, and the error bars indicate the mean ± SD for three independent experiments. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

not be identified if this was caused by defects in the spliceosome assembly and availability or by the spliceosome's inefficient splicing activity.

THUMPD2 depletion led to 4113 significant AS events that affect ∼2000 pre-mRNA species in HEK293T cells, suggesting a wide range of targets for regulating splicing by THUMPD2. These altered-spliced genes were involved in various biological processes, including cell cycle, proliferation, signaling pathways, neuronal development and apoptosis. In consistency, *THUMPD2* depletion further moderately decreased the proliferation of human cells. One-third of the pre-mRNA AS events in human and mouse cells routinely introduce PTCs to trigger the NMD pathway (72). These findings suggest that THUMPD2 was a robust regulator of the global pre-mRNA splicing. However, the possibility of defective cell proliferation affecting the splicing activity or the availability of fewer spliceosomes in the *THUMPD2*-KO cells cannot be ruled out. During the preparation of this manuscript, the KO of *THUMPD2* was also reported to cause a loss of m^2G on U6 and altered the splicing in a human colorectal carcinoma cell line [HCT116] (98), suggesting a general role of THUMPD2 in the regulation of splicing between different cell types. They reported that splicing events could correlated with splicing scores, intron length, or polypyrimidine tract length (PY-tract length) (98). We sought to check our data. Consistently, we found that retained introns have lower acceptor splicing scores, shorter length (significant, Wilcox test) and shorter polypyrimidine tract length. However, retained introns have higher donor splicing scores (significant, Wilcox test) (Supplementary Figures S13A-D).

The spliceosome is required in all cell and tissue types to process the pre-mRNAs properly; however, a dysfunctional spliceosome is typically cell- or tissue-specific, resulting in defects restricted to affecting only a few cell types (78). Interestingly, the retina specifically required proper splicing. For example, mutations in the pre-mRNA splicing factors: PRPF3, 4, 6, 8, 31, SNRNP200 and RP9 were associated with ∼15–20% of autosomal dominant RP (adRP) cases (99). Mutations in the members of the U4/U6.U5 tri-snRNP complex were almost exclusively the cause of RP disease, which impacted the accurate pre-mRNA splicing. The disruption of the kinetics and composition of the spliceosome, accompanied by a reduced rate of pre-mRNA splicing, was the principal mechanism used to elucidate the RP-associated spliceosomopathies (100). Analogically, THUMPD2 depletion caused a loss of m^2G72 in the catalytic center of the U6, which could further ease the rate of pre-mRNA splicing and affect AS events involving the regulation of retinal function. As observed, many AS targets of THUMPD2 were involved in AMD pathogenesis and retina-specific processes, suggesting that the *THUMPD2* mutation associated with AMD identified by GWAS may work by regulating the splicing function of THUMPD2.

Data availability

All data are available in the main text or the appendix data. Raw and processed data of RNA-seq have been deposited at NCBI GEO under accession GSE236689.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

This work is supported by National Key Research and Development Program of China [2020YFA0803400, 2021YFA1100800]; National Natural Science Foundation of China [32022040, 31971230]; This work is also supported by the Shanghai Frontiers Science Center for Biomacromolecules and Precision Medicine in ShanghaiTech University.

We thank the Molecular and Cell Biology Core Facility (MCBCF), the Multi-Omics Core Facility (MOCF), and the Molecular Imaging Core Facility (MICF) at the School of Life Science and Technology, ShanghaiTech University for providing technical support. We also thank the Analytical Chemistry platform (ShanghaiTech University, SIAIS) for technical assistance with protein identification by MS. We thank Prof. Christopher L. Antos and Prof. Ji-Long Liu at ShanghaiTech University for providing zebrafish and drosophila.

Funding

National Key Research and Development Program of China [2020YFA0803400, 2021YFA1100800]; National Natural Science Foundation of China [31971230, 32022040]. Funding for open access charge: National Natural Science Foundation of China.

Conflict of interest statement

None declared.

References

1. Brody,E. and Abelson,J. (1985) The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science*, **228**, 963–967.
2. Grabowski,P.J., Seiler,S.R. and Sharp,P.A. (1985) A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell*, **42**, 345–353.
3. Frendewey,D. and Keller,W. (1985) Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell*, **42**, 355–367.
4. Wang,E.T., Sandberg,R., Luo,S., Khrebtukova,I., Zhang,L., Mayr,C., Kingsmore,S.F., Schroth,G.P. and Burge,C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
5. Pan,Q., Shai,O., Lee,L.J., Frey,B.J. and Blencowe,B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
6. Spingola,M., Grate,L., Haussler,D. and Ares,M. Jr (1999) Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA*, **5**, 221–234.
7. Qin,D., Huang,L., Wlodaver,A., Andrade,J. and Staley,J.P. (2016) Sequencing of lariat termini in *S. cerevisiae* reveals 5' splice sites, branch points, and novel splicing events. *RNA*, **22**, 237–253.
8. Hong,X., Scofield,D.G. and Lynch,M. (2006) Intron size, abundance, and distribution within untranslated regions of genes. *Mol. Biol. Evol.*, **23**, 2392–2404.
9. Brow,D.A. and Guthrie,C. (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*, **334**, 213–218.
10. Wan,R., Bai,R., Zhan,X. and Shi,Y. (2020) How is precursor messenger RNA spliced by the spliceosome? *Annu. Rev. Biochem.*, **89**, 333–358.

11. Wilkinson,M.E., Charenton,C. and Nagai,K. (2020) RNA splicing by the spliceosome. *Annu. Rev. Biochem.*, **89**, 359–388.
12. Kastner,B., Will,C.L., Stark,H. and Lührmann,R. (2019) Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harb. Perspect. Biol.*, **11**, a032417.
13. Jurica,M.S. and Moore,M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell*, **12**, 5–14.
14. Wahl,M.C., Will,C.L. and Lührmann,R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
15. Madhani,H.D. and Guthrie,C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*, **71**, 803–817.
16. Valadkhan,S. (2005) snRNAs as the catalysts of pre-mRNA splicing. *Curr. Opin. Chem. Biol.*, **9**, 603–608.
17. Madhani,H.D. (2013) snRNA catalysts in the spliceosome's ancient core. *Cell*, **155**, 1213–1215.
18. Fica,S.M., Tuttle,N., Novak,T., Li,N.S., Lu,J., Koodathingal,P., Dai,Q., Staley,J.P. and Piccirilli,J.A. (2013) RNA catalyses nuclear pre-mRNA splicing. *Nature*, **503**, 229–234.
19. Brow,D.A. and Guthrie,C. (1989) Splicing a spliceosomal RNA. *Nature*, **337**, 14–15.
20. Wolff,T. and Bindereif,A. (1993) Conformational changes of U6 RNA during the spliceosome cycle: an intramolecular helix is essential both for initiating the U4-U6 interaction and for the first step of slicing. *Genes Dev.*, **7**, 1377–1389.
21. McManus,C.J., Schwartz,M.L., Butcher,S.E. and Brow,D.A. (2007) A dynamic bulge in the U6 RNA internal stem-loop functions in spliceosome assembly and activation. *RNA*, **13**, 2252–2265.
22. Eysmont,K., Matylla-Kulińska,K., Jaskulska,A., Magnus,M. and Konarska,M.M. (2019) Rearrangements within the U6 snRNA core during the transition between the two catalytic steps of splicing. *Mol. Cell*, **75**, 538–548.
23. Steitz,T.A. and Steitz,J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad Sci. USA*, **90**, 6498–6502.
24. Yean,S.L., Wuenschell,G., Termini,J. and Lin,R.J. (2000) Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature*, **408**, 881–884.
25. Lee,Y. and Rio,D.C. (2015) Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.*, **84**, 291–323.
26. Frye,M., Harada,B.T., Behm,M. and He,C. (2018) RNA modifications modulate gene expression during development. *Science*, **361**, 1346–1349.
27. Karijolich,J. and Yu,Y.T. (2010) Spliceosomal snRNA modifications and their function. *RNA Biol.*, **7**, 192–204.
28. Bohnsack,M.T. and Sloan,K.E. (2018) Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. *Biol. Chem.*, **399**, 1265–1276.
29. Morais,P., Adachi,H. and Yu,Y.T. (2021) Spliceosomal snRNA epitranscriptomics. *Front Genet.*, **12**, 652129.
30. Epstein,P., Reddy,R., Henning,D. and Busch,H. (1980) The nucleotide sequence of nuclear U6 (4.7 S) RNA. *J. Biol. Chem.*, **255**, 8901–8906.
31. Harada,F., Kato,N. and Nishimura,S. (1980) The nucleotide sequence of nuclear 4.8S RNA of mouse cells. *Biochem. Biophys. Res. Commun.*, **95**, 1332–1340.
32. Didychuk,A.L., Butcher,S.E. and Brow,D.A. (2018) The life of U6 small nuclear RNA, from cradle to grave. *RNA*, **24**, 437–460.
33. Davis,D.R. (1995) Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.*, **23**, 5020–5026.
34. Sashital,D.G., Venditti,V., Angers,C.G., Cornilescu,G. and Butcher,S.E. (2007) Structure and thermodynamics of a conserved U2 snRNA domain from yeast and human. *RNA*, **13**, 328–338.
35. Pendleton,K.E., Chen,B., Liu,K., Hunter,O.V., Xie,Y., Tu,B.P. and Conrad,N.K. (2017) The U6 snRNA m^6A methyltransferase METTL16 regulates SAM synthetase intron retention. *Cell*, **169**, 824–835.
36. Mendel,M., Delaney,K., Pandey,R.R., Chen,K.M., Wenda,J.M., Vågbø,C.B., Steiner,F.A., Homolka,D. and Pillai,R.S. (2021) Splice site m^6A methylation prevents binding of U2AF35 to inhibit RNA splicing. *Cell*, **184**, 3125–3142.
37. Ishigami,Y., Ohira,T., Isokawa,Y., Suzuki,Y. and Suzuki,T. (2021) A single m^6A modification in U6 snRNA diversifies exon sequence at the 5' splice site. *Nat. Commun.*, **12**, 3244.
38. Guthrie,C. (1991) Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science*, **253**, 157–163.
39. Huppler,A., Nikstad,L.J., Allmann,A.M., Brow,D.A. and Butcher,S.E. (2002) Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. *Nat. Struct. Biol.*, **9**, 431–435.
40. Zhang,X., Yan,C., Hang,J., Finci,L.I., Lei,J. and Shi,Y. (2017) An atomic structure of the Human spliceosome. *Cell*, **169**, 918–929.
41. Wu,Y., Liang,D., Wang,Y., Bai,M., Tang,W., Bao,S., Yan,Z., Li,D. and Li,J. (2013) Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell*, **13**, 659–662.
42. Huang,Q., Yao,P., Eriani,G. and Wang,E.D. (2012) In vivo identification of essential nucleotides in $tRNA^{Leu}$ to its functions by using a constructed yeast $tRNA^{Leu}$ knockout strain. *Nucleic Acids Res.*, **40**, 10463–10477.
43. Suzuki,T., Ikeuchi,Y., Noma,A., Suzuki,T. and Sakaguchi,Y. (2007) Mass spectrometric identification and characterization of RNA-modifying enzymes. *Methods Enzymol.*, **425**, 211–229.
44. Li,Y., Chen,J., Wang,E. and Wang,Y. (1999) T7 RNA polymerase transcription of Escherichia coli isoacceptors $tRNA^{Leu}$. *Sci. China C Life Sci.*, **42**, 185–190.
45. Kibbe,W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.*, **35**, W43–W46.
46. Fitzgerald,D.J., Berger,P., Schaffitzel,C., Yamada,K., Richmond,T.J. and Berger,I. (2006) Protein complex expression by using multigene baculoviral vectors. *Nat. Methods*, **3**, 1021–1032.
47. Trowitzsch,S., Bieniossek,C., Nie,Y., Garzoni,F. and Berger,I. (2010) New baculovirus expression tools for recombinant protein complex production. *J. Struct. Biol.*, **172**, 45–54.
48. Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
49. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
50. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
51. Harrow,J., Frankish,A., Gonzalez,J.M., Tapanari,E., Diekhans,M., Kokocinski,F., Aken,B.L., Barrell,D., Zadissa,A., Searle,S., *et al.* (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.*, **22**, 1760–1774.
52. Law,C.W., Chen,Y., Shi,W. and Smyth,G.K. (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.*, **15**, R29.
53. Shen,S., Park,J.W., Lu,Z.X., Lin,L., Henry,M.D., Wu,Y.N., Zhou,Q. and Xing,Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc. Natl. Acad. Sci. USA*, **111**, E5593–E5601.
54. Yeo,G. and Burge,C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
55. Fislage,M., Roovers,M., Tuszynska,I., Bujnicki,J.M., Droogmans,L. and Versées,W. (2012) Crystal structures of the tRNA:m^2G6 methyltransferase Trm14/TrmN from two domains of life. *Nucleic Acids Res.*, **40**, 5149–5161.

56. Callaway,E. (2020) 'It will change everything': deepMind's AI makes gigantic leap in solving protein structures. *Nature*, **588**, 203–204.
57. Roy,A., Kucukural,A. and Zhang,Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, **5**, 725–738.
58. Yang,J., Yan,R., Roy,A., Xu,D., Poisson,J. and Zhang,Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*, **12**, 7–8.
59. Yang,J. and Zhang,Y. (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.*, **43**, W174–W181.
60. van Tran,N., Ernst,F.G.M., Hawley,B.R., Zorbas,C., Ulryck,N., Hackert,P., Bohnsack,K.E., Bohnsack,M.T., Jaffrey,S.R., Graille,M., *et al.* (2019) The human 18S rRNA m^6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Res.*, **47**, 7719–7733.
61. 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.
62. Yang,W.Q., Xiong,Q.P., Ge,J.Y., Li,H., Zhu,W.Y., Nie,Y., Lin,X., Lv,D., Li,J., Lin,H., *et al.* (2021) THUMPD3-TRMT112 is a m^2G methyltransferase working on a broad range of tRNA substrates. *Nucleic Acids Res.*, **49**, 11900–11919.
63. Liger,D., Mora,L., Lazar,N., Figaro,S., Henri,J., Scrima,N., Buckingham,R.H., van Tilbeurgh,H., Heurgué-Hamard,V. and Graille,M. (2016) Mechanism of activation of methyltransferases involved in translation by the Trm112 'hub' protein. *Nucleic Acids Res.*, **44**, 1482.
64. Bourgeois,G., Létoquart,J., van Tran,N. and Graille,M. (2017) Trm112, a protein activator of methyltransferases modifying actors of the eukaryotic translational apparatus. *Biomolecules*, **7**, 7.
65. Boonanuntanasarn,S., Panyim,S. and Yoshizaki,G. (2008) Characterization and organization of the U6 snRNA gene in zebrafish and usage of their promoters to express short hairpin RNA. *Mar Genomics*, **1**, 115–121.
66. Sun,J.S. and Manley,J.L. (1997) The human U6 snRNA intramolecular helix: structural constraints and lack of sequence specificity. *RNA*, **3**, 514–526.
67. Zhan,X., Yan,C., Zhang,X., Lei,J. and Shi,Y. (2018) Structure of a human catalytic step I spliceosome. *Science*, **359**, 537–545.
68. Wolff,T., Menssen,R., Hammel,J. and Bindereif,A. (1994) Splicing function of mammalian U6 small nuclear RNA: conserved positions in central domain and helix I are essential during the first and second step of pre-mRNA splicing. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 903–907.
69. Maquat,L.E. and Carmichael,G.G. (2001) Quality control of mRNA function. *Cell*, **104**, 173–176.
70. Maquat,L.E. (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.*, **5**, 89–99.
71. Kurosaki,T., Popp,M.W. and Maquat,L.E. (2019) Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.*, **20**, 406–420.
72. Lewis,B.P., Green,R.E. and Brenner,S.E. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 189–192.
73. Wolin,S.L. and Maquat,L.E. (2019) Cellular RNA surveillance in health and disease. *Science*, **366**, 822–827.
74. Popp,M.W. and Maquat,L.E. (2013) Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.*, **47**, 139–165.
75. Rogalska,M.E., Vivori,C. and Valcárcel,J. (2023) Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat. Rev. Genet.*, **24**, 251–269.
76. Guindo- Martínez,M., Amela,R., Bonàs-Guarch,S., Puiggròs,M., Salvoro,C., Miguel-Escalada,I., Carey,C.E., Cole,J.B., Rüeger,S., Atkinson,E., *et al.* (2021) The impact of non-additive genetic associations on age-related complex diseases. *Nat. Commun.*, **12**, 2436.
77. Keuthan,C.J., Karma,S. and Zack,D.J. (2023) Alternative RNA splicing in the retina: insights and perspectives. *Cold Spring Harb. Perspect. Med.*, **13**, a041313.
78. Griffin,C. and Saint-Jeannet,J.P. (2020) Spliceosomopathies: diseases and mechanisms. *Dev. Dyn.*, **249**, 1038–1046.
79. Mordes,D., Luo,X., Kar,A., Kuo,D., Xu,L., Fushimi,K., Yu,G., Sternberg,P. Jr and Wu,J.Y. (2006) Pre-mRNA splicing and retinitis pigmentosa. *Mol. Vis.*, **12**, 1259–1271.
80. Seddon,J.M., Cote,J., Page,W.F., Aggen,S.H. and Neale,M.C. (2005) The US twin study of age-related macular degeneration: relative roles of genetic and environmental influences. *Arch. Ophthalmol.*, **123**, 321–327.
81. Menon,M., Mohammadi,S., Davila-Velderrain,J., Goods,B.A., Cadwell,T.D., Xing,Y., Stemmer-Rachamimov,A., Shalek,A.K., Love,J.C., Kellis,M., *et al.* (2019) Single-cell transcriptomic atlas of the human retina identifies cell types associated with age-related macular degeneration. *Nat. Commun.*, **10**, 4902.
82. Buniello,A., MacArthur,J.A.L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E., *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
83. Voigt,A.P., Mulfaul,K., Mullin,N.K., Flamme-Wiese,M.J., Giacalone,J.C., Stone,E.M., Tucker,B.A., Scheetz,T.E. and Mullins,R.F. (2019) Single-cell transcriptomics of the human retinal pigment epithelium and choroid in health and macular degeneration. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 24100–24107.
84. Aldiri,I., Xu,B., Wang,L., Chen,X., Hiler,D., Griffiths,L., Valentine,M., Shirinifard,A., Thiagarajan,S., Sablauer,A., *et al.* (2017) The dynamic epigenetic landscape of the retina during development, reprogramming, and tumorigenesis. *Neuron*, **94**, 550–568.
85. Janecke,A.R., Liu,X., Adam,R., Punuru,S., Viestenz,A., Strau,V., Laass,M., Sanchez,E., Adachi,R., Schatz,M.P., *et al.* (2021) Pathogenic STX3 variants affecting the retinal and intestinal transcripts cause an early-onset severe retinal dystrophy in microvillus inclusion disease subjects. *Hum. Genet.*, **140**, 1143–1156.
86. Kakakhel,M., Tebbe,L., Makia,M.S., Conley,S.M., Sherry,D.M., Al-Ubaidi,M.R. and Naash,M.I. (2020) Syntaxin 3 is essential for photoreceptor outer segment protein trafficking and survival. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 20615–20624.
87. Giovannone,A.J., Winterstein,C., Bhattaram,P., Reales,E., Low,S.H., Baggs,J.E., Xu,M., Lalli,M.A., Hogenesch,J.B. and Weimbs,T. (2018) Soluble syntaxin 3 functions as a transcriptional regulator. *J. Biol. Chem.*, **293**, 5478–5491.
88. Imamura,M., Takahashi,A., Matsunami,M., Horikoshi,M., Iwata,M., Araki,S.I., Toyoda,M., Susarla,G., Ahn,J., Park,K.H., *et al.* (2021) Genome-wide association studies identify two novel loci conferring susceptibility to diabetic retinopathy in Japanese patients with type 2 diabetes. *Hum. Mol. Genet.*, **30**, 716–726.
89. Bhatwadekar,A.D., Shughoury,A., Belamkar,A. and Ciulla,T.A. (2021) Genetics of diabetic retinopathy, a leading cause of irreversible blindness in the industrialized world. *Genes*, **12**, 1200.
90. Uribe,M.L., Martín -Nieto,J., Quereda,C., Rubio- Fernández,M., Cruces,J., Janssen,G.M.C., de Ru,A.H., van Veelen,P.A. and Hensbergen,P.J. (2021) Retinal proteomics of a mouse model of dystroglycanopathies reveals molecular alterations in photoreceptors. *J. Proteome Res.*, **20**, 3268–3277.
91. Zhang,C., Zhang,Q., Wang,F. and Liu,Q. (2015) Knockdown of poc1b causes abnormal photoreceptor sensory cilium and vision impairment in zebrafish. *Biochem. Biophys. Res. Commun.*, **465**, 651–657.

92. Birtel,J., Eisenberger,T., Gliem,M., Müller,P.L., Herrmann,P., Betz,C., Zahnleiter,D., Neuhaus,C., Lenzner,S., Holz,F.G., *et al.* (2018) Clinical and genetic characteristics of 251 consecutive patients with macular and cone/cone-rod dystrophy. *Sci Rep.*, **8**, 4824.
93. Kantardzhieva,A., Gosens,I., Alexeeva,S., Punte,I.M., Versteeg,I., Krieger,E., Neefjes-Mol,C.A., den Hollander,A.I., Letteboer,S.J., Klooster,J., *et al.* (2005) MPP5 recruits MPP4 to the CRB1 complex in photoreceptors. *Invest. Ophthalmol. Vis. Sci.*, **46**, 2192–2201.
94. Wei,X. and Malicki,J. (2002) nagie oko, encoding a MAGUK-family protein, is essential for cellular patterning of the retina. *Nat. Genet.*, **31**, 150–157.
95. Dönmez,G., Hartmuth,K. and Lührmann,R. (2004) Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA*, **10**, 1925–1933.
96. Hirata,A., Suzuki,T., Nagano,T., Fujii,D., Okamoto,M., Sora,M., Lowe,T.M., Kanai,T., Atomi,H., Suzuki,T., *et al.* (2019) Distinct modified nucleosides in $tRNA^{Trp}$ from the hyperthermophilic archaeon thermococcus kodakarensis and requirement of tRNA $m^2G10/m^2{}_2G10$ methyltransferase (Archaeal Trm11) for survival at high temperatures. *J. Bacteriol.*, **201**, e00448-19.
97. Ginell,S.L. and Parthasarathy,R. (1978) Conformation of N^2-methylguanosine, a modified nucleoside of tRNA. *Biochem. Biophys. Res. Commun.*, **84**, 886–894.
98. Wang,C., Ulryck,N., Herzel,L., Pythoud,N., Kleiber,N., Guérineau,V., Jactel,V., Moritz,C., Bohnsack,M.T., Carapito,C., *et al.* (2023) N^2-methylguanosine modifications on human tRNAs and snRNA U6 are important for cell proliferation, protein translation and pre-mRNA splicing. *Nucleic Acids Res.*, **51**, 7496–7519.
99. Scotti,M.M. and Swanson,M.S. (2016) RNA mis-splicing in disease. *Nat. Rev. Genet.*, **17**, 19–32.
100. Tanackovic,G., Ransijn,A., Thibault,P., Abou Elela,S., Klinck,R., Berson,E.L., Chabot,B. and Rivolta,C. (2011) PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Hum. Mol. Genet.*, **20**, 2116–2130.

Received: August 26, 2023. **Revised:** December 13, 2023. **Editorial Decision:** December 14, 2023. **Accepted:** December 21, 2023