OXFORD UNIVERSITY PRESS | Nucleic Acids Research

Modulating protein production from nonsense-mediated decay targets by SUZ domain-containing proteins.

SCHOLARONE™ Manuscripts

Modulating protein production from nonsense-mediated decay targets by SUZ domain-containing proteins.

Mathias Halbout[1,2], Marina Bury[1,2], Aoife Hanet[1], Isabelle Gerin[1,2], Julie Graff[1,2], Theodore Killian[3], Laurent Gatto[3], Didier Vertommen[4], Guido Bommer[1,2*]

[1] Department of Physiological Chemistry, de Duve Institute, UCLouvain, 1200 Brussels, Belgium

[2] WELBIO, 1200 Brussels, Belgium

[3] Computational Biology Laboratory, de Duve Institute, UCLouvain, 1200 Bruxelles, Belgium

[4] Protein Phosphorylation Unit, de Duve Institute, UCLouvain, 1200 Brussels, Belgium

* To whom correspondence should be addressed. Tel: +32-2 7647568; Fax: +32-2-7647598; Email: guido.bommer@uclouvain.be

ABSTRACT (200 words)

Many transcripts are targeted by nonsense-mediated decay (NMD) leading to their degradation and the inhibition of their translation. We found that several proteins containing a SUZ domain interact with the key NMD factor UPF1, and can increase protein levels from NMD targets. In some instances, this occurs predominantly by relieving translational repression.

Modulation of SUZ RNA binding domain protein 1 (SZRD1) levels changed the abundance of several NMD targets, including some that code for NMD machinery components. This revealed two competing aspects of SZRD1 function: inhibition of NMD of SZRD1-bound transcripts and overall support of NMD by increasing the abundance of a NMD machinery component. The combination of these competing effects may help regulate protein production from a subset of NMD targets.

Interestingly, the other human SUZ domain-containing proteins - namely cAMP-regulated phosphoprotein 21 (ARPP21), R3H-domain containing protein 1 and 2 (R3HDM1, R3HDM2) – show a reciprocal tissue expression pattern compared to SZRD1. We propose that this is reinforced by miR-128, which targets SZRD1 and is produced from the same primary transcripts as ARPP21 and R3HDM1. Taken together, we show that fine-tuning of protein expression from distinct subsets of NMD targets could be achieved by SUZ domain-containing proteins.

INTRODUCTION

Nonsense-mediated decay (NMD) was initially discovered as a pathway leading to the degradation of transcripts with premature stop codons generated by somatic mutations or by non-productive recombination of immunoglobulin loci in eukaryotic cells (1-3). However, it was later revealed that this pathway not only degrades defective transcripts with premature stop codons, but also a large number of physiological transcripts with particular architectural characteristics such as a long 3'UTR or upstream open-reading frames (4-7). This means that NMD not only serves as a quality control function but is also a way to regulate the expression of a subset of genes (8,9).

Recognition of premature stop codons relies in part on exon junction complexes (EJC) containing the core components eIF4A3, RBM8A, MLN51, and MAGOH, as well as several peripheral proteins such as UPF3A, UPF3B and RNPS1 (10). These protein complexes are deposited concomitantly with splicing and are normally removed during the first round of translation (11). Yet, when premature stop codons are present more than 50 nucleotides upstream of an exon-exon junction, the EJC is not removed during translation (12). In this pathway, the eukaryotic release factor eRF3 within the terminated ribosome recruits the helicase UPF1 (Up-frameshift 1) leading to the assembly of the so-called SURF complex, consisting of the kinase **S**MG1 (Suppressor with morphogenetic effect on genitalia protein 1), **U**PF1 and the eukaryotic ribosome release factors e**RF**3 and e**RF**1 (13). Phosphorylation of UPF1 represents a central event in NMD and converts the SURF complex into the decay-inducing complex (DECID) (13). This transition can be facilitated by the helicase DHX34, which binds to both SMG1 and UPF1 (14,15). This binding leads to conformational changes that allow the recruitment of UPF2 and additional phosphorylation of UPF1 by SMG1 (15). Close vicinity between a terminated ribosome and an EJC facilitates the UPF1-UPF2 interaction, since UPF2 binds to the EJC component UPF3B (16), However, it has been shown that UPF2 and UPF3B can also interact directly with eRF3 independently of the presence of an EJC (17). This allows initiation of NMD in the absence of an EJC, which can be observed on transcripts with very long 3'UTR (4,5), albeit with a lower efficiency of target mRNA degradation (18,19). In this constellation, the terminated ribosome is localized at a distance from the polyA tail bound by PABPC1s (polyA binding protein cytoplasmic 1). This distance is expected to impair rapid removal of the ribosome, and suggest that timely removal of terminated ribosomes is a critical factor in preventing NMD. Yet, reality might be more complicated than this, since recent work revealed that ribosome stalling at termination codons is not a prerequisite for NMD (20). Furthermore, recent evidence indicates that NMD does not only occur during the first round of translation: a study relying on the analysis of protein translation of single transcripts revealed that the presence of architectural characteristics (e.g. long 3'UTRs or premature termination codons) increases the probability of degradation by NMD but that several rounds of translation are usually required (21).

UPF1 and its phosphorylation play a central role in triggering transcript degradation during NMD (13). Phosphorylated UPF1 serves as a binding platform for several effectors that contribute to the degradation and translational inhibition of NMD target transcripts, like the deadenylase complex CCR4-NOT complex and the endonuclease SMG6 (22-25). Degradation of transcripts is the main consequence of NMD. However, several lines of evidence show that NMD can also lead to an inhibition

of mRNA translation. Phosphorylated UPF1 has been found to interact with components of the eukaryotic translation initiation factor eIF3 (likely via eIF3a) thereby preventing the assembly of the 80S ribosome (26). Furthermore, the CCR4-NOT complex is not only a well-known deadenylase complex but also has been described as a translational repressor (27). Deadenylation of transcript by CCR4-NOT exonuclease leads to the 3'-5' degradation of mRNA; thus, leading to a decrease in protein production. In addition, it has been demonstrated that the recruitment of the CCR4-NOT complex to transcripts can also lead to targeted translational repression independently of mRNA degradation. In fact, several laboratories have demonstrated that miRNA-induced translation repression of mRNAs occurs via the recruitement of the CCR4-NOT complex (28-30). Of note, there are likely several other mechanisms that can be employed to induce targeted translational repression, for example through the cap-binding protein 4E-T (31,32).

Beyond UPF1, the requirement of individual NMD factors is much less clear. For example, some subsets of NMD targets are dependent on UPF2 whereas others require UPF3B (33,34). It is currently unclear what determines whether a transcript's degradation is preferentially dependent on UPF2 or UPF3B. Furthermore, we are only beginning to understand how NMD is regulated. Some *cis*-elements and RNA-binding proteins have been reported to influence whether a transcript with a long 3'UTR can evade NMD or not (35-37). In addition, NMD seems to be auto-regulated since transcripts coding for UPF1, UPF3B and SMG1 have all been described as NMD targets (38-40). Additionaly, PABPC1 has been reported to block NMD activity via his interaction with eIF4G and eRF3 (26,41,42). Interestingly NMD targets show a lower translational efficiency in comparison to non-NMD targets likely due to lower translational initiation and elongation efficiency (43). Yet, it is unclear whether this is causing NMD or whether this is a consequence of the occurrence of NMD.

NMD is not only regulated by proteins. It has been reported that miR-128 can modulate NMD by targeting the two NMD machinery components UPF1 (44) and UPF3B (45), as well as the EJC component MLN51/CASC3 (44). Some data indicated that upregulation of miR-128 during brain development *in vivo* and neuronal differentiation *in vitro* contributes to decreased UPF1 and MLN51 protein levels (44). This change correlated with the upregulation of a series of transcripts that are predicted to be targeted by NMD (44). Inhibition of miR-128 affected neuronal differentiation, suggesting that the inhibition of NMD by miR-128 might play a role in neuronal differentiation.

The miR-128 family is conserved in mammals, fish and birds (46). In the human genome, two genomic loci produce mature miR-128 (miR-128-1 on human chromosome 2 and miR-128-2 on chromosome 3, henceforth referred to as miR-128). In both cases, the miR-128 precursor sequence is localized in an intron of transcripts coding for two related proteins, either R3HDM1 or ARPP21 (47). In mammals, miR-128 is highly expressed in the brain, skeletal muscle and thymus (48). Both tumor-suppressive (49-56) and oncogenic functions have been described for miR-128 (57-59). Genetic inactivation of miR-128-2 and miR-128-1 in mice led to a reduction of miR-128 levels in the brain of 80% and 20%, respectively (60). Phenotypically, the profound reduction of miR-128-2 in miR-128-2 knockout mice led to hyperactivity and lethal epilepsy at 2-3 months of age in mice (60). The underlying increased neuronal excitability has been attributed in part to the activation of the ERK2 branch of MAP kinase

signaling (60), while no obvious changes in NMD were reported (60). Mechanistically, the excitability was suggested to be caused by the proteins PEA15A and SZRD1, which were found to be increased in miR-128-2 knockout neurons and have been reported to play a role in modulating MAP kinase signaling, albeit in a very context-dependent manner. For example, PEA15A facilitates ERK1/2 activation (61) and allows ERK1/2 to efficiently activate the stress kinase RSK2 (62), but it inhibits MAP kinase signaling in other situations (63). The picture is even less clear in the case of SZRD1, which – when overexpressed – can strongly activate a MAP kinase-dependent reporter gene construct (64) and modulate a wide-range of signaling pathways (65). However, how SZRD1 might exert these functions is unclear.

We were intrigued by the protein SZRD1 because it not only seemed to play a role in miR-128 function (60), but also because it contains a so-called SUZ domain that is otherwise only present in two paralogs of proteins produced from the same primary transcript as miR-128 (47) (i.e. ARPP21 and R3HDM1), as well as a paralog called R3HDM2. This indicated that transcriptional activation of the primary transcripts of miR-128 might lead to the coordinated decrease of one SUZ domain-containing protein (i.e. SZRD1) and the increase of another SUZ domain-containing protein (i.e. ARPP21 and R3HDM1). ARPP21, R3HDM1 and R3HDM2 contain an R3H domain and a SUZ domain. The R3H domain is found in many proteins and has been described to mediate the interaction with RNA molecules in a non-sequence-specific manner (66,67). In contrast, the SUZ domain is only present in ARPP21, R3HDM1, R3HDM2 and SZRD1. It is approximately 70 amino acids long and has first been described in the *Caenorhabditis elegans* (*C.elegans*) protein SZY-20, an SZRD1 homolog that plays a role in centrosome function. Based on the interaction with poly-uridine-coated agarose beads (68), it was postulated that SZY-20 may bind to RNA, but this interaction was not further characterized. In addition, the protein SZRD1 contains a C-terminal SUZ-C domain, which is present in the proteins LARP6, GEMIN7 and CSDE1 (69-71) that all interact with the protein STRAP.

We decided to identify interaction partners of SZRD1, determine the relative contribution of its domains to protein-protein interactions and glean insights into SZRD1 function. We found that SZRD1 interacts with UPF1 in cells. Interestingly, changes of SZRD1 and ARPP21 can modulate protein production of NMD target transcripts. In some cases, this seems to be predominantly due to a SUZ domain-dependent relief from translational repression. Our data indicate that tissue-specific expression of SUZ domain-containing proteins represents a way to regulate protein output from transcripts targeted by NMD.

MATERIAL AND METHODS

Sequences for primers used for cloning are given in table S1; antibodies are listed in table S2; plasmids are listed in table S3; and qPCR primers are in table S4. The nucleotide sequences of the inserts in all plasmids generated in this study were verified by Sanger sequencing.

Generation of expression constructs. The open reading frame of SZRD1 was amplified using cDNA from HCT116 cells as template with the primers *SZRD1long-FW* and *SZRD1-RV* for the long forms of

SZRD1 (corresponding to NP_001108072.1 ('+CAG') or NP_001258798.1 ('-CAG')), and the primers *SZRD1short-FW* and *SZRD1-RV* for the short forms of SZRD1 (corresponding to AAH23988.1 ('+CAG') or BAC77397.1 ('-CAG')) to generate constructs driving the expression of proteins with a C-terminal SFB tag (consisting of S peptide, FLAG tag and streptavidin-binding peptide). The resulting PCR products were digested with the restriction endonucleases NheI and BsrGI, and inserted into the corresponding sites of the lentiviral vector pGTB5002 (see table S3) in frame with an SFB tag (72,73). Constructs driving the expression of untagged versions of different SZRD1 forms were generated by insertion of PCR products with endogenous stop codons obtained with the reverse primer *SZRD1-STOP-RV* into the same vector. Mutant forms of SZRD1 were obtained by PCR-driven overlap extension (74). The flanking primers were the same used to amplify the wild-type forms, and the internal primers are listed in table S1.

The open reading frame of human UPF1 was amplified from cDNA obtained from HCT116 cells using the primers *UPF1-FW* and *UPF1-RV* (corresponding to NP_002902.2). The resulting PCR product was digested and inserted into the vector pGTB5002 using the restriction enzymes NheI and BsrGI. The open reading frame of mouse STRAP was amplified from mouse spleen cDNA with the primers *STRAP-FW* and *STRAP-RV* (corresponding to NP_035629.2). The open reading frame of human R3HDM1 and mouse ARPP21 were amplified from EST clones (R3HDM1, Id:5556247; ARPP21 Id: 5686446; GE Healthcare) using the primers *R3HDM1-FW* and *R3HDM1-RV* (corresponding to NP_056176.2) or the primers *ARPP21-FW* and *ARPP21-RV,* respectively. The resulting ARPP21 PCR product lacked part of the SUZ due to the alternative splicing event, and represents a minor splice variant (according to a large-scale RNAseq datasets) (75). An expression construct for the most common isoform (corresponding to NP_001348953.1) was generated by inserting the missing exon by Gibson assembly using the gene block gb-ARPP21SUZ. The PCR products obtained for STRAP, R3HDM1 and ARPP21 ORFs were digested with NheI and Acc65I, and inserted into the NheI and BsrGI sites of the vector pGTB5002.

The R3HDM2 ORF (NP_001338140.1) was amplified from cDNA of CCRF/CEM cells in two overlapping PCR products using the primer pair R3HDM2Nt-FW and R3HDM2Nt-RV to amplify the N-terminal part, and the primer pair R3HDM2Ct-FW and R3HDM2Ct-RV to amplify the C-terminal part. The two PCR products were cloned into the vector pJET 1.2-blunt (Fermentas), verified by sequencing and then connected by fusion PCR using the primer pair R3HDM2full-FW and R3HDM2full-RV. The resulting PCR product was digested and inserted into the NheI and NotI restriction enzyme sites of the vector pGTB5002.

Vectors used for transient expression of lambdaN-tagged proteins were based on the pCDNA3.1 vector (Life Technologies). The lambdaN tag was inserted by ligating the annealing oligos *lambdaN-FW* and *lambdaN-RV* into the NheI site resulting in the plasmid pJG211. Then, the SZRD1 short and long ORF were taken from pAH47 and pAH45, respectively, using the restriction enzymes NheI-BsrGI and inserted into pJG211 plasmid. The same procedure was followed to insert the mutated forms of SZRD1 that are described above. ARPP21 and R3HDM1 were amplified from pMH117 and pMH50, respectively using the primers *lambdaNARPP21-FW* and *lambdaNARPP21-RV* for ARPP21 and *lambdaNR3HDM1-FW* and *lambdaNR3HDM1-RV* for R3HDM1. For the amplification of the ARPP21

ORF containing the deletion in the SUZ domain, we used the plasmid pOH463 (see table S3) as a template and amplified the ORF with the same primers as for full length ARPP21. The PCR products were digested by NheI and Acc65I and inserted into corresponding sites in pJG211 . R3HDM2 was amplified from the plasmid pMH130 (see table S3) with the primers *lambdaNR3HDM2-FW* and *lambdaNR3HDM2-RV*, digested by NheI and NotI and inserted into corresponding sites in the plasmid pJG211.

Generation of lentiviral shRNA constructs. To express shRNAs, we used the plasmid pLVXpuro (Clontech), in which we inserted an EGFP expression cassette followed by a shRNA expression cassette that closely resembles the improved miR-30 expression scaffold miRE (76) to generate the plasmid pJG97. shRNA sequences were identified using the splashRNA tool (77). All shRNAs cassettes were amplified from synthetic oligonucleotides (IDT) using the primers *shRNA-FW* and *shRNA-RV* and inserted into the plasmid pJG97 via the XhoI and EcoRI sites. A nonsilencing control shRNA was shuttled from the vector pGIPZ-negative (Openbiosystem). The sequences of oligonucleotide templates are given in table S1. With regard to SZRD1 shRNAs, #1 and #2 target the 3'UTR and therefore reduce all SZRD1 isoforms. In contrast, shRNAs #3 to #6 target exon 2, which is present only in the long isoforms of SZRD1. To generate inducible shRNA expression constructs, shRNAs were amplified from the synthetic oligonucleotide corresponding to shRNA #1 (CAGTGCCAGCAATAACAGT, see table S1) or from the vector pGIPZ-negative with the primers *miR30PCRXhoIF_s* and *miR30PCREcoRIF_rev,* and inserted into the XhoI and EcoRI sites of the plasmid pTRIPZ-empty (Openbiosystems).

Generation of reporter gene constructs. A reporter construct driving expression of a *Renilla* luciferase gene fused on its C-terminus with part of the wild type human beta-globin sequence (Renilla-HBB wt) or a version containing a premature stop codon (Renilla- Hbb NS39) was a kind gift from Jana Loeber and Andreas Kulozik (78). Reporter constructs that allow recruitment of lambdaN-tagged proteins contain destabilizing elements and/or miRNA-binding sites were produced similar to what is described by Gehring and colleagues (79). Briefly, a synthetic linker (BoxB adaptor) was inserted using the NotI site. A total of six BoxB sites were inserted by three consecutive insertions of the *BoxB genblock* (containing 2 BoxB sites) into a single MluI site. In addition, a poly A/U rich site sequence (ARE) (*ARE-MluI*) or its control (*AREctrl-MluI*) were inserted downstream of the BoxB sites using the MluI restriction enzyme. To assess the effects of miRNAs, four complementary or seed-mutated let-7a binding sites were inserted in the same location as the ARE sequence. These sequences were obtained by PCR from a plasmid kindly provided by Thomas Michiels from UCLouvain, Belgium (unpublished data) using the primers *pADC-miRs-FW and* pADC-miRs-*RV*.

The sequence corresponding to the region of the SZRD1 3'UTR containing miR-128 binding sites (between 115 nt and 210 nt after the stop codon) was generated by fusing multiple combinations of overlapping primer pairs containing either wild type or mutant mir-128 binding sites (see table S1), and amplified with the primers *miR-128-FW* and *miR-128-RV* allowing insertion into the XbaI and EcoRI sites of the plasmid pGL3 (see table S3) control. The sites predicted to interact with miR128 were modified from 5'-CACUGUGA-3' to 5'-CTCAGAGA-3'. We further analyzed constructs containing the combinations of wild type and mutant sites depicted in figure 1C.

Insertion of an SFB-tag into the UPF1, eIF4A3 and RNPS1 loci ('knock-in'). To generate CRISPR/Cas9 constructs cutting immediately downstream of the open reading frames, we inserted annealed primer pairs *UPF1-guide*, *RNPS1-guide* or *eIF4A3-guide* into the BbsI site of the vector pX330 (Addgene # 42230, a kind gift from Feng Zheng, MIT, Cambridge, MA, USA) (80).

To select for correct insertions, we generated a cassette consisting of an SFB-tag sequence, a self-cleaving P2A site and a puromycin resistant gene, similar to what has been described by Sheridan and Bentley (81), into the pGolden-AAV vector (Addgene #51424, a kind gift from Younglun Luo) (82) generating the vector pJG180. Homology arms containing the sequences immediately upstream and downstream of the stop codons were amplified from genomic DNA of the U2OS cell line with the primer pairs indicated in table S1 (*UPF1HL-FW* and *UPF1HL-RV* for the left homology arm of UPF1; *UPF1HR-FW* and *UPF1HR-RV* for the right homology arm of UPF1; *RNPS1HL-FW* and *RNPS1HL-RV* for the left homology arm of RNPS1; *RNPS1HR-FW* and *RNPS1HR-RV* for the right homology arm of RNPS1; *eIF4A3HL-FW* and *eIF4A3HL-RV* for the left homology arm of eIF4A3; *eIF4A3HR-FW* and *eIF4A3HR-RV* for the right homology arm of eIF4A3). The selection cassette was amplified with the primers *SFB-P2A-puro-FW* and *SFB-P2A-puro-RV* using the vector pJG180 as a template. Subsequently, homology arms and selection cassettes were fused and inserted into the vector pGolden-AAV using Gibson assembly.

To generate helper-free recombinant adeno-associated virus we transiently transfected the above mentioned donor vectors together with the plasmids pHelper and pAAV-RC into HEK293 cells (Agilent #240071) using calcium phosphate precipitation. After 6 hours, the medium was changed and 72 hours later medium and cells were recovered and viruses were liberated by 3 freeze-thaw cycles.

HEK293 were transfected with CRISPR/Cas9 constructs at 70% confluence in 6-well plates using Lipofectamine 2000 (2 μg of ADN for 6 μl of Lipofectamine 2000). After six hours, the medium was changed and 1/2 of the AAV supernatant was added. 72 hours later, cells were plated in two 10 cm dishes with 0.5 μg/ml of puromycin. Single-cell clones were isolated by limiting dilution. Insertion of the SFB-tag was assessed by Western blot and by sequencing of the insertion site (Fig. S6A and B).

Cell culture, lentiviral transduction. Human embryonic kidney cells HEK293, HEK293-T, U2OS and HeLa were cultured in DMEM medium containing 4.5 g/l D-glucose, 10 % fetal calf serum, 2 mM Ultraglutamine I (Lonza), and 100 U/ml Penicillin/Streptomycin (Lonza) at 37°C and 5% CO_2. Jurkat cells were cultured in RPMI-1640 medium (Lonza: 12-702F), 10% fetal calf serum, 2 mM Ultraglutamine I (Lonza), and 100 U/ml Penicillin/Streptomycin (Lonza) at 37°C and 5% CO_2. Jurkat cells were maintained between 0.2 and 1.5 million cells/ml. To inhibit proteasomal degradation cells were treated with 10 μM MG132 (Milipore : 133407-82-6) for the indicated times.

To generate recombinant lentiviruses used for the stable overexpression or knockdown of proteins, we transiently transfected HEK293-T cells with lentiviral vectors and packaging plasmids using calcium phosphate precipitation (83). Briefly, HEK293-T cells were transfected at 80 % confluence in 6 cm dishes with 8.4 μg of psPAX2 (Addgene #12260) and 4.2 μg of pMD2.G (Addgene #12259), both kind gifts from Didier Trono (University of Geneva, Switzerland), together with 8.4 μg of the lentiviral vectors described above. Medium was changed 6 h after transfection, and 24 h later tissue culture supernatants containing viruses were harvested, filtered through a 0.4 μm filter and diluted to infect target cells in the

presence of polybrene (8 µg/ml; Sigma). After 24 h, cells were subcultured and infected cells were selected with 2 µg/ml of puromycin for 2 days or until all cells on a non-infected control plate were dead. In cases, were two lentiviruses were used to infect cells, vectors were generated containing a hygromycin resistance cassette but were otherwise identical. This allowed a parallel selection with 300 µg/ml of hygromycin and 2 µg/ml puromycin.

Transfection of miR-mimics and LNA-miR-inhibitors. We transfected pre-MIRs (AMBION) with Lipofectamine 2000 (10 pmol of pre-MIR per µl of Lipofectamine) at a final concentration of 5 nM, in HEK293, HEK293-T and HeLa cells following the manufacturer's recommendations (Life Technologies). After 48 h, cells were washed with cold PBS and harvested using the RIPA buffer [50nMTris pH 7.4, 150 mM NaCl, 0.1 % SDS, 0.5 % sodium deoxycholate and 1% Igepal 640] with phosphatase inhibitors [5 mM EDTA, 10 mM NaF, 1 mM orthovanadate pH 10.0, 2 mM sodium pyrophosphate and 2 mM β-glycerophosphate] and a protease inhibitor cocktail (cOmplete: Roche).

miRCURY LNA miR inhibitors (Qiagen) were electroporated into the cell line JURKAT using the Nucleofector 4D electroporation device (Lonza) following the AMAXA protocol V4XC-1012. After 48h, transfected cells were washed with cold PBS, pelleted by centrifugation and resuspended in RIPA buffer as described above.

Affinity purification via streptavidin-coated beads. Cells growing in 10 cm plates were washed twice with ice-cold PBS and harvested in 300 µl NETN lysis buffer [150 mM NaCl, 1 mM EDTA, 20 mM Tris-HCl pH8.0 and 1 % NP40] containing phosphatase inhibitors [10 mM sodium fluoride, 1 mM sodium orthovanadate pH 9.0, 2 mM sodium pyrophosphate and 2 mM β-glycerophosphate] and protease inhibitors (cOmplete: Roche). Lysates were incubated on ice for 10 min and briefly sonicated to shear genomic DNA. Subsequently, lysates were clarified by centrifugation at 27,000 g for 30 min at 4°C and the supernatants were recovered. We measured protein concentrations with the bicinchoninic acid (BCA) assay and incubated equal amounts of proteins with streptavidin-coupled Sepharose beads(GE Healthcare: 17-5113-01) or magnetic beads (Thermo Fisher: 88817; Figure 3C – F), which had been pre-equilibrated with NETN buffer. Samples were incubated on a rotating wheel for at least one hour at 4°C. Approximately 10 µl of beads were used per mg of protein. Where indicated, 5 µl of RNAse A (Thermo scientific: EN0531) and 5 µl of RNAse T1 (Thermo scientific: EN0541) per ml of NETN buffer were added during the pulldown.

Beads were washed three times by adding 750 µl of NETN buffer and separated from the supernatant by centrifugation at 500 g for 3 min at 4°C or by exposure to a strong magnet for 30 seconds. Retained proteins were removed with denaturing loading dye and heated for 5 min at 85°C. As an input control, a fixed fraction of the amount used in the pulldown was processed in parallel, representing between 5 and 12 % as specified in the figure legends. Initial experiments were performed to identify interacting partners of SZRD1 followed by a tandem affinity purification protocol (73), where proteins retained on the streptavidin beads were eluted with 1 mg/ml biotin and loaded onto S-protein-coated agarose beads (Novagen 69704). After elution with denaturing loading dye and gel electrophoresis, gels were stained with Coomassie brilliant blue G250 (Pageblue, Thermo Scientific), bands were excised and proteins were identified after in-gel digestion with trypsin using an LTQ mass spectrometer (84).

Reporter gene assays. Cells were plated one day before transfection in 24-well plates at 75 000 cells per well for HEK293T cells and for U2OS cells, and 35 000 cells per well for HeLa cells. Transfections were performed using a total of 200 ng of DNA per well with 1 µl of Lipofectamine 2000 following the manufacturer's protocol. The total amount of DNA consisted of 50 ng of the firefly luciferase control plasmid pCLneoFirefly, 50 ng of the Renilla luciferase constructs derived from the plasmid pCLneoRenilla (wt or NS39, see Figure 4F) and 100 ng of the constructs driving the indicated effector proteins or empty vector controls. When overexpressing tethered SZRD1, ARPP21 or R3HDM2 with a lambdaN tag, we used 10ng of the pDNA3.1 constructs driving expression of these proteins and 90 ng of the empty vector (pJG211). In the case of lambdaN-tagged R3HDM1, we used 100 ng of the construct since R3HDM1 overexpression was undetectable at lower levels. After 24 hours, cells were harvested and analyzed using a commercial dual luciferase kit (Dual-luciferase, Promega #E1960) with the help of a luminescence plate reader (GloMax Discover, Promega). To assess relative expression levels of endogenous proteins and overexpressed proteins, luciferase reporter assay lysates were diluted two times with water, heated after addition of 1/6 volume of 6x concentrated denaturing loading buffer [0.415 M of SDS; 0.9 M of bromophenol blue; 60 mM of Tris pH 6.8; 47% of glycerol and 0.6 M of DTT] (85) for 10 min at 72°C, and analyzed by western blot as described below.

When luciferase mRNA levels were analyzed, a set of parallel transfections was performed in 12 well plates using two times more of each component for the transfection. RNA purification, reverse transcription and quantitative PCR were performed as described below.

When miRNA mimics were cotransfected in reporter assays, 35 000 HeLa cells per well of a 24-well were plated one day before transfection. Transfection conditions were similar to the ones described above (i.e. 50 ng pGL3-control plasmid with or without SZRD1 3'UTR, 50 ng of an SV40-promoter driven Renilla luciferase construct (Promega) and 100 ng of an empty expression vector (pJG211) to reach 200ng) except for the addition of the indicated mimics at a final concentration of 3.3 nM.

RNA extraction, reverse transcription and quantitative PCR. RNA extraction, RT-PCR and qPCR were performed as described in (86). In brief, cells were harvested with TRIzol® (Life technologies: 15596018) and RNA extracted according to the manufacturer's recommendations. Subsequently, 1.2 µg total RNA was reverse transcribed with RevertAid reverse transcriptase (Fermentas) and random hexamer primers in a total volume of 20 µl (86). The resulting cDNAs were diluted ten-fold and 5 µl were used to perform a quantitative PCR with a commercial SYBR green master mix (Takyon mix, Eurogentec) according to the protocol given by the manufacturer. Results were analyzed using the delta-delta Ct method (87). Primers used for qPCR are listed in table S4 and amplicons comprise exon-exon junctions unless indicated otherwise.

To quantify miR-128, the hairpin based TaqMan MicroRNA assays from 'Applied Biosystems' (kit: P02840360) and the corresponding reverse transcription kit were used.

Western blot analysis. Western blots were performed as described previously (84). In brief, cells were harvested using RIPA buffer or NETN buffer as described above. Cell lysates were sonicated and clarified by centrifugation for 30 min at 27,000 g at 4°C. Protein concentrations were determined using the BCA assay. Equal protein amounts were loaded on 10 % or 12 % BIS-TRIS polyacrylamide gels. Electrophoresis was performed in MOPS buffer [50 mM MOPS, 50 mM Tris, 1 mM EDTA pH 8.0 and

0.1% of SDS]. Proteins were transferred onto PVDF membranes (Immobilon P, Milipore) using a tank transfer system (Bio-Rad) and membranes were blocked with 5 % non-fat dry milk (Regilait, Delhaize, Belgium) in T-TBS buffer [0.5 % Tween-20 in 20 mM Tris-base pH 7.2, 150 mM NaCl]. Incubation with the primary antibody was performed using T-TBS buffer containing 2% bovine serum albumin (Fisher Scientific, Belgium) overnight at 4°C using the concentrations indicated in table S2. Horse-radish peroxidase-coupled secondary antibodies against mouse (A5278, Sigma; dilution 1/10 000) or rabbit (NA934V, GE Healthcare dilution 1/20 000) immunoglobulins were diluted in 5% non-fat dry milk T-TBS solution and incubated for 1 h at room temperature. Washing steps after first and secondary antibodies were performed with T-TBS solution three times for 5 min at room temperature. Signals were detected using the chemiluminescence reagent (WBKLS0500, Milipore) and autoradiography films (Fuji X-ray film) or a digital image acquisition system (LAS4000, GE Healthcare).

RNAseq analysis. For overexpression experiments, HeLa cells were infected with a lentivirus driving expression of the long SZRD1 isoform (with the addition of a serine residue, pAH45), human ARPP21 or an empty control lentivirus. After 24 hours, cells were divided into medium containing 2 μg/ml puromycin. The medium was changed to medium without antibiotic 24h later. Another 24 h later, two plates were harvested with TRIzol for RNA extraction and NETN lysis buffer for protein extraction. For knockdown experiments, we transduced HeLa cells with a recombinant lentivirus driving docycyclin-inducible expression of SZRD1 shRNA #1 from the backbone pTRIPZ –empty (Openbiosystems). As a negative control, we used constructs expressing a non-silencing control shRNA (pTRIPZ-negative, Openbiosystems). Expression of shRNAs was induced by addition of doxycyclin for 60 h before collection of cells in TRIzol.

After RNA extraction, 15 μg of RNA were incubated with 1 μl of DNAse (RQ1 RNase-Free DNase, M6101 Promega) for 30 min at 37°C with the indicated buffer and RNA was cleaned up using the Qiagen RNeasy kit (Qiagen #74104). Purity and concentrations were determined with an Agilent 2100 bioanalyzer and 1 μg was sent for RNAseq analysis (after polyA selection) to Beckman Coulter Genomics.

The quality of the raw reads was assessed with FastQC (version 0.11.8) (88). The reads were trimmed of the universal adaptor with Trimmomatic (version 0.36) (89) and aligned using hisat2 (version 2.1.0) (90) to the Genome Reference Consortium Human Build 38 reference assembly (91). The resulting aligned reads were sorted and indexed using samtools (version 1.9) (92) and counts were then generated using htseq-count (version 0.11.1) (93). The RNAseq count data were subsequently analyzed using the Bioconductor package (94) DESeq2 (version 1.26.0) (95) and volcano plots were generated with the EnhancedVolcano R package (96). Transcripts were sorted according to the t-statistic and a gene set enrichment analysis was performed using the fgsea package (97) using a list of human NMD taken from the study by (6) as well as the molecular signature database (98).

Northern blot. Northern blot was performed as previously described (99). In brief, 10 μg of total RNA was separated on a 15% polyacrylamide gel containing 8 M urea and 1x TBE buffer [90 mM of Tris-base, 90 mM boric acid and 2 mM of EDTA pH 8.0]. Electrophoresis was performed in 1x TBE buffer and transfer onto nylon membrane (Hybond N+, GE Healthcare) was performed with two-times diluted TBE in a semi-dry transfer system at 100 mA per gel for 90 minutes. RNA was crosslinked to

membranes with UV light (Stratalinker, Agilent). Probes were radioactively labeled using 30 units of T4-polynucleotide kinase (Fermentas) per 50 pmol of oligonucleotides in the presence of 50 µCi of [γ-^{32}P] ATP for 1 hour. Unincorporated nucleotides were removed using a G25 gel filtration column (GE Healthcare). Membranes were hybridized in hybridization buffer (RapidHyb, GE Healthcare) at 10^6 cpm/ml at 42 °C, overnight. After washing three times for 5 min in washing buffer [300 mM of NaCl, 30 mM of sodium citrate, 0.1% of SDS, at pH 7.0] at room temperature, the signal was exposed to a storage phosphor imaging screen and detected on a Typhoon TRIO from GE Healthcare.

RESULTS

SZRD1 protein isoforms interact with proteins involved in RNA biology

The *SZRD1* gene codes for transcripts with up to four exons. Alternative splicing events can lead to four different transcripts that are predicted to produce four different protein isoforms (Fig. 1A & B). Inclusion or exclusion of exon 2 results into long and short isoforms of SZRD1. The long and short isoforms have start codons in different open reading frames convergeing in the same open reading frame from exons 3 and 4. Thus, these isoforms differ in their N-terminal part, encoded by exon 1 and 2, and are identical in the C-terminal part, encoded by exon 3 and 4. The long isoform is conserved in a wide-range of organisms down to *C. elegans* (Fig. S1A,B), whereas the short form is restricted to a limited number of vertebrates. Additional complexity of SZRD1 transcripts is caused by the inclusion of three additional nucleotides (CAG) on the 5' end of exon 3 during splicing in about half of the transcripts. This leads to the insertion or exclusion of one additional amino acid, an arginine in the short form of SZRD1 and a serine in the long form of SZRD1 (Fig. 1B).

All four transcripts are present at comparable levels in human tissues (75). Yet, it was unclear whether these isoforms were translated into proteins. This particularly concerns the short isoforms since productive translation of these isoforms is predicted to start with a downstream start codon, whereas usage of the first start codon of these transcripts would be expected to lead to a degradation by NMD. We therefore assessed endogenous SZRD1 protein levels with a polyclonal antibody expected to recognize both long and short isoforms. We observed two distinct bands that migrated at an apparent molecular weight of 20 and 23 kDa (Fig. 1C), close to 15 and 17 kDa predicted for the short and long isoforms, respectively. Next, we wanted to test whether these two bands were the consequence of the inclusion or not of exon 2. To this end, we knocked down either all SZRD1 isoforms using two different shRNAs targeting the last common exon (Fig. 1C, #1 and #2) or only the long isoform using four different shRNAs targeting exon 2 of SZRD1 (Fig. 1C, #3 to #6) in HEK293 cells. We observed that shRNAs #1 and #2 strongly decreased both bands, confirming that both bands are SZRD1 proteins. In contrast, shRNAs targeting exon 2 only decreased the band of higher predicted molecular weight with varying efficiency, whereas the lower bands remained unaffected. Of note, none of the shRNAs affected levels of the loading control β-actin. We conclude that the upper band corresponds to long SZRD1 protein isoforms produced from transcripts containing exon 2, whereas the lower band corresponds to short SZRD1 protein isoforms that are produced from transcripts that do not contain exon 2.

Very little is known about the function of SZRD1 and its isoforms. To glean insights into the function of SZRD1 proteins, we first decided to identify interacting protein partners. We overexpressed SZRD1 proteins with a C-terminal SFB-tag consisting of the S-peptide, a FLAG tag and the streptavidin-binding peptide (73,100). SZRD1 and interacting proteins were purified in a tandem affinity purification consisting of binding to streptavidin-coated Sepharose beads, eluting with biotin and binding of eluted proteins to S-protein-coated agarose beads. Bands were excised and interacting proteins were identified by mass spectrometry (data not shown).

Subsequently, we focused on putative interaction partners with predicted molecular weights corresponding to the excised band sizes, and validated their interaction by performing streptavidin-bead pulldowns followed by western blot analysis (Fig. 1D). We confirmed that UPF1 (up-frameshift 1), STRAP (Serine/threonine kinase receptor-associated protein), PABPC1 (Polyadenylate-binding protein) and C1QBP (Complement component 1 Q subcomponent-binding protein) co-purified with SFB-tagged SZRD1. These proteins were retained on streptavidin-beads in cell lines expressing SFB-tagged SZRD1 isoforms, but were absent in pulldowns performed in cell lines expressing an empty vector control (Fig. 1D). Expression levels of different SFB-tagged SZRD1 proteins were similar between cell lines facilitating the comparison of interaction efficiency between SZRD1 isoforms. UPF1 and PABPC1 interacted more readily with the short form of SZRD1, and C1QBP interacted with the short form of SZRD1 containing the arginine insertion. Of note, we consistently observed that overexpression of SZRD1 led to an increase of UPF1 protein levels (Fig. 1D & E), which was further investigated later in this work.

Taken together, at this stage we concluded that SZRD1 proteins in cells have some common (like UPF1, STRAP and PABPC1) and some isoform-specific interaction partners (e.g. C1QBP with the short form), and that these interactions vary depending on the insertion or not of the extra CAG at transcript level. Intriguingly, the confirmed interaction partners have known functions in RNA metabolism: UPF1 is known to play a central role in NMD among other functions (101). STRAP is implicated in different aspects of RNA metabolism (69-71). PABPC1 interacts with the polyA tail of mRNAs (102) and C1QBP is a multifunctional protein with a role in RNA splicing and homologous recombination (103-106).

To investigate whether the interaction of SZRD1 with these proteins is dependent on the presence of intact RNA, we purified SFB-tagged SZRD1 proteins from cells and treated the cell lysates or not with a mixture of RNAse A and RNAse T1. To simplify the experimental setup, we focused on the long and short isoforms of SZRD1 carrying the serine and arginine insertion, respectively (Fig. 1B. '+CAG'). In the presence of RNAses, PABPC1 and C1QBP were co-purified with SZRD1 to a weaker extent (Fig. 1E + vs – RNAse A/T1). In contrast, the amounts of STRAP remained the same and the amount of interacting UPF1 protein was increased upon RNAse A/T1 treatment (Fig. 1 E). This indicated that the interactions of SZRD1 with PABPC1 and C1QBP might be indirect via an RNA scaffold, whereas the interactions with UPF1 and STRAP are likely independent of the presence of RNA in cells.

Our initial experiments demonstrated that epitope-tagged SZRD1 interacts with endogenous UPF1 and STRAP. To complement these data, we also performed the reverse experiment where we overexpressed UPF1 and STRAP with an N-terminal SFB tag (Fig. 1F & G). In these experiments, we

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

recovered endogenous SZRD1 only in cells that express epitope-tagged UPF1 (Fig. 1F) or STRAP (Fig. 1G), but not in cell lines transduced with an empty vector ('control'). In all these experiments, UPF1 was preferentially recovered with the short isoform, whereas STRAP was pulled down with both isoforms (Fig. 1D-F). Taken together, our experiments revealed that SZRD1 interacts with STRAP and UPF1 in an RNA-independent manner in cells.

The SUZ domain is required for the interaction with UPF1 and the SUZ-C domain is required for the interaction with STRAP.

Next, we investigated to what extent the conserved SUZ and SUZ-C domains in SZRD1 are required for its protein-protein interactions (Fig. 2A & B). The SUZ-C domain is approximately 30 amino acids long, and, besides SZRD1, the human genome codes for three additional proteins with this domain (GEMIN7, CSDE1 and LARP6) (Fig. 2A). Interestingly, all three proteins have been described to interact with STRAP (69-71). Hence, we hypothesized that the SUZ-C domain would be required for the interaction between SZRD1 and STRAP. To test this, we overexpressed SFB-tagged mutant SZRD1 proteins in HEK cells. We decided to either delete the entire SUZ-C domain ('ΔC') or to replace three conserved amino acids ('GPD') by three alanine residues ('AAA') (see Fig. 2A). Affinity purifications with streptavidin-beads were performed and analyzed by western blot. This revealed that only wild-type SZRD1 co-purified with STRAP, whereas the deletion of the SUZ-C domain and the indicated three amino acid substitution abolished the interaction with STRAP (Fig. 2C & D). Of note, mutations in the SUZ-C domain abolished the interaction with STRAP but did not perturb the interaction with UPF1 (Fig. 2C and D). These experiments demonstrate that the SUZ-C domain of SZRD1 is required for its interaction with STRAP in cells but is dispensable for the interaction with UPF1.

The SUZ domain comprises approximately 70 amino acids and has been suggested to be involved in RNA binding (68). Besides SZRD1, the human genome contains three other loci coding for proteins with a SUZ domain (ARPP21, R3HDM1, R3HDM2; Fig. 2B), two of which code miR-128 from an intron. We decided to test the requirement of the SUZ domain of SZRD1 for the interactions with STRAP or UPF1. To do this, we overexpressed SFB-tagged SZRD1 proteins carrying mutations of conserved amino acids in the SUZ domain. We generated four different mutations (Fig. 2B): two deletions of four amino acids ('Δ1' and 'Δ2') and two single amino acid substitutions ('Y→F and R→Q'). We then analyzed whether these proteins interacted with UPF1 and STRAP using affinity purification with streptavidin beads followed by western blot analysis. In this case, mutations in the SUZ domain did not affect the interaction with STRAP indicating that the SUZ domain of SZRD1 is dispensable for its interaction with STRAP. In contrast, deletion of the conserved sequence ILKR ('Δ1') abolished the interaction of the long and short SZRD1 isoforms with UPF1 (Fig. 2E & F). Mutations in the C-terminal part of the SUZ-domain reduced the interaction of the long SZRD1 isoform, whereas the interaction with the short SZRD1 isoform was less affected (Fig. 2E). As already noted in Fig. 1D & E, we observed a consistent increase in UPF1 protein levels (Fig. 2C-F) when SZRD1 was overexpressed. Interestingly, the introduction of mutations in the SUZ domain of SZRD1 abolished this increase in UPF1 protein levels (Fig. 2E&F), although only deletion of the ILKR motif (i.e. mutation Δ1) abolished interaction with UPF1. Overall, our data indicated that the four residues ILKR in the SUZ domain are required for the

interaction between SZRD1 and UPF1, and that the SZRD1 via its SUZ domain might somehow affect UPF1 protein levels.

It had previously been reported that UPF1 and STRAP interact with each other (107). Given that both proteins also interacted with SZRD1 via separate, distinct domains, we considered that SZRD1 may link UPF1 to STRAP (Fig. 2G, upper panel). To test this, we knocked down SZRD1 in cell lines expressing SFB-tagged UPF1 or STRAP. When we pulled down proteins with streptavidin beads, we co-purified SFB-tagged STRAP with endogenous UPF1, and SFB-tagged UPF1 with endogenous STRAP (Fig. 2H & I). However, this interaction was not affected by a reduction of SZRD1 levels in cell lines expressing SZRD1-specific shRNA.

We conclude that STRAP and UPF1 interact with SZRD1 via distinct domains in cells. The interaction of these proteins with SZRD1 is not responsible for their interaction with each other (Fig. 2G, lower panel).

SZRD1 overexpression inhibits nonsense-mediated decay

Next, we wanted to understand whether overexpression of SZRD1 leads to changes in any particular group of RNAs. To this end, we assessed transcriptional changes upon overexpression of SZRD1 in HeLa cells by RNAseq (Fig. S2A). To glean insights into functional groups that may be affected by SZRD1 overexpression, we performed a gene set enrichment analysis (98) with the Molecular Signature Database (98) and with sets of genes that had been found to be upregulated when UPF1 or other NMD components were knocked down (6). This analysis revealed that NMD target gene sets were by far the most enriched gene sets among the transcripts upregulated upon SZRD1 overexpression ($p < 10^{-139}$) (Fig. 3A&B, and table S5). Using RT-qPCR we confirmed that several *bona fide* NMD targets were upregulated upon overexpression of the short or long isoform of SZRD1 (Fig. 3C-G), consistent with a model where SZRD1 overexpression inhibits UPF1's function in NMD.

Besides its roles in NMD, UPF1 is also involved in Staufen-mediated decay (108) and histone mRNA degradation (109). When interrogating the abundance of histone mRNAs (Fig. S2B) and transcripts that are known to be regulated by Staufen-mediated decay (i.e. FBLIM1, OASL in Fig. S2C) (108) in our RNAseq dataset, we did not observe any systematic changes upon SZRD1 overexpression except for an increase in some histone mRNAs.

Of note, the list of NMD targets comprises a group of long non-coding RNA that serve as transcriptional precursors for the production of small nucleolar RNAs (i.e. snoRNAs produced from the transcripts GAS5, ZFAS1 and SNHG15 in Fig. S2D) (110,111). Previously, the SZRD1 interactant STRAP has been reported to play a role in the assembly of snoRNA-containing ribonucleoprotein complexes (112). To test whether SZRD1 might somehow affect snoRNA production we performed northern blots to quantify the abundance of snord12, snora9 and snord79 that are derived from the transcripts ZFAS1, SNHG15 and GAS5, respectively. Overexpression of SZRD1 did not affect the abundance of these snoRNAs (Fig. S2E) suggesting that the corresponding precursor transcripts, ZFAS1, SNHG15 and GAS5, were not increased due to increased transcription, but rather due to reduced degradation.

Recruitment of SZRD1 to transcripts facilitates its effect on NMD

To further investigate the action of SZRD1, we used a reporter gene construct that expresses *Renilla* luciferase in the same open reading frame as the human genomic β-globin gene sequence (Fig. 3H, 'WT') (78). To have a quantifiable readout of NMD in cells, we used a similar reporter gene construct that carries a premature stop codon at position 39 of the β-globin sequence (78), which triggers degradation by NMD (Fig. 3H, 'NMD') (113). *Renilla* luciferase activity was normalized to the activity of a firefly luciferase construct driven by the same promoter. When we transfected these constructs together with expression constructs for long and short SZRD1 isoforms in HEK293T (Fig. 3I) and HeLa cells (Fig. 3J), we observed a more than 2-fold increase in Renilla luciferase activity produced from the NMD-sensitive reporter construct, whereas the wild type construct was unaffected. Taken together, RT-qPCR (Fig. 3C-F) and reporter gene assay results indicated that SZRD1 overexpression impairs NMD.

Given the high levels of SZRD1 overexpression, the observed inhibition of NMD could be caused by a general disturbance of RNA degradation pathways or by recruitment of SZRD1 to specific transcripts. To assess effects of SZRD1, we performed reporter gene assays where SZRD1 was tethered to the 3' UTR of transcripts. To do this, six BoxB sites were inserted in the 3'UTR of transcripts of reporter gene constructs similar to what has been described by Gehring and colleagues (79) (Fig. 4A). The presence of these sites allows the specific recruitment of co-expressed proteins carrying a lambdaN-tag to BoxB-containing transcripts. In a first step, we wanted to test whether the recruitment of SZRD1 inhibits NMD or whether SZRD1 blocks NMD in a manner that does not require its recruitment to the transcript. We overexpressed the short SZRD1 isoform with a N-terminal lambdaN-tag at different concentrations together with reporter gene constructs (Fig. 4B) that contain ('WT-BoxB' and 'NMD-BoxB') or do not contain ('WT-noBoxB' and 'NMD-noBoxB') BoxB sites. We observed a dose-dependent increase in Renilla luciferase activity from the construct that contains BoxB sites and is subject to NMD ('NMD-BoxB', white bars Fig. 4B). This effect was only observed at higher concentrations of SZRD1 when in presence of the NMD-noBoxB reporter (light grey bars, Fig. 4B). Western blot analysis of samples obtained from these reporter gene assays demonstrated that these effects are observed when SZRD1 levels are increased approximately 3-fold above baseline levels (Fig. S3A).

Overall, we conclude that recruitment to transcripts facilitates the effect of SZRD1 on NMD.

The SUZ domain is required for the inhibition of NMD by overexpressed SZRD1

To elucidate the mechanism of action of SZRD1 on NMD activity in cells, we decided to identify the domains of SZRD1 that are required for this inhibition. To achieve this, we generated constructs driving expression of lambdaN-tagged versions of the SZRD1 mutants described in figure 2. As expected, lambdaN-tagged long and short SZRD1 isoforms increased Renilla activity from the NMD reporter gene construct with a premature stop codon ('NMD-BoxB') (Fig. 4C & D) in HEK293T cells. In these experiments, the tethered short form was 3 to 4-fold less efficient than the tethered long form, although the untagged versions had a similar effect (Fig. 3I). None of the SZRD1 isoforms affected the reporter gene construct with a stop codon in its physiological position ('WT-BoxB', Fig. S3B & C). All mutants were expressed at comparable levels (Fig. S3D & E), but mutations in the SUZ domain almost completely abolished the inhibition of NMD, except for the mutation 'R→Q', which still showed a residual

effect (Fig. 4C & D, bars 3 to 6). In contrast, the inhibitory effect of SZRD1 was maintained or even stronger when we introduced a mutation in the SUZ-C domain that abolishes the interaction with STRAP ('AAA') (Fig. 4C & D). Thus, our data indicate that the SUZ domain is required for the increase of *Renilla* luciferase levels generated by NMD-sensitive reporter gene constructs, whereas the SUZ-C domain is not.

Interestingly, the capacity to inhibit NMD did not directly correlate with the ability of SZRD1 to interact with UPF1 in cells. Mutations in the C-terminal part of the SUZ domain of the short isoform of SZRD1 (namely Δ2, YF and RQ) barely affected the interaction with UPF1 (Fig. 2E), but abolished its capacity to increase *Renilla* luciferase levels produced from NMD-sensitive reporter gene constructs (Fig. 4D).

Taken together, we observe that the effect of SZRD1 on NMD targets depends on the SUZ domain. Interestingly, the interaction of SZRD1 with UPF1 does not suffice to inhibit NMD. This indicates that SZRD1 does not simply take the place of an essential interaction partner of UPF1.

SZRD1 inhibits NMD predominantly by relieving translational inhibition

Transcript degradation is the major outcome of NMD. To determine whether SZRD1 inhibits RNA degradation in other contexts (i.e. outside of NMD), we generated reporter constructs carrying AU-rich elements (ARE) in their 3'UTR, which are known to induce rapid degradation of transcripts independent of NMD (114). These constructs carry not only BoxB sites but also an AU-rich element in the 3'UTR ('WT-BoxB-ARE', Fig. 4E) similar to what has been done by Gehring and colleagues (79). At baseline, transfection of 'WT-BoxB-ARE' and 'NMD-BoxB-noARE' reporter gene constructs showed 5- to 10-fold lower Renilla luciferase activities than the 'WT-BoxB-noARE' construct, which lacks premature stop codon and ARE element (Fig. 4F & G) (79). Cotransfection with constructs driving expression of the long or the short isoforms of SZRD1 significantly increased the activity of the 'NMD-BoxB-noARE' reporter gene construct compared to the empty vector control. In contrast, the activity of the 'WT-BoxB-ARE' reporter gene was completely unaffected indicating that SZRD1 does not affect ARE-dependent mRNA degradation.

To discriminate between an effect of SZRD1 on RNA degradation and RNA translation, we quantified mRNA levels of the *Renilla* and firefly luciferase transcripts by RT-qPCR in parallel to the luciferase activity measurements at steady state (Fig. 4H & I; primers highlighted in cyan on Fig. 4E). Interestingly, the increase in relative *Renilla* luciferase mRNA levels due to SZRD1 was smaller than the increase of relative *Renilla* luciferase activity (Fig. 4H & I vs 4F & G). This can be visualized more easily, when we present the ratios of the normalized luciferase activities and the normalized mRNA levels (Fig. 4J & K). While these data are more clear-cut for the long SZRD1 isoform (Fig. 4J) than in the case of the short SZRD1 isoform (Fig. 4K), these results are consistent with a model where SZRD1 overexpression increases Renilla luciferase levels predominantly via an increase in reporter gene translation.

The NMD machinery reduces gene expression by causing the degradation of transcripts and by inhibiting translation (26,115,116). A combination of transcript degradation and inhibition of translation has also been observed when miRNAs bind to mRNA transcripts (117). Thus, we wondered

whether SZRD1 might affect protein production from transcripts targeted by miRNAs. To test this, we inserted four let7a binding sites ('let7a_bs') or binding sites where key nucleotides had been mutated ('let7a_mu') in the 3'UTR of reporter constructs containing six BoxB sites (Fig. S3F). When we transfected these constructs into HEK293T cells, we observed approximately 50% less *Renilla* luciferase activity from the let7a reporter construct in comparison to the construct where the binding sites were mutated (Fig. S3G & H). Coexpression of short and long lambda-N-tagged SZRD1 isoforms did not change the activity of the let-7a reporter construct (Fig. S3G & H). This indicates that recruitment of SZRD1 does not block posttranscriptional regulation by let-7 in this reporter gene construct.

Taken together, we observe that recruitment of SZRD1 to NMD target transcripts increases reporter gene activity predominantly by relieving translational repression. In contrast, not effect was observed on reporter gene constructs with an ARE or let-7 binding sites in the 3'UTR.

SZRD1 does not form a stable complex with the exon junction complex (EJC)

NMD is a cytoplasmic process that is most efficient when translation terminates on transcripts that retain EJCs. While core elements of these complexes are deposited during splicing in the nucleus, the identity and abundance of ancillary factors varies strongly in the course of mRNA maturation, nuclear export and the first round of translation (reviewed in (10)). Given that nuclear assembly and maturation of the EJC plays an important role in NMD, we wondered whether SZRD1 might interact with components of the EJC and might thereby affect NMD. In preliminary experiments, we had noted that overexpressed SZRD1 did show some interaction with the exon junction complex components eIF4A3 and MLN51, but not with RNPS1 or Y14 (Fig. S4A). Yet, these interactions were exclusively observed with the short SZRD1 isoform. We decided to investigate whether these proteins interacted at physiological concentrations. To do this, we first knocked-in a C-terminal SFB tag into the genetic loci of the EJC components eIF4A3 and RNPS1, as well as UPF1 (Fig. 5A). This allowed us to avoid non-physiological interactions due to overexpression of proteins and ensured that all proteins were pulled down with comparable efficiency. When we pulled down SFB-tagged endogenous proteins with streptavidin-beads, we saw that endogenous UPF1 co-purified with endogenous SZRD1 (Fig. 5B). In contrast, we did not observe any co-purification of SZRD1 with the exon junction complex components RNPS1 or eIF4A3. At the same time, the EJC components MLN51 and Y14 did interact with eIF4A3 and RNPS1 suggesting that EJC components could be identified using our approach (Fig. 5B & S4B). Likewise, UPF1 did co-purify with MLN51, UPF2 and UPF3B suggesting that the interactions between NMD and EJC components were also maintained. Thus, we conclude that at physiological levels, SZRD1 is not part of a stable complex with components of the EJC.

SZRD1 increases protein levels of a subset of NMD targets including a key NMD factor

RNAseq analysis of cells overexpressing high levels of SZRD1 clearly showed an increase of most NMD target transcripts (Fig. 3B). To investigate to what extent endogenous NMD targets are affected by physiological amounts of SZRD1, we used RT-qPCR to assess mRNA levels of transcripts that are known to be degraded by NMD. Knockdown of SZRD1 by two different shRNAs did not consistently affect transcripts levels of *bona fide* NMD targets hnRNPA2B1, GAS5, ZFAS1 and SC-35 (Fig. 6A - D)

(33,118-120). One reason for this observation may be that SZRD1 is not recruited to these transcripts. In line with this reasoning, we did not observe any change of the activity of the NMD-sensitive reporter relative to the wild type reporter gene constructs in cell lines where we knocked down SZRD1 (Fig. 6E). On the other hand, the lack of an effect of SZRD1 knockdown on mRNA levels of NMD targets may also be explained if endogenous SZRD1 acted primarily to relieve translational repression rather than preventing transcript degradation.

To assess whether endogenous SZRD1 might affect translational repression, we needed to analyze the abundance of proteins produced from NMD-sensitive transcripts. While many NMD target transcripts are known, assessing the resulting protein production is not trivial, since these transcripts in many cases do not produce proteins (e.g. ZFAS1 and GAS5) or represent splice variants, whose protein products are difficult to quantify by western blot (e.g. SC-35 or hnRNPA2B1). We therefore focused on a recent quantitative proteomics study that identified proteins that are translated from transcripts targeted by NMD and that were upregulated upon UPF1 knockdown (121). Among the top four candidates, antibodies recognizing GABARAPL1 and PEA15 were commercially available and able to detect endogenous protein levels.

We then analyzed protein levels by western blot and mRNA levels by RT-qPCR in cell lines expressing different SZRD1 shRNAs, one UPF1 shRNA or a non-silencing control shRNA (Fig. 6F-Q & S5). When SZRD1 was knocked down in HEK293 cells, we observed a 50 % reduction in GABARAPL1 protein levels (while PEA15 was undetectable), whereas GABARAPL1 mRNA levels were slighly increased (Fig. 6I & J). When we performed the same experiment in U2OS cells, we observed a 50% reduction of PEA15 protein levels whereas GABARAPL1 only showed a 15% non-significant reduction (Fig. 6M & N). In this case, the reduction of PEA15 protein levels was associated with a 25% reduction of PEA15 mRNA levels in in U2OS cells (Fig. 6N & Q). While these observations are consistent with a model where SZRD1 inhibits protein production from NMD target transcripts, the observed effects are not very strong and vary between cell lines. This indicated that SZRD1 might not simply be a general inhibitor of NMD.

Interestingly, when we analysed the abundance of NMD machinery components, we observed that SZRD1 knockdown systematically reduced the abundance of UPF3B. UPF3B protein and RNA levels were reduced by 70 % and 50 % respectively (Fig. 6G & L), whereas other NMD components were less affected. When we specifically knocked down the long isoform of SZRD1, we observed an approximately 50 % reduction in UPF3B levels, indicating that the long isoform of SZRD1 has a major contribution to SZRD1 function in HEK293 cells (Fig. S5K).

At first sight, the parallel regulation of NMD components and NMD targets might be counterintuitive, since a reduction of NMD components would be expected to lead to an increase in the abundance of NMD targets, and not a decrease. However, it has previously been shown that transcripts coding for several NMD machinery components are themselves targeted by NMD. For example, transcripts coding for UPF1, UPF3B and SMG1 have all been described as NMD targets (38,39). Knockdown of UPF1 in U2OS and HEK293, to a varying degree, increased the abundance of mRNA levels of UPF3B and other NMD components (39) (Fig. 6H & O, S5D & F). This suggests that the transcripts coding for these NMD components are indeed subject to NMD in our experimental system.

Therefore, the abundance of these proteins may be reduced upon SZRD1 knockdown because their transcripts are subjected to more efficient NMD.

Next, we wondered whether overexpression of SZRD1 would lead to reciprocal effects in comparison to knockdown of SZRD1. When we analyzed cell lines overexpressing short or long isoforms of SZRD1 (Fig. 6R & S), we observed a clear increase in UPF1, SMG1, UPF2, UPF3B, and GABARAPL1 protein levels. In contrast, GAPDH was largely unaffected and PEA15 was undetectable in the HEK293 cell line. The increase of NMD components and GABARAPL1 was almost completely abolished by mutations in the SUZ domain ('Δ1' and 'Δ2'), which we had previously shown to abolish the effect on NMD (Fig. 4C & D). Thus the increase of UPF1 observed in earlier experiments (Fig. 1D & E and 2C - F), the increase of UPF3B , and the increase of some NMD targets upon overexpression of SZRD1 might be caused by a SUZ-domain-dependent inhibition of NMD.

Our data indicate that SZRD1 can exert two partially antagonistic effects on NMD. On the one side, it can inhibit NMD when recruited to transcripts, and on the other side, it may contribute to NMD efficiency by maintaining UPF3B levels. To evaluate the overall relevance of these competing effects, we quantified global transcript abundances in cell lines where we acutely overexpressed either SZRD1 shRNA #1 or a nonsilencing control. We then used GSEA to analyze behaviour of NMD targets upon SZRD1 knockdown. This revealed a strong enrichment of an NMD target gene set among the genes that were upregulated upon SZRD1 knockdown (Figure 6T, FDR 9.7×10^{-12}). While this result indicates that the abundance of a sizeable number of NMD targets tends to be increased upon SZRD1 knockdown, we note that many of them remain unchanged and some are among the transcripts that are decreased upon SZRD1 (albeit not significantly enriched among the latter group). We also noted that the abundance of a series of other gene sets were significantly changed upon SZRD1 knockdown, indicating that SZRD1 plays functions that extend beyond the regulation of NMD (Tables S5&S6) possibly via its interaction with STRAP.

Taken together, while recruitment of SZRD1 to transcripts can increase protein production from some NMD targets, these local might be offset by changes of UPF3B levels leading to transcriptome-wide changes in the abundance of NMD-targets (Fig. 6U).

The protein ARPP21 combines translational activation of NMD targets via the SUZ domain with more general stabilization of mRNAs

The capacity of SZRD1 to increase protein levels from NMD targets requires the SUZ domain. Besides in SZRD1, SUZ domains have only been found in a family of related proteins that consists of R3HDM1, R3HDM2 and ARPP21. Recently, Rehfeld and colleagues demonstrated that ARPP21 can increase the translation of some transcripts when bound to uridine-rich sequences in the 3'UTR in part via an interaction with eIF4A1 and eIF4G (122) (Fig. 7A). This was reported to antagonize the effect of miR-128 on some of its targets. In the study, an effect on NMD targets was not specifically assessed. However, in an affinity purification of R3HDM1 and ARPP21 interaction partners reported in the supplementary data of the study, UPF1 was listed as potential interaction partner (122). Likewise, in a SILAC-based study to identify UPF1 interaction partners, R3HDM1 has been identified as interaction partner of UPF1 (107).

In analogy to our observations concerning SZRD1, we hypothesized that other SUZ-domain-containing proteins might also affect the protein levels from a subset of NMD (Fig. 7B). To test this, we first wanted to confirm that some of these proteins indeed interacted with UPF1. We pulled down SFB-tagged UPF1 and analyzed interacting proteins by western blot. Consistent with the data presented by Flury and colleagues and in the supplementary dataset of Rehfeld and colleagues (107,122), we observed that SFB-tagged UPF1 pulls down R3HDM1, and that this interaction does not require the presence of intact RNA (Fig. 7C). This confirmed that R3HDM1, another SUZ domain-containing proteins, also can interact in cells with UPF1.

In a next step, we wanted to know whether ARPP21 might modulate the abundance NMD targets. To do this, we overexpressed ARPP21 in HeLa cells, which normally express very low levels of this protein. We then performed RNAseq to quantify global transcriptional changes. Using GSEA, we evaluated the behavior of NMD targets and observed that this group of transcripts was significantly enriched among the transcripts that were upregulated upon ARPP21 overexpression (Fig. 7D). This indicated that ARPP21 can indeed modulate the abundance of NMD targets.

Last but not least, we used reporter gene assays to test whether these SUZ-domain containing proteins might affect NMD in a similar way as SZRD1. In a first set of experiments, we wanted to know whether ARPP21 might elicit different effects when recruited to a NMD-sensitive transcript in comparison to a transcript that is not subject to NMD. To test this, we used *Renilla* luciferase constructs based on the wild-type β-globin gene ('WT') or a β-globin gene with a premature stop codon ('NMD'), with or without six BoxB sites in the 3'UTR allowing the recruitment of lambdaN-tagged ARPP21 (described in Fig. 4A). Overexpression of λ-ARPP21 led to a 1.6-fold increase of relative reporter activity from the wild type reporter gene construct, similar to what has been observed by Rehfeld and colleagues (122). In contrast, we observed an 8.3-fold increase of the NMD reporter construct (Fig. 7E). No change was observed when λ-ARPP21 was co-transfected with reporter constructs that lacked BoxB sites. Thus, our data indicate that recruitment of λ-ARPP21 to reporter transcripts can increase protein production, but that this increase is much more profound for NMD-sensitive reporter gene transcripts.

To further elucidate the mechanism, we investigated the effect of λ-ARPP21 recruitment to reporter gene transcripts that were subject to NMD (NMD-BoxB-noARE), that are subject to ARE-dependent degradation ('WT-BoxB-ARE') or that aren't suject to either of these processes ('WT-BoxB-noARE') (see Fig. 4E). Similar to the experiments shown in Fig. 4F-K, we not only assessed luciferase reporter activity but also reporter gene mRNA levels. In addition to wild type ARPP21 (λ-ARPP21), we used the protein product of a rare splice variant of ARPP21 that lacks the second half of the SUZ domain (λ-ARPP21ΔSUZ; see Fig. 2B, underlined in black). Recruitment of wild type ARPP21 to the 3'UTR led to an increase of *Renilla* luciferase activity from the wild type transcripts (Fig. 7F, bars for WT-BoxB-noARE) and from the ARE-containing transcripts (Fig. 7F, bars for WT-BoxB-ARE). In both cases, the increase in *Renilla* luciferase activity was comparable to the increase in mRNA levels (Fig. 7G & H). Of note, the observed mRNA increase was independent of the presence of the SUZ-domain (Fig. 7G) and also observed when λN-tagged R3HDM1 and R3HDM2 were overexpressed (Fig. S6).

Strikingly, the situation was very different when ARPP21 was recruited to a NMD-sensitive construct. Both wild-type and SUZ-domain lacking ARPP21 led to comparable increases in transcript levels (Fig. 7G, bars for NMD-BoxB-noARE). In contrast, reporter gene activity was strongly increased by the wild type protein than by the deletion mutant (7.6-fold versus 2.6-fold) (Fig. 7F & H, bars for NMD-BoxB-noARE). The significantly higher protein levels from comparable amounts of transcripts indicate that recruitment of ARPP21 can increase translation of this NMD-target in a SUZ-domain-dependent manner. This effect seems to occur in addition to an increase in reporter mRNA levels, which was independent of the SUZ domain.

Reciprocal regulation of SZRD1 and ARPP21 via miR-128 may alter levels of NMD targets in a tissue specific manner.

miR-128 is produced from the introns of spliced transcripts that code for the proteins R3HDM1 and ARPP21 (Fig. 8A). Several miRNA binding site prediction algorithms (123,124) identified four putative binding sites for miR-128 in the 3'UTR of SZRD1, which are well conserved during evolution (Fig. S7A). Starting 120 nucleotides after the stop codon they are spaced at a distance of approximately 25 nucleotides between each other. We were intrigued by this potential regulation, since SZRD1, R3HDM1, ARPP21 and R3HDM2 are the only proteins in the human genome that contain the conserved SUZ domain (68). A regulation of SZRD1 by miR-128 could therefore indicate that the transcriptional activation of the R3HDM1/ARPP21-miR-128 loci would replace one SUZ-domain protein (i.e. SZRD1) with other ones (i.e. R3HDM1/ARPP21). In fact, miR-128 may help maintain low SZRD1 expression levels in the central nervous system, where ARPP21 and R3HDM1 levels are highest (Fig. S7 B-E) (75). Consistent with this, western blot analysis for SZRD1 in neurons from miR-128-2 knockout mice indicated that SZRD1 might be a miR-128 target (60). Yet, the specific requirement of the binding sites in the 3'UTR has not been assessed.

We therefore generated a constitutively active firefly luciferase reporter gene construct carrying the region of the 3'UTR of the SZRD1 3'UTR comprising the four predicted miR-128 binding sites (Fig. 8B, 'wt site'). As a control, we either used a construct without the SZRD1 3'UTR (Fig. 8B, 'no 3'UTR) or constructs where three nucleotides in all miR-128 binding sites were mutated (Fig. 8B, 'mut site'). We then transfected these constructs in the presence of a miR-128 mimic or a control mimic into HeLa cells. Firefly luciferase reporter gene activity was measured and normalized to the activity of a co-transfected *Renilla* luciferase reporter gene driven by the same promoter (and normalized to the value of the control miR-ctrl). In the presence of the miR-128 mimic, we observed a threefold decrease in the activity of the luciferase construct containing all four miR-128 binding sites, whereas we did not see any difference when all four sites were mutated or when the SZRD1 3'UTR was absent (Fig. 8B). This strongly indicated that miR-128 can exert a function by binding to some of the predicted binding sites in 3'UTR of SZRD1. In a second step, we tested whether miR-128 affects endogenous SZRD1 protein levels. To this end, we transfected miR-128 mimics or negative control mimics in HEK293 cells and quantified endogenous SZRD1 via western blot. We observed a three-fold decrease in SZRD1 protein levels when miR-128 mimic was transfected in comparison to two miR-mimic controls (Fig. 8C). Lastly, we wanted to investigate whether endogenous levels of miR-128 suffice to regulate SZRD1 protein levels. To this

end, we turned to a cell line that expresses elevated levels of miR-128 at baseline. Consistent with elevated expression of miR-128 in the thymus, we observed high levels of miR-128 in the acute T-cell lymphoma cell line Jurkat (Fig. S7F). Electroporation of a locked nucleic acid (LNA) inhibitor of miR-128 in Jurkat cells led to a 60% increase in SZRD1 levels in comparison to cells transfected with a negative-control LNA inhibitor (Fig. 8D). Overall, our experiments clearly demonstrate that miR-128 acts on the 3'UTR of SZRD1 and that acute modulation of miR-128 leads to reciprocal changes of SZRD1 protein levels in cells.

Both ARPP21 and SZRD1 can increase protein production when recruited to NMD reporter gene transcripts. It is therefore tempting to speculate that the exchange of one protein by the other one may serve to modulate protein levels of distinct subsets of NMD target transcripts. Indeed, SZRD1 knockdown and ARPP21 overexpression lead to only partially overlapping increases of NMD targets (Fig. 6T & 7D). Of note, the GSEAs also revealed that overexpression of ARPP21 and knockdown of SZRD1 lead to distinct gene expression changes that are likely to be a consequence of mechanisms outside of NMD (Table S6&7). This indicates that the switch from ARPP21 to SZRD1 expression might also serve to exchange these functions.

DISCUSSION

Recruitment of SUZ-domain containing proteins can increase protein production from NMD targets

When we overexpressed SZRD1 to supraphysiological levels (Fig. S2A), transcript levels of NMD targets were universally increased in cells (Fig. 3A&B) consistent with a general inhibition of NMD. However, when we expressed lambdaN-tagged SZRD1 at close to physiological levels, a more differentiated picture emerged. First, we observed that much lower concentrations were required to increase reporter production when SZRD1 was recruited to transripts, which indicated that SZRD1 needs to be in proximity to the transcript, and does not simply bind away some NMD factor.

Secondly, we noted that SZRD1 recruitment had a much stronger effect on reporter activity (6.72-fold) than on mRNA levels (1.46-fold) (Fig. 4). This observation was consistent with a model, where recruitment of SZRD1 in a SUZ-domain dependent manner relieves the translational repression occurring during NMD. While other models might explain the observed discrepancy between changes in protein and mRNA levels, this model is supported by our investigations of ARPP21 function. Recruitment of lambdaN-tagged ARPP21 to NMD reporter gene constructs led to an increase in protein and mRNA levels. Interestingly, deletion of the SUZ domain reduced the increase in protein levels by more than 2-fold, whereas mRNA levels were completely unaffected (Fig. 7). This demonstrates that ARPP21 can increase protein translation from these transcripts in a SUZ domain manner. Of note this SUZ-domain-dependent effect on protein levels was not observed in ARE-containing transcripts and in control cell lines. Taken together, our data support the notion that SUZ domain-containing proteins ARPP21 and SZRD1 can relieve translational inhibition during NMD.

Previous studies have reported that the SUZ domain in the *C. elegans* homolog of SZRD1 can bind to single stranded RNA, but it is unclear whether this would occur in a sequence specific manner (68). In some of our experiments, we used lambdaN-tagged SZRD1 or ARPP21 which are recruited to

BoxB sites in the 3'UTR of NMD targets thereby obviating a potential requirement of the SUZ-domain to bind these proteins to mRNA (Fig. 4 & 7). In this situation key residues in the SUZ domain were still required to overcome translational inhibition of these transcripts (Fig. 4C & D and Fig. 7). This demonstrates that the function of the SUZ domain is not limited to the recruitment of SZRD1 to mRNA, but rather that the SUZ domain likely is implicated in the modulation of translation of NMD targets.

SZRD1 interacts with UPF1 and prevents translational inhibition of NMD targets. The easiest model to explain this effect would be that SZRD1 simply displaces a component of the NMD machinery and thereby prevents NMD. This model would predict that binding of SZRD1 to UPF1 should be sufficient to overcome translational inhibition. When we tested expression constructs for the short form of SZRD1 carrying different mutations in the SUZ domain, several of these mutants still retained interaction with UPF1, but completely lost their ability to alleviate translational repression (Fig. 2E & 4D). Since binding of SZRD1 to UPF1 is not sufficient to alleviate translational represstion, SZRD1 does not simply disrupt interaction of UPF1 with other proteins. Consistent with this, we did not observe any reduced interaction of UPF2, UPF3B and SMG1 with SFB-tagged UPF1, when we overexpressed SZRD1 in our UPF1 knock-in cells (data not shown).

An alternative possibility would be that SZRD1 needs to interact with a protein other than UPF1 to exert its function. Based on the activity of SZRD1 mutants (Fig. 2E & 4D), it is tempting to speculate that this unknown interaction partner would bind to the C-terminal part of the SUZ domain. Consistent with this hypothesis, deletion of the C-terminal part of the SUZ domain in ARPP21 completely abolished the effect of ARPP21 on protein translation of reporter gene constructs (Fig. 7F & G).

Taken together, we conclude that recruitment of SUZ-domain containing proteins can increase protein production from NMD targets at least in part by counteracting the translational repression occurring during NMD. Future studies will be required to dissect the molecular mechanisms underlying this effect.

Regulation of protein production of a subset of NMD-targets by SZRD1.

NMD plays a key role in the elimination of potentially dangerous proteins that could be formed from transcripts resulting from aberrant splicing or from the transcription of mutated genes. However, many physiological transcripts are also subject to NMD (8) due to the presence of long 3'UTRs or upstream open-reading frames. In some of these cases, degradation by NMD may play a role in limiting the half-life of transcripts coding for physiological relevant proteins. Regulation of protein production from these transcripts might therefore be physiologically relevant.

So far, most of what we know about the regulation of NMD concerns the regulation of transcript degradation. Binding of several proteins has been shown to prevent NMD. For example, binding of PTBP1 (36) or binding of the protein hnRNPL (37) to the 3' UTR can protect transcripts from degradation via NMD. Curiously, there is little overlap between transcripts protected from NMD by hnRNPL and PTBP1 suggesting that these proteins protect a distinct subset of transcripts from degradation by NMD. Many regulators likely remain to be discovered since *cis* elements in the first 100 base pairs of the 3'UTRs have been shown to confer resistance to degradation by NMD to transcripts with long 3'UTRs (35), but the molecular mechanism underlying these observations remains unclear.

NMD not only leads to transcript degradation, but also to inhibition of translation. Some mechanisms for this translational repression have been characterized (26,107,125) and others can be inferred by the recruitment of the CCR4/Not1 complex during NMD (42,126). Yet, in depth studies are complicated by the fact that translation of transcripts is required to trigger NMD. We found that recruitment of SZRD1 and ARPP21 increases protein production from reporter gene transcripts that are subject to NMD (Fig. 4F-K, 7F-H) in a SUZ-domain dependent manner. This indicated that translational inhibition during NMD can be regulated with a certain specificity.

Our SZRD1 knockdown experiments indicate that SZRD1 might effect protein production from NMD targets via two competing effects. On the one hand, recruitment to the transcript will increase protein production by inhibiting NMD, at least in part via a relief from translational repression. On the other hand SZRD1 is required to maintain protein levels of the NMD target UPF3B (Fig. 6F-Q) (39). Thus an increase of UPF3B and other NMD machinery components might improve NMD efficiency (Fig.6U) and thereby competing with the local effect of SZRD1. Consistent with this model, NMD target genes were actually enriched among the transcripts that were upregulated upon SZRD1 knockdown (Fig. 6T).

Overall, our data suggest that SZRD1 can increase protein production when it is recruited to NMD transcripts. On a whole transcriptome scale, this effect seems to be outweighed by its capacity to maintain UPF3B levels. Together these two competing effects may allow to regulate protein production from a specific subset of NMD targets.

Does the downregulation of SZRD1 reinforce the effect of miR-128 on other targets?

A tissue-specific miR-128 knockout in mouse brain causes lethal seizures at a few months of age (60). In this study, increased activity of the ERK2 branch of the MAP kinase pathway was observed, which is known to regulate neuronal excitability (60). This change was attributed to an increase in several proteins that carry miR-128 binding sites in the 3'UTR. One of these proteins is a *bona fide* regulator of MAP kinase signaling, PEA15, and the other was SZRD1. This observation is remarkable in light of our finding that SZRD1 can modulate the abundance of PEA15 protein levels (Fig. 6N & Q). In fact, this opens the possibility that the increase of PEA15 levels in miR-128 knockout cells is not only due to the loss of a directed effect of miR-128, but also due to an increase of SZRD1 levels (Fig. S8). Interestingly, this might not be the only situation where the loss of SZRD1 may reinforce the effect of miR-128. Both UPF1 and UPF3B have been shown to be miR-128 targets (44,45). In the present study, we observed that UPF3B (and to a lesser extent UPF1) protein levels are increased by SZRD1 (Fig. 6 G & L and Fig. S5 A & G). Again, this opens the possibility that some of the modulation UPF3B (and UPF1) levels by miR-128 might be caused by a downregulation of SZRD1 (Fig. S8).

Reciprocal regulation of ARPP21 and SZRD1 might allow context-dependent regulation of protein production from NMD targets.

When recruited to reporter gene transcripts, SZRD1 mainly increases translation of NMD-sensitive transcripts (Fig. 4J & K). In contrast, recruitment of ARPP21, R3HDM2 and R3HDM1 leads to a general increase in reporter gene transcript abundance (Fig. S6), which is combined with increased protein

translation of NMD-sensitive transcripts mainly in the case of ARPP21 (Fig. 7 F-H and Fig. S6). Several arguments suggest that the SUZ domain containing proteins modulate translation of NMD targets by a similar mechanism. First, SZRD1 and ARPP21 both require a functional SUZ domain for this effect. Second, we have observed that both SZRD1 and R3HDM1 interact with UPF1. While we have not been able to test the interaction between ARPP21 and UPF1, we note that UPF1 was among the proteins that co-purified with ARPP21 in the report of Rehfeld and colleagues (supplementary data of (122)).

Gene expression data from human tissues highlights a low expression of SZRD1 in the brain, where expression of other SUZ domain proteins is highest (Fig. S7 B-E). We have shown that miR-128, produced from introns of R3HDM1 or ARPP21, reduces SZRD1 protein levels (Fig. 8). While the strong differences in transcript levels between organs are likely to be caused by differences in gene transcription, our and published data demonstrate that miR-128 can reduce SZRD1 levels in cells that express high levels of ARPP21 and miR-128 (60). Thus, miR-128 might help reinforce a reciprocal regulation of miR-128 and SZRD1. Both ARPP21 and SZRD1 can modulate protein production from NMD targets. It is therefore tempting to speculate that a shift from SZRD1 to ARPP21 might lead to the regulation of protein production from distinct subsets of NMD targets.

Recruitment of SZRD1 increases protein production from NMD-sensitive reporter transcripts. In contrast, we observed no effect on ARE-containing reporters or reporter containing a let-7 binding site. This demonstrates that SZRD can affect NMD targets in a way that leaves other pathways of mRNA degradation and translational repression unaffected. Previous studies have demonstrated that ARPP21 can increase protein output from transcripts that contain binding sites for miR-128 in their 3'UTR (122). At first sight our observation that SZRD1 does not change protein production from a reporter gene transcript with let-7 binding sites might seem to conflict with the idea that SZRD1 would affect miRNA-dependent silencing. However, such an effect might depend on a specific constellation of other RNA binding factors in the 3'UTR. Hence, we cannot exclude the possibility that SZRD1 may relieve translational inhibition induced by another miRNA that remains to be identified (if it exists).

Beyond an effect on NMD targets, GSEA of transcriptome changes upon SZRD1 knockdown and overexpression revealed that expression levels of several gene sets were modulated by these interventions. Given that SZRD1 interacts with STRAP independent of its effect on NMD, it is likely that the interaction with STRAP is responsible for part of these effects (Table S5-6). Likewise the overexpression of ARPP21 led to gene expression changes beyond the modulation of NMD (Table S7). These observations indicate that the shift of expression from SZRD1 to ARPP21 likely leads to fundamental changes that might be unrelated from its effects on NMD. However, we need to note that the RNAseq analyses only represent part of the picture. Our data indicate that on some transcripts SZRD1 and ARPP21 might predominantly relieve translational repression. As such, the resulting changes in protein abundance might contribute to the overall function of SUZ-domain containing proteins, but might be undetectable in a transcriptomic analysis.

To make things even more complicated, two additional layers of regulation exist. First, ARPP21 has previously been shown to increase translation of transcripts targeted by miR-128, thereby antagonizing the function of the miRNA that is produced from its gene locus (122). This complicates things since upregulation of ARPP21 might partially prevent a reduction in SZRD1 protein levels caused

by miR-128. Secondly, the transcript of ARPP21 contains miR-128 binding sites that limit protein production from this transcript (60). Both layers of regulation might prevent a complete shift from ARPP21 to SZRD1. To fully understand the function of these proteins, it will be important to understand to which transcripts ARPP21 and SZRD1 bind. Eventually, this will allow us to correlate ARPP21 and SZRD1 protein expression with changes in the abundance of the proteins produced from these transcripts.

We conclude that expression of SUZ domain-containing proteins can alter the protein levels produced from NMD targets. This reveals a novel layer of regulation within this group of transcripts that will need to be explored in future studies.

AVAILABILITY

RNAseq data are available via Arrayexpress.

ACCESSION NUMBERS

RNAseq data has been deposited at Arrayexpress under the accession number E-MTAB-9627.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

M.H. & G. T. B. conceived the study. M. H., M.B., A. H., I.G., J.G. performed molecular biology and cell biological experiments presented in the figures. D.V. identified proteins by mass spectrometry. L.G. and T.K. analyzed RNAseq data. M.H., A.H. & G.T.B. wrote the paper with input from all authors.

FUNDING

CONFLICT OF INTEREST

The authors do not have any conflict of interest to declare.

TABLE AND FIGURES LEGENDS

Figure 1. Different isoforms of SZRD1 interact with several RNA binding proteins.

(A) Schematic representation of the *SZRD1* gene locus and the resulting transcripts. Inclusion or exclusion of exon 2 (in yellow) lead to long and short isoforms, respectively. Two potential start codons (ATG) in exon 1 are highlighted in red. In addition, the inclusion or exclusion of three nucleotides (CAG) before exon 3 results in an additional amino acid at the protein sequence (see B). **(B)** Schematic representation of *SZRD1* transcripts and the resulting proteins. The long transcripts contain exon 2 and need to be translated starting from the first AUG codon, whereas the short transcripts lack exon 2 and need to be translated from the second AUG codon to allow protein production. Since the two start codons are not in the same reading frame, the N-terminal part of short and long SZRD1 isoforms are different. Inclusion of three nucleotides (+CAG) before exon 3 during splicing leads to the insertion of a serine or an arginine residue in the long and short form, respectively. 'SUZ' and 'SUZ-C' designates conserved domains. **(C)** shRNAs specifically targeting all *SZRD1* isoforms (exon 4, '#1','#2') or only the long isoform (exon 2; '#3','#4','#5','#6') were stably expressed in HEK293 cells. Western blot analysis was performed with antibodies predicted to recognize short and long *SZRD1* isoforms, as well as β-actin as loading control. Samples labeled with 'control' express a non-silencing control shRNA. **(D)** Protein lysates from HEK293 cells overexpressing the indicated SZRD1 proteins ('SZRD1') with a C-terminal SFB tag or an empty vector ('control') were subjected to affinity purification using streptavidin-Sepharose beads, and analyzed by western blot using the indicated antibodies. Input samples correspond to 10 % of the amount used in the pulldown. Exposure time was identical for input and pulldown samples except for western blots analysing STRAP, where pulldown exposure was 5x shorter (asterisk). **(E)** Protein lysates from HEK293 cells overexpressing SZRD1 proteins (containing the additional arginine or lysine residue) with a C-terminal SFB tag identical to panel (D) were subjected to affinity purification using streptavidin-Sepharose beads in the presence ('+') or absence ('-') of RNAse A/T1. Western blot analysis and presentations are identical to panel D. **(F-G)** Protein lysates from HEK293 cells overexpressing UPF1 with an N-terminal SFB tag ('UPF1') (F), STRAP with an N-terminal SFB tag ('STRAP') (G) or an empty vector ('control') were subjected to affinity purification followed by Western blotting as described in panel D.

Figure 2. **SZRD1 interacts with UPF1 via the SUZ domain and STRAP via the SUZ-C**.

(A&B) Alignment of human SUZ-C domain (A) and SUZ domain (B) sequences of the indicated proteins (top panel), and graphical representation of the conservation of the indicated proteins in four vertebrates (*Homo sapiens*, *Mus musculus*, *Gallus gallus* and *Xenopus tropicalis*) (lower panel). Mutations and deletions ('Δ') used in the other panels are indicated. **(C-D)** Protein lysates from HEK293 cells overexpressing the short SZRD1 protein containing the arginine insertion (C) or the long SZRD1 protein containing the serine insertion (D) with a C-terminal SFB tag ('WT'), or a deletion of the SUZ-C domain ('ΔC') or the corresponding proteins with the indicated mutation ('AAA') were subjected to affinity purification using magnetic streptavidin beads followed by western blot analysis. Input corresponds to 10% of the amount used in the pulldown. Cell lines labeled with 'control' were transduced with an empty control vector. **(E-F)** Pulldown of overexpressed short (E) and long (F) SZRD1 proteins containing a C-

terminal SFB tag was performed as described in panel C & D, except that mutations were introduced in the SUZ domain. Exposure time was identical for input and pulldown samples except for western blots analysing STRAP, where pulldown exposure was 5x shorter (asterisk). **(G)** Schematic representation of SZRD1 interacting with UPF1 and STRAP. The upper panel shows the possibility of SZRD1 serving as a bridge between STRAP and UPF1. The lower panel shows the possibility that these three proteins interact independent of each other. **(H-I)** HEK293 cells expressing STRAP (H) or UPF1 (I) with an N-terminal SFB tag or empty vector control cells ('ctrl') were engineered to express an shRNA targeting SZRD1 ('+') or a nonsilencing shRNA ('-'). Protein lysates were subjected to affinity purification using streptavidin-sepharose beads followed by western blot analysis. Input corresponds to 10% of the amount used in the pulldown.

Figure 3. SZRD1 overexpression inhibits Nonsense Mediated Decay.

(A) RNAseq was performed 48h after infection of HeLa with lentiviruses driving SZRD1 overexpression or an empty vector control. GSEA with a gene set of NMD targets (6) revealed strong enrichment among transcripts increased upon SZRD1 overexpression. Genes are sorted according to the signed t-statistic. Upregulated genes are on the left; FDR , false discovery rate. **(B)** Volcano plot of gene expression changes upon overexpression of SZRD1. Members of the NMD target gene set are highlighted in red. **(C-F)** mRNA levels of four transcripts known to be degraded by the NMD machinery were measured by RT-qPCR in HEK293 cells expressing the indicated SZRD1 proteins (long (+CAG) and short (+CAG)) or in empty vector control cells ('ctrl'). Expression levels were normalized to the average of *B2M* and *TBP* mRNA levels. Values are means +/-SEM of three independent experiments, where each condition was performed in triplicates. Asterisks indicate p<0.05 in post-hoc testing after one-way ANOVA. **(G)** To quantify the degree of overexpression, we mixed protein lysates from untransfected cells with 1 % of a lysate from cells overexpressing SZRD1 (used in C-F). This revealed more than 100-fold overexpression of SZRD1 isoforms. 'LC' represents a nonspecific band used as loading control. **(H)** Schematic representation of *Renilla* luciferase reporter gene constructs used in panels I and J ('WT', or 'NMD'). Each construct contains the *Renilla* ORF (in white) driven by the CMV promoter followed by a genomic β-globin sequence comprising two introns and three exons (represented with the β-globin coding sequence highlighted in brown). In the 'NMD' construct, a premature stop codon ('STOP') before the second intron triggers NMD. For all the experiments, a firefly luciferase construct driven by the same promoter was used to normalize luciferase activities. **(I-J)** Luciferase activities were measured in HEK293T (I) or HeLa (J) cells transfected with the indicated reporter gene constructs ('WT' black bars and 'NMD' grey bars) and with plasmids driving expression of the short or long forms of SZRD1 proteins (containing the arginine or serine insertions, respectively), or the corresponding empty plasmid ('ctrl'). Values represent the ratio of *Renilla* to firefly luciferase activity and are means +/-SEM of three independent experiments, where each condition was performed in triplicates. Ratios are presented relative to the ones observed with WT reporter and control plasmid. Asterisks indicate p<0.05 in post-hoc testing after two-way ANOVA.

Figure 4. Recruitment of SZRD1 increases protein production from NMD targets in a SUZ-domain dependent manner.

(A) Schematic representation of *Renilla* reporter gene constructs used in (B) to recruit SZRD1 to reporter transcripts containing BoxB sites ('WT-BoxB' and 'NMD-BoxB') or not ('WT-noBoxB' and 'NMD-noBoxB'). Each construct contains the *Renilla* ORF (in white) followed by a genomic β-globin sequence (coding sequence in brown). In the 3'UTR, six BoxB sites ('BoxB') or not ('BoxB-ctrl') were inserted. A premature stop codon ('STOP') in the NMD constructs triggers nonsense mediated decay. **(B)** Relative luciferase activities were measured in HEK293T cells transfected with the indicated reporter gene constructs containing or no BoxB sites, together with the indicated amounts of constructs driving expression of lambdaN-tagged short SZRD1 or an empty vector. **(C-D)** Relative luciferase activities were measured in HEK293T cells cotransfected with the indicated reporter gene constructs ('NMD-BoxB') and expression vectors for lambdaN-tagged long (C) or short (D) SZRD1 forms containing or not mutations in the SUZ and SUZ-C domains described in Figure 3A & B. 'empty' indicates transfection of an empty expression vector. **(E)** Schematic representation of *Renilla* reporter gene constructs used in panels F to K. Each construct contains the Renilla ORF (in white) followed by a genomic β-globin sequence (coding sequence in brown). In 3'UTR, six BoxB sites ('BoxB') and an AU-rich element ('ARE') or a control sequence ('AREctrl') were inserted. A premature stop codon ('STOP') in the NMD constructs triggers nonsense mediated decay. The cyan arrows indicate positions of primers used to quantify mRNA levels by qPCR. **(F-K)** Relative luciferase activities (F and G) and mRNA levels (H and I) were measured in HEK293T cells transfected with the indicated reporter gene constructs in presence (grey bars) or absence (black bars, 'ctrl') of plasmids driving expression of the long ('λ-SZRD1 long') (F and H) or the short form of SZRD1 (+CAG) with a lambdaN tag in the N-terminus ('λ-SZRD1 short') (G and I). Luciferase activities are presented as the ratio of *Renilla* to firefly luciferase activity. mRNA levels were measured by RT-qPCR and are presented as ratio of *Renilla* to firefly luciferase mRNA calculated by the delta-delta-Ct method. **(J-K)** Ratios of the normalized luciferase activities and normalized mRNA levels observed for the indicated reporter constructs in the presence of the lambdaN tagged long (J; obtained from panels F & H) or short (K; obtained from panels G & I) SZRD1 forms. In all panels, values are means +/-SEM of at least three independent experiments, where each condition was assessed in triplicates. Asterisks indicate p<0.05 in post-hoc testing after two-way ANOVA. Ratios are presented relative to the ones observed with WT-BoxB (B-D) or the WT-BoxB-noARE reporter (F-K) and the control plasmid.

Figure 5. SZRD1 is not part of the exon junction complex (EJC).

(A) Schematic representation of the selection cassette used to insert an SFB tag before the stop codon of the endogenous protein coding sequence. CRISPR/CAS9 cuts at the end of the ORF on the target gene and the transfection of a donor sequence with 700 to 1500 bp homology arms allows the insertion of an SFB tag followed by a P2A self-cleaving peptide and a puromycin resistance gene. **(B)** Protein lysates from HEK293 cells carrying a C-terminal SFB tag knocked into the endogenous loci of UPF1, eIF4A3 and RNPS1, or an unmodified control cell line ('ctrl') were subjected to affinity purification using magnetic streptavidin beads followed by western blot analysis with the indicated antibodies. Input corresponds to 10% of the amount used in the pulldown. Exposure time was identical for input and

pulldown samples except for western blots analysing the FLAG epitope, where pulldown exposure was 5 times shorter (*). Complementary analyses are shown in Fig. S4.

Figure 6. SZRD1 affects NMD targets via direct and indirect mechanisms.

(A-D) mRNA levels of four known NMD targets were measured by RT-qPCR in HEK293 cells expressing a non-silencing control shRNA ('ctrl'), two different shRNAs ('1' and '2') targeting *SZRD1* or one shRNA targeting *UPF1*. Expression levels were normalized to the average of *B2M* and *TBP* mRNA levels. **(E)** Relative luciferase activities were measured in HEK293T cells transfected with the indicated reporter gene constructs ('WT', black bars and 'NMD', grey bars) and expressing shRNAs against *SZRD1* or *UPF1*, or a nonsilencing control ('ctrl') (left panel). The ratios of the normalized luciferase values obtained for the 'WT' versus the 'NMD' construct for each cell line are represented in the right panel. **(F)** HEK293 cell lines expressing a non-silencing shRNA ('control'), two different shRNAs targeting SZRD1 ('1', '2') or one shRNA targeting UPF1 were analyzed by western blot using the indicated antibodies. **(G-I)** Quantification of UPF3B (G) and GABARAPL1 (I) protein levels observed shown in F, normalized to the abundance of GAPDH. **(H-J)** Quantification of *UPF3B* (H) and *GABARAPL1* (J) mRNA levels as described in (A-D). These experiments were performed in parallel to the protein quantification in F. **(K)** U2OS cell lines expressing non-silencing shRNA ('control'), two different shRNAs targeting SZRD1 ('1', '2') or one shRNA targeting UPF1 were analyzed by western blot using the indicated antibodies. **(L-N)** Quantification of UPF3B (L), GABARAPL1 (M) and PEA15 (N) protein levels measured in K were normalized to the abundance of GAPDH. **(O-Q)** Quantification of UPF3B (O), GABARAPL1 (P) and PEA15 (Q) mRNA levels as described in (A-D) but here in U2OS cell lines. These experiments were performed in parallel to the protein quantification in K. **(R-S)** Protein lysates from HEK cells overexpressing C-terminally SFB-tagged short (R) or long (S) SZRD1 protein ('wt', containing the additional nucleotides CAG), the indicated mutants of the SUZ domain ('Δ1' and 'Δ2') or an empty vector ('control') were analyzed by western blot using the indicated antibodies. **(T)** RNAseq was performed 60h after induction of an SZRD1 shRNA or a control shRNA. GSEA with a gene set of NMD targets (6) revealed strong enrichment among transcripts increased upon SZRD1 knockdown. Genes are sorted according to the signed t-statistic. Upregulated genes are on the left; FDR , false discovery rate. **(U)** Schematic representation of the competing effects of SZRD1 on the protein production of NMD targets.

In panels where mRNA levels or luciferase activities are presented, values are means +/-SEM of three independent experiments, where each condition was performed in triplicates. When protein levels were quantified, values are means +/- SEM of three independent experiments. Asterisks indicate p<0.05 in post-hoc testing after one-way ANOVA.

Figure 7. The protein ARPP21 combines translational activation of NMD targets via the SUZ domain with a general stabilization of mRNAs.

(A) Schematic representation of ARPP21 function described by Rehfeld and colleagues. (**B**) We hypothesized that recruitment of ARPP21 might increase protein production from NMD-targets in a SUZ-domain-dependent manner similar to the action of SZRD1. **(C)** HEK293 cells expressing UPF1 with an N-terminal SFB tag ('SFB-UPF1') or not ('ctrl'), were engineered to express an shRNA targeting

R3HDM1 ('+') or a non-silencing shRNA ('-'). Protein lysates were subjected to affinity purification using streptavidin-coupled Sepharose beads followed by western blot analysis using the indicated antibodies. Input samples correspond to 10 % of the amount used in the pulldown. Exposure time was identical for input and pulldown. **(D)** RNAseq was performed 48h after infection with lentiviruses driving ARPP21 expression or an empty vector control. GSEA with a gene set of NMD targets (6) revealed strong enrichment among transcripts increased upon ARPP21 overexpression. Upregulated genes are on the left; FDR , false discovery rate. **(E)** Relative luciferase activities measured in HEK293T cells transfected with the indicated reporter gene constructs ('WT' and 'NMD') with 6x BoxB sites or not in presence (blue bars) or absence (black bars, 'ctrl') of plasmid driving the expression of ARPP21 containing a lambdaN tag in the N-terminus ('λ-ARPP21 full'). Luciferase activities are presented as the ratio of *Renilla* to Firefly luciferase activities. Values are means +/- SEM of three independent experiments, where each condition was performed in triplicates. **(F-H)** HEK293T cells were transfected with the indicated reporter gene constructs in presence (colored bars) or absence (black bars) of plasmids driving the expression of the N-terminally lambdaN-tagged ARPP21 (dark blue bars), or a splice variant of ARPP21 that lacks the SUZ domain (bright blue bars). Normalized (i.e. *Renilla*/firefly) luciferase activities (F) and mRNA levels (G) were measured as described in Figure 4H-I. Panel H represent the ratios of normalized luciferase activities relative to normalized mRNA levels. Values are means +/-SEM of three independent experiments, where each condition was performed in triplicates. Asterisks indicate p<0.05 post-hoc testing after two-way ANOVA. Ratios are presented relative to the ones observed with WT-BoxB (E) or the WT-BoxB-noARE reporter (F-H) and the control plasmid.

Figure 8. Reciprocal regulation of SZRD1 and ARPP21/R3HDM1 via miR-128.

(A) Schematic representation of the genes *R3HDM1* or *ARPP21* containing the intronic miR-128 (left panel), and the gene coding for *SZRD1* (right panel). All three proteins contain a SUZ domain (indicated in blue). The *SZRD1* transcript has four predicted binding sites for miR-128 (highlighted in red). **(B)** miR-128 or control mimics were co-transfected with firefly luciferase reporter constructs containing part of 3' UTR of human SZRD1 encompassing the four predicted miR-128 sites ('wt site') or mutated versions of these binding sites ('mut site'). Firefly luciferase activity was normalized to the activity of a co-transfected *Renilla* luciferase reporter gene construct driven by the same promoter. Values are means +/-SEM of three independent experiments, where each condition was performed in triplicates. Asterisks indicate p<0.05 in post-hoc testing after two-way ANOVA. **(C)** HEK293 cells were transfected with miR-128 or two controls mimics (#1 and #2). SZRD1 and GAPDH protein levels were assessed 48 h post transfection by western blot with specific antbodies. In the lower panel, the ratio of SZRD1 and GAPDH signals is presented as means +/- SEM of three independent experiments. Asterisks indicated p<0.05 in post-hoc testing after one-way ANOVA. **(D)** Jurkat cells were transfected with a locked-nucleic acid miR-128 inhibitor ('128') or a control inhibitor ('ctrl') at 50 nM or 100n M, as indicated. SZRD1 and GAPDH levels were assessed by western blot and are presented as described in panel C.

REFERENCES

1. Chang, J.C., Temple, G.F., Trecartin, R.F. and Kan, Y.W. (1979) Suppression of the nonsense mutation in homozygous beta 0 thalassaemia. *Nature*, **281**, 602-603.

2. Maquat, L.E., Kinniburgh, A.J., Rachmilewitz, E.A. and Ross, J. (1981) Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia. *Cell*, **27**, 543-553.

3. Jack, H.M., Berg, J. and Wabl, M. (1989) Translation affects immunoglobulin mRNA stability. *Eur J Immunol*, **19**, 843-847.

4. Buhler, M., Steiner, S., Mohn, F., Paillusson, A. and Muhlemann, O. (2006) EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. *Nat Struct Mol Biol*, **13**, 462-464.

5. Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z. and Muhlemann, O. (2008) Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol*, **6**, e92.

6. Colombo, M., Karousis, E.D., Bourquin, J., Bruggmann, R. and Muhlemann, O. (2017) Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA*, **23**, 189-201.

7. Karousis, E.D., Nasif, S. and Muhlemann, O. (2016) Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip Rev RNA*, **7**, 661-682.

8. Nasif, S., Contu, L. and Muhlemann, O. (2018) Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. *Semin Cell Dev Biol*, **75**, 78-87.

9. 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.

10. Le Hir, H., Sauliere, J. and Wang, Z. (2016) The exon junction complex as a node of post-transcriptional networks. *Nat Rev Mol Cell Biol*, **17**, 41-54.

11. Lejeune, F., Ishigaki, Y., Li, X. and Maquat, L.E. (2002) The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J*, **21**, 3536-3545.

12. Schlautmann, L.P. and Gehring, N.H. (2020) A Day in the Life of the Exon Junction Complex. *Biomolecules*, **10**.

13. Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G. and Ohno, S. (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev*, **20**, 355-367.

14. Hug, N. and Caceres, J.F. (2014) The RNA helicase DHX34 activates NMD by promoting a transition from the surveillance to the decay-inducing complex. *Cell reports*, **8**, 1845-1856.

15. Melero, R., Hug, N., Lopez-Perrote, A., Yamashita, A., Caceres, J.F. and Llorca, O. (2016) The RNA helicase DHX34 functions as a scaffold for SMG1-mediated UPF1 phosphorylation. *Nat Commun*, **7**, 10585.

16. Chamieh, H., Ballut, L., Bonneau, F. and Le Hir, H. (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat Struct Mol Biol*, **15**, 85-93.

17. Lopez-Perrote, A., Castano, R., Melero, R., Zamarro, T., Kurosawa, H., Ohnishi, T., Uchiyama, A., Aoyagi, K., Buchwald, G., Kataoka, N. *et al.* (2016) Human nonsense-mediated mRNA decay factor UPF2 interacts directly with eRF3 and the SURF complex. *Nucleic Acids Res*, **44**, 1909-1923.

18. Boehm, V., Haberman, N., Ottens, F., Ule, J. and Gehring, N.H. (2014) 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep*, **9**, 555-568.

19. Metze, S., Herzog, V.A., Ruepp, M.D. and Muhlemann, O. (2013) Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. *RNA*, **19**, 1432-1448.

20. Karousis, E.D., Gurzeler, L.A., Annibaldis, G., Dreos, R. and Muhlemann, O. (2020) Human NMD ensues independently of stable ribosome stalling. *Nat Commun*, **11**, 4134.

21. Hoek, T.A., Khuperkar, D., Lindeboom, R.G.H., Sonneveld, S., Verhagen, B.M.P., Boersma, S., Vermeulen, M. and Tanenbaum, M.E. (2019) Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. *Mol Cell*, **75**, 324-339 e311.

22. Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F. and Ohno, S. (2012) N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res*, **40**, 1251-1266.

23. Loh, B., Jonas, S. and Izaurralde, E. (2013) The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev*, **27**, 2125-2138.

24. Gatfield, D. and Izaurralde, E. (2004) Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. *Nature*, **429**, 575-578.

25. Nicholson, P., Josi, C., Kurosawa, H., Yamashita, A. and Muhlemann, O. (2014) A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. *Nucleic Acids Res*, **42**, 9217-9235.

26. Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W. and Maquat, L.E. (2008) Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell*, **133**, 314-327.

27. Miller, J.E. and Reese, J.C. (2012) Ccr4-Not complex: the control freak of eukaryotic cells. *Crit Rev Biochem Mol Biol*, **47**, 315-333.

28. Chen, Y., Boland, A., Kuzuoglu-Ozturk, D., Bawankar, P., Loh, B., Chang, C.T., Weichenrieder, O. and Izaurralde, E. (2014) A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. *Mol Cell*, **54**, 737-750.

29. Coller, J. and Parker, R. (2005) General translational repression by activators of mRNA decapping. *Cell*, **122**, 875-886.

30. Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E. and Filipowicz, W. (2014) Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol Cell*, **54**, 751-765.

31. Jonas, S. and Izaurralde, E. (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet*, **16**, 421-433.

32. Ferraiuolo, M.A., Basak, S., Dostie, J., Murray, E.L., Schoenberg, D.R. and Sonenberg, N. (2005) A role for the eIF4E-binding protein 4E-T in P-body formation and mRNA decay. *J Cell Biol*, **170**, 913-924.

33. Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W. and Kulozik, A.E. (2005) Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol Cell*, **20**, 65-75.

34. Chan, W.K., Huang, L., Gudikote, J.P., Chang, Y.F., Imam, J.S., MacLean, J.A., 2nd and Wilkinson, M.F. (2007) An alternative branch of the nonsense-mediated decay pathway. *EMBO J*, **26**, 1820-1830.

35. Toma, K.G., Rebbapragada, I., Durand, S. and Lykke-Andersen, J. (2015) Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. *RNA*, **21**, 887-897.

36. Ge, Z., Quek, B.L., Beemon, K.L. and Hogg, J.R. (2016) Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. *Elife*, **5**.

37. Kishor, A., Ge, Z. and Hogg, J.R. (2019) hnRNP L-dependent protection of normal mRNAs from NMD subverts quality control in B cell lymphoma. *EMBO J*, **38**.

38. Yepiskoposyan, H., Aeschimann, F., Nilsson, D., Okoniewski, M. and Muhlemann, O. (2011) Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA*, **17**, 2108-2118.

39. Huang, L., Lou, C.H., Chan, W., Shum, E.Y., Shao, A., Stone, E., Karam, R., Song, H.W. and Wilkinson, M.F. (2011) RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. *Mol Cell*, **43**, 950-961.

40. Longman, D., Hug, N., Keith, M., Anastasaki, C., Patton, E.E., Grimes, G. and Caceres, J.F. (2013) DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and Caenorhabditis elegans. *Nucleic acids research*, **41**, 8319-8331.

41. Joncourt, R., Eberle, A.B., Rufener, S.C. and Muhlemann, O. (2014) Eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay by two genetically separable mechanisms. *PLoS One*, **9**, e104391.

42. Fatscher, T., Boehm, V., Weiche, B. and Gehring, N.H. (2014) The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay. *RNA*, **20**, 1579-1592.

43. Zhang, Z., Zhou, L., Hu, L., Zhu, Y., Xu, H., Liu, Y., Chen, X., Yi, X., Kong, X. and Hurst, L.D. (2010) Nonsense-mediated decay targets have multiple sequence-related features that can inhibit translation. *Mol Syst Biol*, **6**, 442.

44. Bruno, I.G., Karam, R., Huang, L., Bhardwaj, A., Lou, C.H., Shum, E.Y., Song, H.W., Corbett, M.A., Gifford, W.D., Gecz, J. *et al.* (2011) Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. *Mol Cell*, **42**, 500-510.

45. Lou, C.H., Shao, A., Shum, E.Y., Espinoza, J.L., Huang, L., Karam, R. and Wilkinson, M.F. (2014) Posttranscriptional control of the stem cell and neurogenic programs by the nonsense-mediated RNA decay pathway. *Cell Rep*, **6**, 748-764.

46. Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res*, **47**, D155-D162.

47. Meunier, J., Lemoine, F., Soumillon, M., Liechti, A., Weier, M., Guschanski, K., Hu, H., Khaitovich, P. and Kaessmann, H. (2013) Birth and expression evolution of mammalian microRNA genes. *Genome Res*, **23**, 34-45.

48. Guidi, M., Muinos-Gimeno, M., Kagerbauer, B., Marti, E., Estivill, X. and Espinosa-Parrilla, Y. (2010) Overexpression of miR-128 specifically inhibits the truncated isoform of NTRK3 and upregulates BCL2 in SH-SY5Y neuroblastoma cells. *BMC Mol Biol*, **11**, 95.

49. Godlewski, J., Nowicki, M.O., Bronisz, A., Williams, S., Otsuki, A., Nuovo, G., Raychaudhury, A., Newton, H.B., Chiocca, E.A. and Lawler, S. (2008) Targeting of the Bmi-1 oncogene/stem cell renewal factor by microRNA-128 inhibits glioma proliferation and self-renewal. *Cancer Res*, **68**, 9125-9130.

50. Papagiannakopoulos, T., Friedmann-Morvinski, D., Neveu, P., Dugas, J.C., Gill, R.M., Huillard, E., Liu, C., Zong, H., Rowitch, D.H., Barres, B.A. *et al.* (2012) Pro-neural miR-128 is a glioma tumor suppressor that targets mitogenic kinases. *Oncogene*, **31**, 1884-1895.

51. Chen, D.Z., Wang, W.W., Chen, Y.L., Yang, X.F., Zhao, M. and Yang, Y.Y. (2019) miR128 is upregulated in epilepsy and promotes apoptosis through the SIRT1 cascade. *Int J Mol Med*, **44**, 694-704.

52. Ye, Y., Zhi, F., Peng, Y. and Yang, C.C. (2018) MiR-128 promotes the apoptosis of glioma cells via binding to NEK2. *Eur Rev Med Pharmacol Sci*, **22**, 8781-8788.

53. Zhao, D., Han, W., Liu, X., Cui, D. and Chen, Y. (2017) MicroRNA-128 promotes apoptosis in lung cancer by directly targeting NIMA-related kinase 2. *Thorac Cancer*, **8**, 304-311.

54. Shang, C., Hong, Y., Guo, Y., Liu, Y.H. and Xue, Y.X. (2016) miR-128 regulates the apoptosis and proliferation of glioma cells by targeting RhoE. *Oncol Lett*, **11**, 904-908.

55. Lian, B., Yang, D., Liu, Y., Shi, G., Li, J., Yan, X., Jin, K., Liu, X., Zhao, J., Shang, W. *et al.* (2018) miR-128 Targets the SIRT1/ROS/DR5 Pathway to Sensitize Colorectal Cancer to TRAIL-Induced Apoptosis. *Cell Physiol Biochem*, **49**, 2151-2162.

56. Han, H., Wang, L., Xu, J. and Wang, A. (2018) miR-128 induces pancreas cancer cell apoptosis by targeting MDM4. *Exp Ther Med*, **15**, 5017-5022.

57. Mets, E., Van Peer, G., Van der Meulen, J., Boice, M., Taghon, T., Goossens, S., Mestdagh, P., Benoit, Y., De Moerloose, B., Van Roy, N. *et al.* (2014) MicroRNA-128-3p is a novel oncomiR targeting PHF6 in T-cell acute lymphoblastic leukemia. *Haematologica*, **99**, 1326-1333.

58. Adlakha, Y.K. and Saini, N. (2014) Brain microRNAs and insights into biological functions and therapeutic potential of brain enriched miRNA-128. *Mol Cancer*, **13**, 33.

59. Donzelli, S., Fontemaggi, G., Fazi, F., Di Agostino, S., Padula, F., Biagioni, F., Muti, P., Strano, S. and Blandino, G. (2012) MicroRNA-128-2 targets the transcriptional repressor E2F5 enhancing mutant p53 gain of function. *Cell Death Differ*, **19**, 1038-1048.

60. Tan, C.L., Plotkin, J.L., Veno, M.T., von Schimmelmann, M., Feinberg, P., Mann, S., Handler, A., Kjems, J., Surmeier, D.J., O'Carroll, D. *et al.* (2013) MicroRNA-128 governs neuronal excitability and motor behavior in mice. *Science*, **342**, 1254-1258.

61. Ramos, J.W., Hughes, P.E., Renshaw, M.W., Schwartz, M.A., Formstecher, E., Chneiweiss, H. and Ginsberg, M.H. (2000) Death effector domain protein PEA-15 potentiates Ras activation of extracellular signal receptor-activated kinase by an adhesion-independent mechanism. *Mol Biol Cell*, **11**, 2863-2872.

62. Vaidyanathan, H., Opoku-Ansah, J., Pastorino, S., Renganathan, H., Matter, M.L. and Ramos, J.W. (2007) ERK MAP kinase is targeted to RSK2 by the phosphoprotein PEA-15. *Proc Natl Acad Sci U S A*, **104**, 19837-19842.

63. Formstecher, E., Ramos, J.W., Fauquet, M., Calderwood, D.A., Hsieh, J.C., Canton, B., Nguyen, X.T., Barnier, J.V., Camonis, J., Ginsberg, M.H. *et al.* (2001) PEA-15 mediates cytoplasmic sequestration of ERK MAP kinase. *Dev Cell*, **1**, 239-250.

64. Matsuda, A., Suzuki, Y., Honda, G., Muramatsu, S., Matsuzaki, O., Nagano, Y., Doi, T., Shimotohno, K., Harada, T., Nishida, E. *et al.* (2003) Large-scale identification and

characterization of human genes that activate NF-kappaB and MAPK signaling pathways. *Oncogene*, **22**, 3307-3318.

65.   Zhao, N., Zhang, G., He, M., Huang, H., Cao, L., Yin, A., Wang, P. and Wang, L. (2017) SZRD1 is a Novel Protein that Functions as a Potential Tumor Suppressor in Cervical Cancer. *J Cancer*, **8**, 2132-2141.

66.   Grishin, N.V. (1998) The R3H motif: a domain that binds single-stranded nucleic acids. *Trends Biochem Sci*, **23**, 329-330.

67.   He, G.J., Zhang, A., Liu, W.F. and Yan, Y.B. (2013) Distinct roles of the R3H and RRM domains in poly(A)-specific ribonuclease structural integrity and catalysis. *Biochim Biophys Acta*, **1834**, 1089-1098.

68.   Song, M.H., Aravind, L., Muller-Reichert, T. and O'Connell, K.F. (2008) The conserved protein SZY-20 opposes the Plk4-related kinase ZYG-1 to limit centrosome size. *Dev Cell*, **15**, 901-912.

69.   Carissimi, C., Baccon, J., Straccia, M., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J. and Pellizzoni, L. (2005) Unrip is a component of SMN complexes active in snRNP assembly. *FEBS Lett*, **579**, 2348-2354.

70.   Hunt, S.L., Hsuan, J.J., Totty, N. and Jackson, R.J. (1999) unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA. *Genes Dev*, **13**, 437-448.

71.   Vukmirovic, M., Manojlovic, Z. and Stefanovic, B. (2013) Serine-threonine kinase receptor-associated protein (STRAP) regulates translation of type I collagen mRNAs. *Mol Cell Biol*, **33**, 3893-3906.

72.   Gerin, I., Ury, B., Breloy, I., Bouchet-Seraphin, C., Bolsee, J., Halbout, M., Graff, J., Vertommen, D., Muccioli, G.G., Seta, N. *et al.* (2016) ISPD produces CDP-ribitol used by FKTN and FKRP to transfer ribitol phosphate onto alpha-dystroglycan. *Nat Commun*, **7**, 11534.

73.   Kim, H., Chen, J. and Yu, X. (2007) Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science*, **316**, 1202-1205.

74.   Heckman, K.L. and Pease, L.R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc*, **2**, 924-932.

75.   Consortium, G.T. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet*, **45**, 580-585.

76.   Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D.Y., Barbosa, I.A., Kwon, J.S., Guan, Y. *et al.* (2013) An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep*, **5**, 1704-1713.

77.   Pelossof, R., Fairchild, L., Huang, C.H., Widmer, C., Sreedharan, V.T., Sinha, N., Lai, D.Y., Guan, Y., Premsrirut, P.K., Tschaharganeh, D.F. *et al.* (2017) Prediction of potent shRNAs with a sequential classification algorithm. *Nat Biotechnol*, **35**, 350-353.

78.   Boelz, S., Neu-Yilik, G., Gehring, N.H., Hentze, M.W. and Kulozik, A.E. (2006) A chemiluminescence-based reporter system to monitor nonsense-mediated mRNA decay. *Biochem Biophys Res Commun*, **349**, 186-191.

79.   Gehring, N.H., Hentze, M.W. and Kulozik, A.E. (2008) Tethering assays to investigate nonsense-mediated mRNA decay activating proteins. *Methods Enzymol*, **448**, 467-482.

80.   Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.

81.   Sheridan, R.M. and Bentley, D.L. (2016) Selectable one-step PCR-mediated integration of a degron for rapid depletion of endogenous human proteins. *Biotechniques*, **60**, 69-74.

82.   Luo, Y., Lin, L., Bolund, L. and Sorensen, C.B. (2014) Efficient construction of rAAV-based gene targeting vectors by Golden Gate cloning. *Biotechniques*, **56**, 263-268.

83.   Jordan, M., Schallhorn, A. and Wurm, F.M. (1996) Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic acids research*, **24**, 596-601.

84.   Collard, F., Baldin, F., Gerin, I., Bolsee, J., Noel, G., Graff, J., Veiga-da-Cunha, M., Stroobant, V., Vertommen, D., Houddane, A. *et al.* (2016) A conserved phosphatase destroys toxic glycolytic side products in mammals and yeast. *Nat Chem Biol*, **12**, 601-607.

85.   Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

86.   Gerin, I., Noel, G., Bolsee, J., Haumont, O., Van Schaftingen, E. and Bommer, G.T. (2014) Identification of TP53-induced glycolysis and apoptosis regulator (TIGAR) as the

phosphoglycolate-independent 2,3-bisphosphoglycerate phosphatase. *Biochem J*, **458**, 439-448.

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

88.   Steven W, W. and Simon, A. (2018) FastQ Screen: A tool for multi-genome mapping and quality control *F1000Research*, **7**, 1338.

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

90.   Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*, **37**, 907-915.

91.   Schneider, V.A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.C., Kitts, P.A., Murphy, T.D., Pruitt, K.D., Thibaud-Nissen, F., Albracht, D. *et al.* (2017) Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res*, **27**, 849-864.

92.   Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

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

94.   Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T. *et al.* (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*, **12**, 115-121.

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

96.   Blighe, K.R., S.; Lewis, M. (2020), *R package version 1.6.0,* , https://github.com/kevinblighe/EnhancedVolcano. .

97.   Korotkevich, G., Sukhov, V. and Sergushichev, A. (2019) Fast gene set enrichment analysis. *bioRxiv*, **doi: 10.1101/060012**.

98.   Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, **102**, 15545-15550.

99.   Gerin, I., Clerbaux, L.A., Haumont, O., Lanthier, N., Das, A.K., Burant, C.F., Leclercq, I.A., MacDougald, O.A. and Bommer, G.T. (2010) Expression of miR-33 from an SREBP2 intron inhibits cholesterol export and fatty acid oxidation. *J Biol Chem*, **285**, 33652-33661.

100.   Geng, L., Huntoon, C.J. and Karnitz, L.M. (2010) RAD18-mediated ubiquitination of PCNA activates the Fanconi anemia DNA repair network. *J Cell Biol*, **191**, 249-257.

101.   Leeds, P., Peltz, S.W., Jacobson, A. and Culbertson, M.R. (1991) The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev*, **5**, 2303-2314.

102.   Sato, H. and Maquat, L.E. (2009) Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin beta. *Genes Dev*, **23**, 2537-2550.

103.   Bai, Y., Wang, W., Li, S., Zhan, J., Li, H., Zhao, M., Zhou, X.A., Li, S., Li, X., Huo, Y. *et al.* (2019) C1QBP Promotes Homologous Recombination by Stabilizing MRE11 and Controlling the Assembly and Activation of MRE11/RAD50/NBS1 Complex. *Mol Cell*, **75**, 1299-1314 e1296.

104.   McGee, A.M., Douglas, D.L., Liang, Y., Hyder, S.M. and Baines, C.P. (2011) The mitochondrial protein C1qbp promotes cell proliferation, migration and resistance to cell death. *Cell Cycle*, **10**, 4119-4127.

105.   Petersen-Mahrt, S.K., Estmer, C., Ohrmalm, C., Matthews, D.A., Russell, W.C. and Akusjarvi, G. (1999) The splicing factor-associated protein, p32, regulates RNA splicing by inhibiting ASF/SF2 RNA binding and phosphorylation. *EMBO J*, **18**, 1014-1024.

106.   Yoshikawa, H., Ishikawa, H., Izumikawa, K., Miura, Y., Hayano, T., Isobe, T., Simpson, R.J. and Takahashi, N. (2015) Human nucleolar protein Nop52 (RRP1/NNP-1) is involved in site 2 cleavage in internal transcribed spacer 1 of pre-rRNAs at early stages of ribosome biogenesis. *Nucleic Acids Res*, **43**, 5524-5536.

107.   Flury, V., Restuccia, U., Bachi, A. and Muhlemann, O. (2014) Characterization of phosphorylation- and RNA-dependent UPF1 interactors by quantitative proteomics. *J Proteome Res*, **13**, 3038-3053.

108. Gong, C. and Maquat, L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature*, **470**, 284-288.

109. Choe, J., Ahn, S.H. and Kim, Y.K. (2014) The mRNP remodeling mediated by UPF1 promotes rapid degradation of replication-dependent histone mRNA. *Nucleic Acids Res*, **42**, 9334-9349.

110. Watkins, N.J. and Bohnsack, M.T. (2012) The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA*, **3**, 397-414.

111. Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A. and Jensen, T.H. (2014) Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev*, **28**, 2498-2517.

112. Krastev, D.B., Slabicki, M., Paszkowski-Rogacz, M., Hubner, N.C., Junqueira, M., Shevchenko, A., Mann, M., Neugebauer, K.M. and Buchholz, F. (2011) A systematic RNAi synthetic interaction screen reveals a link between p53 and snoRNP assembly. *Nat Cell Biol*, **13**, 809-818.

113. Zhang, J., Sun, X., Qian, Y. and Maquat, L.E. (1998) Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. *RNA*, **4**, 801-815.

114. Shaw, G. and Kamen, R. (1986) A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell*, **46**, 659-667.

115. Jonas, S., Weichenrieder, O. and Izaurralde, E. (2013) An unusual arrangement of two 14-3-3-like domains in the SMG5-SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. *Genes Dev*, **27**, 211-225.

116. Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J. and Izaurralde, E. (2008) SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA*, **14**, 2609-2617.

117. Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E. and Filipowicz, W. (2005) Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science*, **309**, 1573-1576.

118. Gerbracht, J.V., Boehm, V., Britto-Borges, T., Kallabis, S., Wiederstein, J.L., Ciriello, S., Aschemeier, D.U., Kruger, M., Frese, C.K., Altmuller, J. *et al.* (2020) CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex. *Nucleic Acids Res*.

119. McGlincy, N.J., Tan, L.Y., Paul, N., Zavolan, M., Lilley, K.S. and Smith, C.W. (2010) Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. *BMC Genomics*, **11**, 565.

120. Tani, H., Torimura, M. and Akimitsu, N. (2013) The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. *PLoS One*, **8**, e55684.

121. Sieber, J., Hauer, C., Bhuvanagiri, M., Leicht, S., Krijgsveld, J., Neu-Yilik, G., Hentze, M.W. and Kulozik, A.E. (2016) Proteomic Analysis Reveals Branch-specific Regulation of the Unfolded Protein Response by Nonsense-mediated mRNA Decay. *Mol Cell Proteomics*, **15**, 1584-1597.

122. Rehfeld, F., Maticzka, D., Grosser, S., Knauff, P., Eravci, M., Vida, I., Backofen, R. and Wulczyn, F.G. (2018) The RNA-binding protein ARPP21 controls dendritic branching by functionally opposing the miRNA it hosts. *Nat Commun*, **9**, 1235.

123. Agarwal, V., Bell, G.W., Nam, J.W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**.

124. Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Vlachos, I.S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T. and Hatzigeorgiou, A.G. (2013) DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res*, **41**, W169-173.

125. Hwang, J., Sato, H., Tang, Y., Matsuda, D. and Maquat, L.E. (2010) UPF1 association with the cap-binding protein, CBP80, promotes nonsense-mediated mRNA decay at two distinct steps. *Mol Cell*, **39**, 396-409.

126. Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W. and Kulozik, A.E. (2008) Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J*, **27**, 736-747.
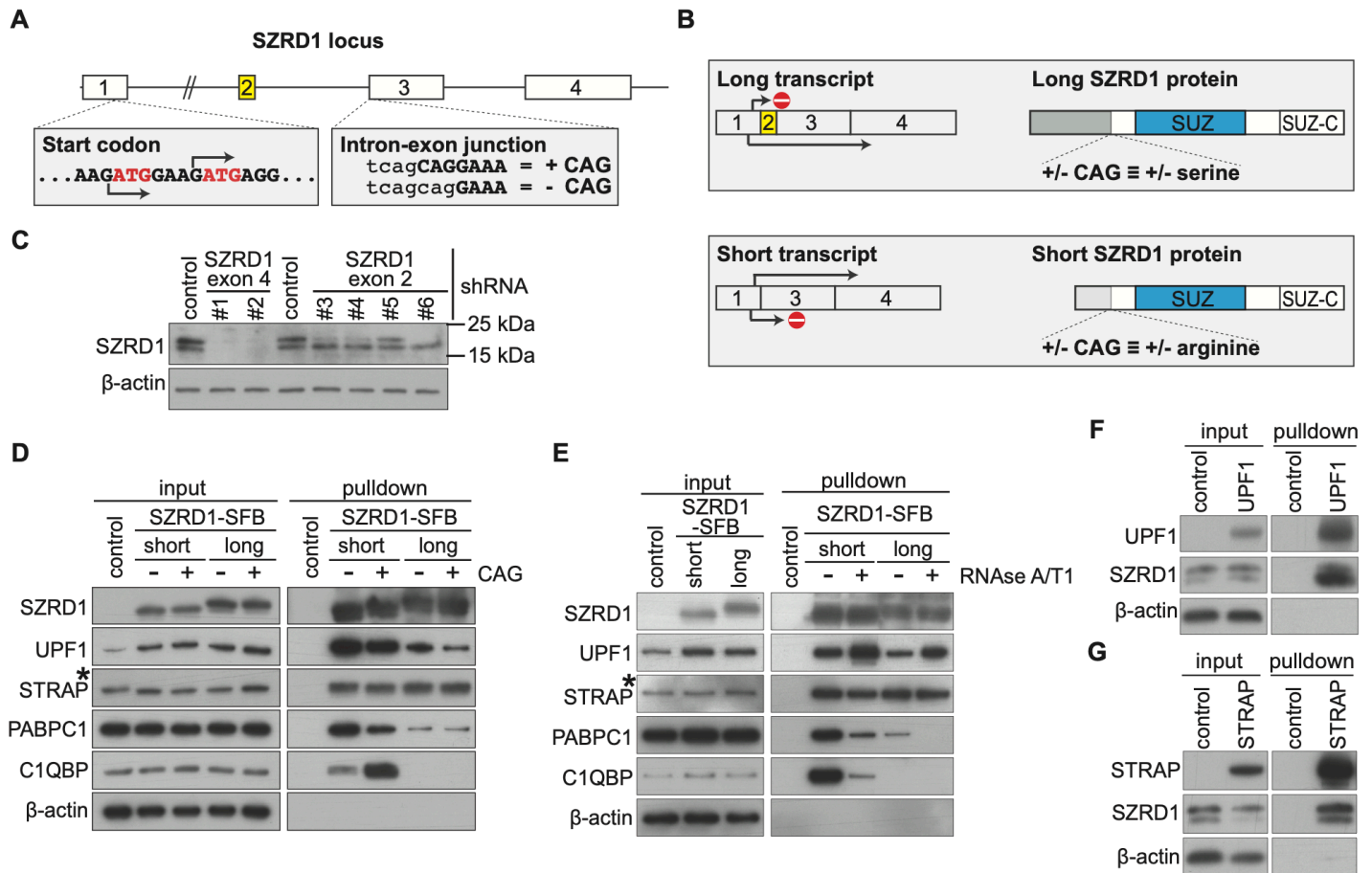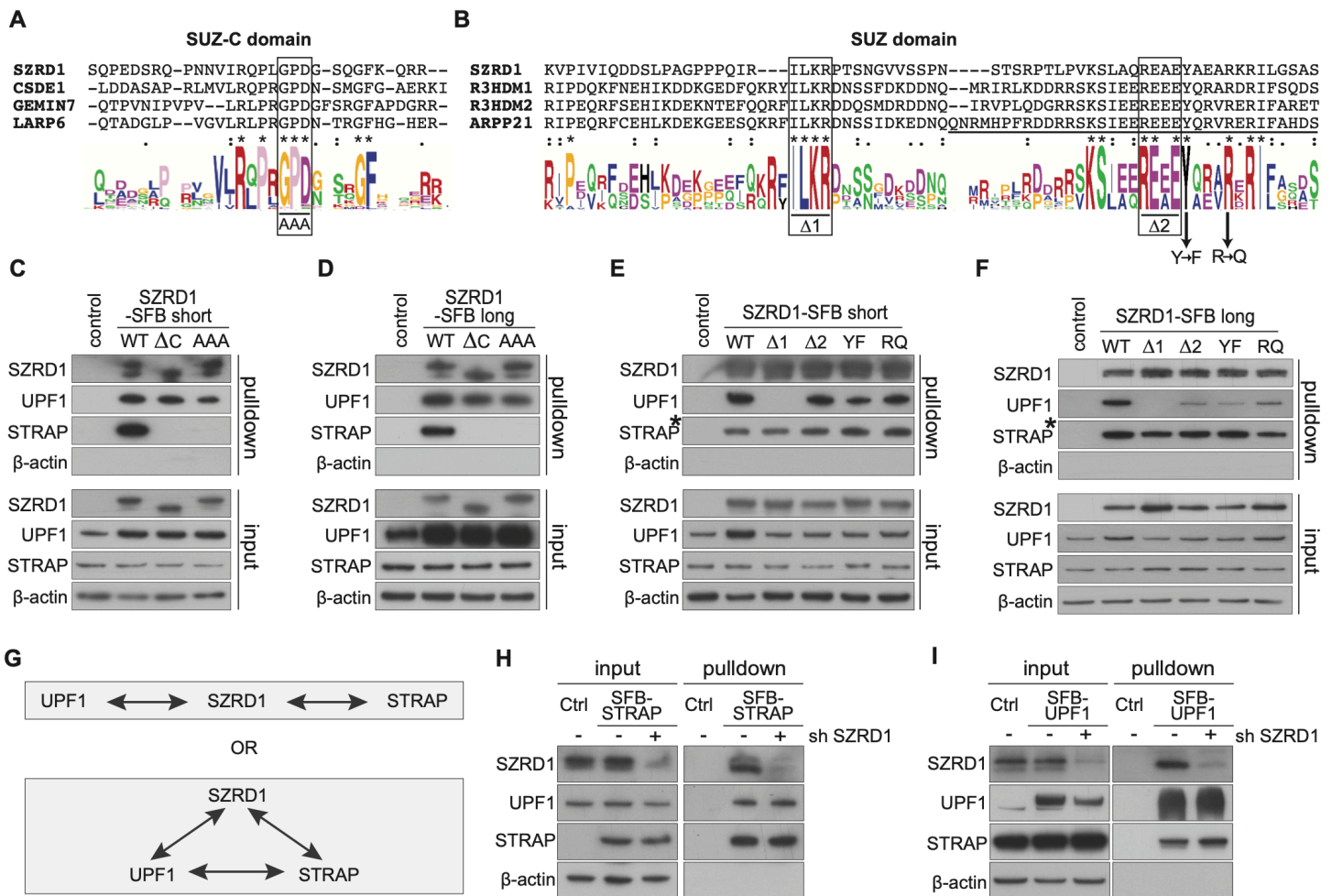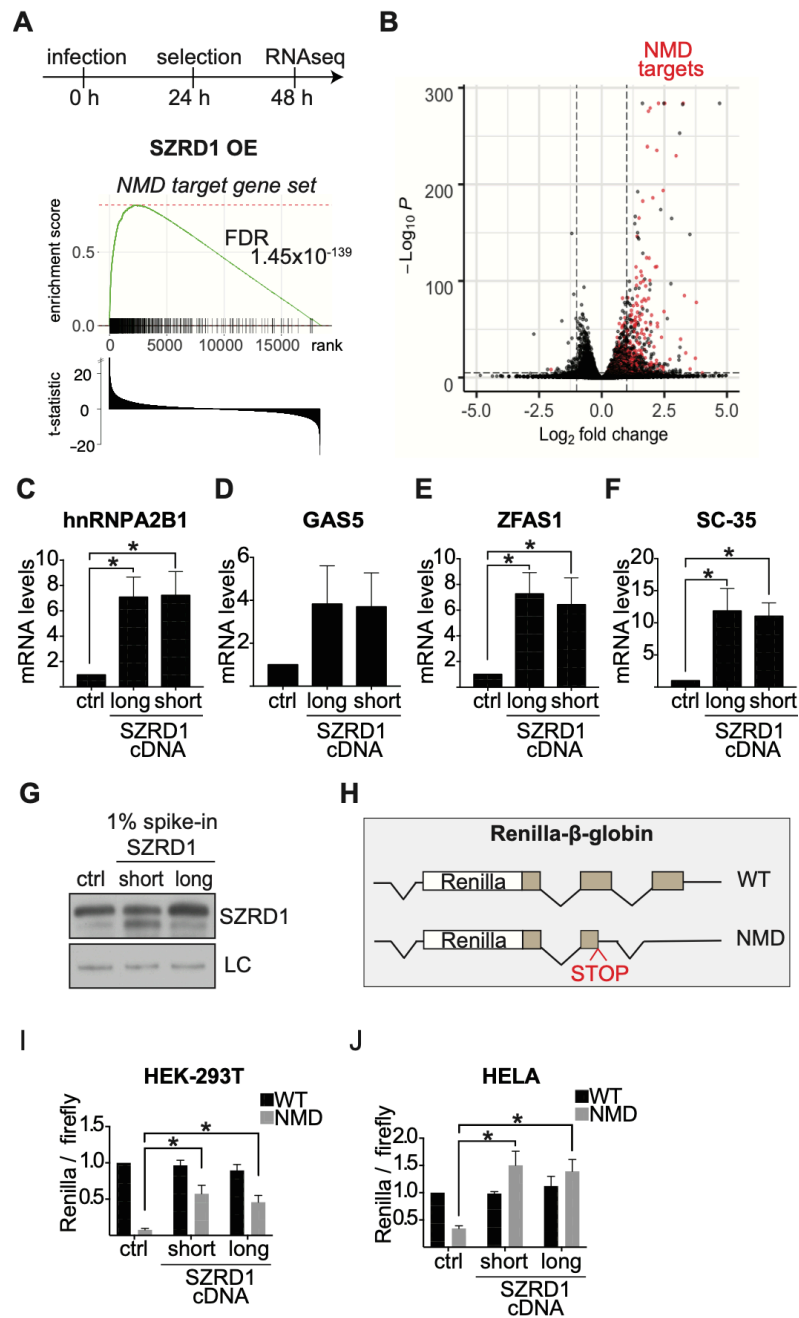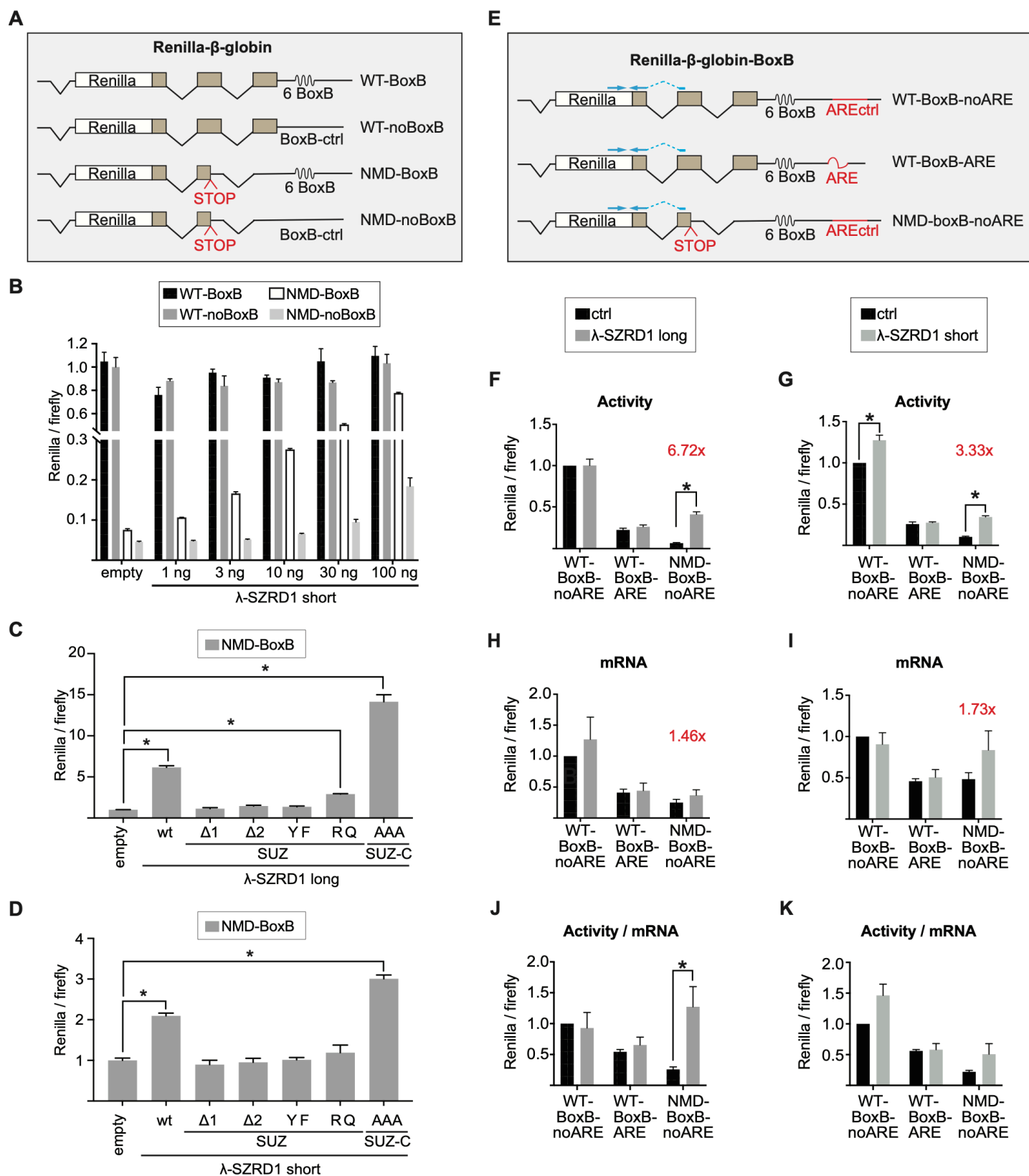
Figure 1

Figure 2

A

infection   selection   RNAseq

0 h   24 h   48 h

SZRD1 OE

NMD target gene set

FDR
1.45×10^{-139}

enrichment score

0.5

0.0

0   5000   10000   15000   rank

t-statistic

20

0

−20

B

NMD
targets

$-\mathrm{Log}_{10}\,P$

300

200

100

−5.0   −2.5   0   2.5   5.0

Log_2 fold change

C   **hnRNPA2B1**

mRNA levels

10
8
6
4
2

ctrl   long   short
SZRD1
cDNA

D   **GAS5**

mRNA levels

6

4

2

ctrl   long   short
SZRD1
cDNA

E   **ZFAS1**

mRNA levels

10
8
6
4
2

ctrl   long   short
SZRD1
cDNA

F   **SC-35**

mRNA levels

20
15
10
5

ctrl   long   short
SZRD1
cDNA

G

1% spike-in
SZRD1

ctrl   short   long

SZRD1

LC

H

Renilla-β-globin

Renilla   WT

Renilla   NMD
STOP

I   **HEK-293T**

Renilla / firefly

1.5
1.0
0.5

ctrl   short   long
SZRD1
cDNA

■WT
▨NMD

J   **HELA**

Renilla / firefly

2.0
1.5
1.0
0.5

ctrl   short   long
SZRD1
cDNA

■WT
▨NMD

Figure 3

Figure 4

A

Endogenous tag method

CRISPR/CAS9

target gene ···· [ORF [ ][ ] 3'UTR] ····

donor [LH | SFB | P2A | PURO | RH]

B

	input	pulldown
	knock-In (SFB)	knock-In (SFB)

Ctrl UPF1 eIF4A3 RNPS1   Ctrl UPF1 eIF4A3 RNPS1

FLAG *

SZRD1

UPF2

UPF3B

MLN51

Y14

GAPDH

Figure 5

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 6

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A

Published ARPP21 model:
increased protein output via 3'UTR
and translational initiation complex

5' UUUUU ———— AAA 3'

eIF4F
complex

Rehfeld et al. 2018

B

Our hypothesis

SZRD1

ARPP21 **?**

Modulation of
subsets of
NMD targets
via SUZ domain

■ SUZ domain

C

PULLDOWN

Ctrl   SFB-UPF1

−   −   +   −   RNAse A/T1
−   −   −   +   shRNA R3HDM1

FLAG

R3HDM1

β-actin

INPUT

Ctrl   SFB-UPF1

−   +   −   +   shRNA R3HDM1

FLAG

R3HDM1

β-actin

D

ARPP21 OE

NMD target gene set

FDR
1.34×10^{-9}

E

■ ctrl
■ λ-ARPP21 full

Activity

6xBoxB   +   −   +   −
         WT   WT NMD NMD

F

■ ctrl
■ λ-ARPP21 full
■ λ-ARPP21ΔSUZ

Activity

WT-   WT-   NMD-
BoxB-   BoxB-   BoxB-
noARE   ARE   noARE

G

mRNA

WT-   WT-   NMD-
BoxB-   BoxB-   BoxB-
noARE   ARE   noARE

H

Activity / mRNA

WT-   WT-   NMD-
BoxB-   BoxB-   BoxB-
noARE   ARE   noARE

Figure 7

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

A

B

C

HEK293

D

Jurkat

Figure 8

36
37
38
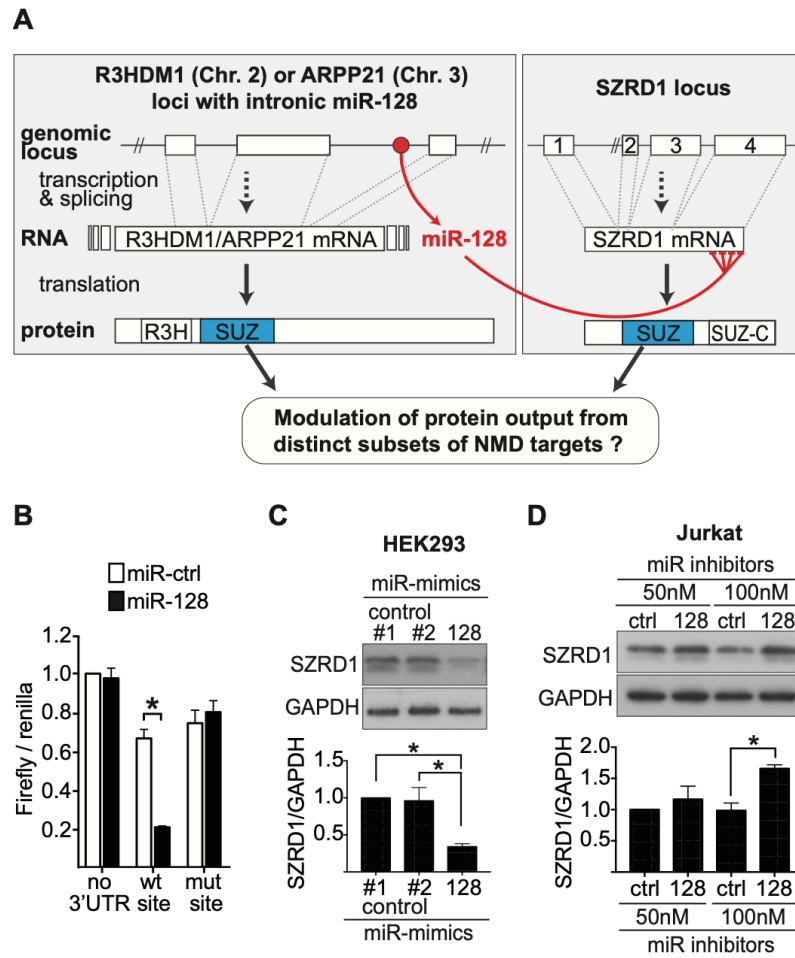39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Modulating protein production from nonsense-mediated decay targets by SUZ domain-containing proteins.

Mathias Halbout[1,2], Marina Bury[1,2], Aoife Hanet[1,2], Isabelle Gerin[1,2], Julie Graff[1,2], Theodore Killian[3], Laurent Gatto[3], Didier Vertommen[4], Guido Bommer[1,2*]

[1] Department of Physiological Chemistry, de Duve Institute, UCLouvain, 1200 Brussels, Belgium

[2] WELBIO, 1200 Brussels, Belgium

[3] Computational Biology Laboratory, de Duve Institute, UCLouvain, 1200 Bruxelles, Belgium

[4] Protein Phosphorylation Unit, de Duve Institute, UCLouvain, 1200 Brussels, Belgium

* To whom correspondence should be addressed. Tel: +32-2 7647568; Fax: +32-2-7647598; Email: guido.bommer@uclouvain.be

Supplementary tables S1 – S7

Supplementary figures S1 – S8

Table S1.
List of primers used for cloning plasmids.

Table S2.
List of antibodies.

Table S3
List of plasmids.

Table S4.
List of primers used for qPCR.

Table S5.
Enrichment scores for gene sets in SZRD1 overexpression RNAseq dataset.

Table S6
Enrichment scores for gene sets in SZRD1 knockdown RNAseq dataset.

Table S7.
Enrichment scores for gene sets in ARPP21 overexpression RNAseq dataset.

A

long SZRD1 isoforms

```
Homo sapiens  MED-EEVAESWEEAADSGE--IDRRLEKKLKITQKE----------SRKSKSPPKVPIVIQDDSLP-AGPPPQIRILKRP...
Mus musculus  MED-EEVAESWEEAADSGE--IDRRLEKKLKITQKE----------SRKSKSPPKVPIVIQDDSLP-TGPPPQIRILKRP...
X. tropicalis MEE-DEVAESWEEAADSGE--IDRRLEKKLKISQRE----------NSKSKSPPKAPVVIQDDSLP-SGPPPQIRILKRP...
Danio rerio   MDD-EEVAESWEEAADSGE--MERRLEEKLRISQKERLS------SGSSSRSPMRTAIVIQDDSLP-AAPPPQIRILKRP...
C. elegans    MSKENVVADSWDDADADPVKELMDKVE-KVKLLQRKEEKKEAFFEKVKAEESSGVVSKLQTEEGLGPSAEEPKRVFLRRP...
```

B

SZRD1 transcripts

SZRD1 long

SZRD1 short

```
Mus musculus  ATGGAAGATGAGGAGGTCGCTGAGAGCTGGGAGGAGGCGGCAGACAGCGGG...
G. gallus     ATGGAAGATGAGGAGGTGGCCGAGAGCTGGGAGGAGGCGGCCGACAGCGGG...
X. tropicalis ATGGAAGAAGATGAAGTCGCGGAGAGTTGGGAAGAAGCGGCCGATAGCGGA...
Danio rerio   ATGGATGATGAGGAGGTCGCCGAGAGCTGGGAAGAGGCTGCCGACAGCGGG...
```
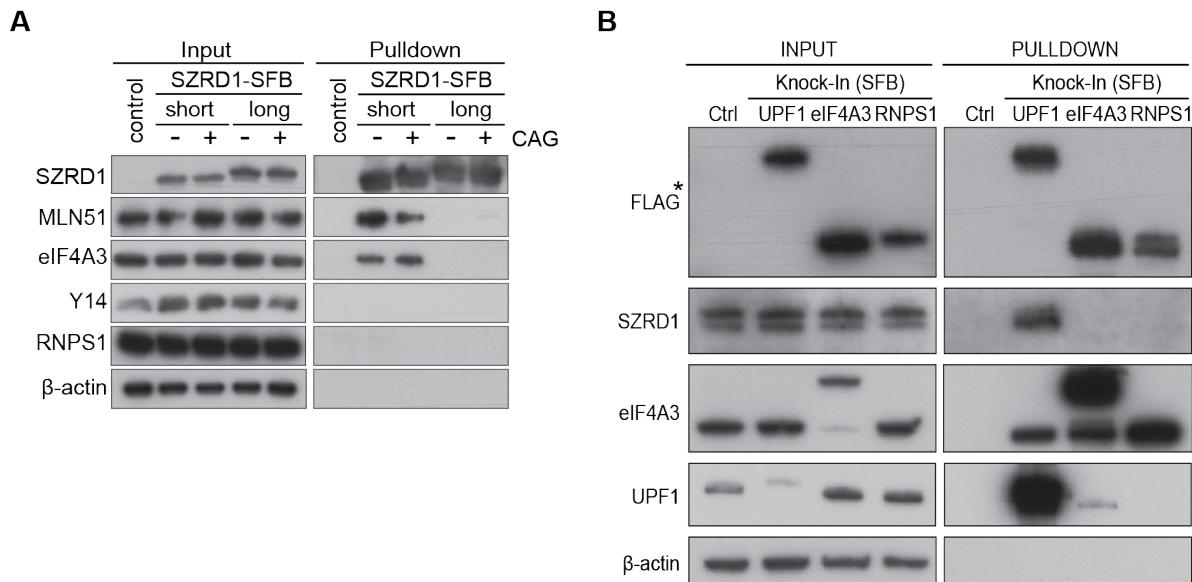
Figure S1.
(A) Amino acid sequence alignment of the N-terminal part of the long SZRD1 isoform in *Homo sapiens*, *Mus musculus*, *Xenopus tropicalis* ('*X. trop*icalis'), Danio rerio and *Caenorhabditis elegans ('C. elegans')* with differences highlighted in brown. **(B)** Nucleotide alignment of the beginning of the *SZRD1* coding sequence from the indicated species ('*G. gallus*' = *Gallus gallus*) with differences highlighted in brown. Amino acids sequences of human long and short SZRD1 are indicated. Vertical lines delimitate the codons in the open reading frame of the long SZRD1 isoform. All alignments were done with the use of CLUSTAL W (1,2).
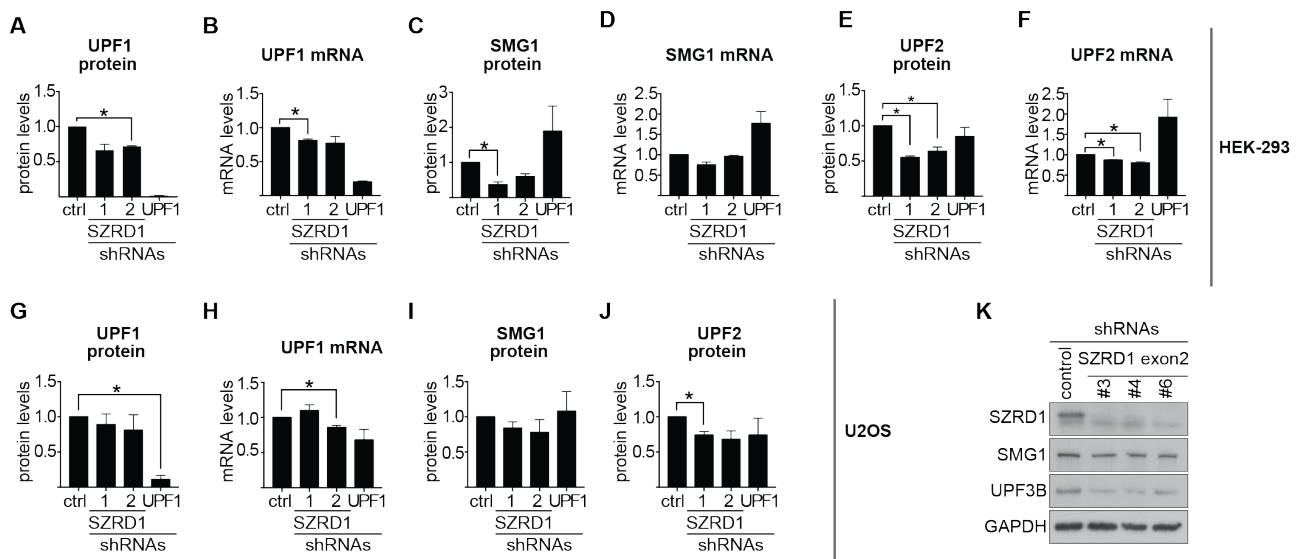
Figure S2.
(A) Western blot analysis with anti-SZRD1 antibody for the samples used in the RNAseq experiment shown in Figure 3 A & B. **(B-D)** Histone (B), SMD target transcripts (C) and snoRNA precursor abundances (D) were extracted from the RNAseq dataset obtained from HeLa cells upon transduction with recombinant lentiviruses driving expression of SZRD1 ("SZRD1 OE", grey squares) or an empty control lentivirus ("Control OE", black circles). **(E)** Northern blot analysis of SNORD and SNORA RNAs produced from the transcripts GAS5, ZFAS1 and SNHG15 (shown in D). Hybridization with a tRNA-E probe and an ethidium bromide stain ("RNA loading") are shown as loading controls.
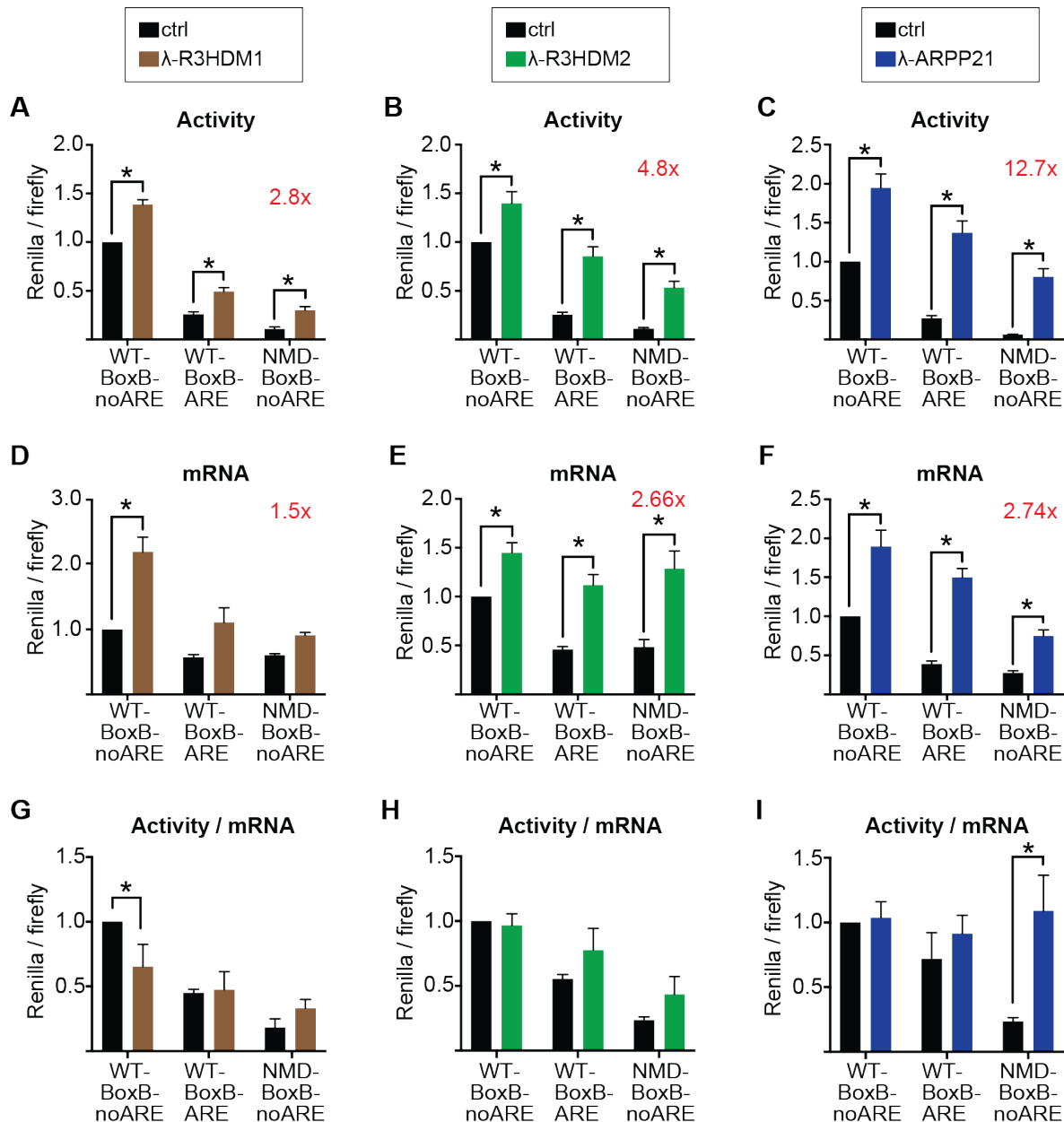
Figure S3.
(A) Western blot analysis showing the levels of lambdaN-tagged short SZRD1 in comparison to endogenous levels of SZRD1 in samples from the experiments presented in figure 4B. **(B-C)** Relative luciferase activities were measured in HEK293T cells cotransfected with the indicated reporter gene constructs ("WT-BoxB") and expression vectors for lambdaN-tagged long (B) or short (C) SZRD1 forms containing or not mutations in the SUZ and SUZ-C domains (see figure 4C and D). 'empty' indicates transfection of an empty expression vector. **(D-E)** Western blot analysis showing the expression of lambdaN-tagged SZRD1 long (D) or short (E) from the reporter gene lysis buffer presented in figure 4C and D. **(F)** Schematic representation of reporter gene constructs used in G and H. Each construct contains the *Renilla* ORF (in white) followed by a genomic β-globin sequence (in brown) and six BoxB sites ('6 BoxB') in the 3'UTR. In addition, four let7a binding sites ('let7a bs') or four mutated binding sites ('let7a mu') were inserted downstream the BoxB sites. A premature stop codon ('STOP') in the NMD constructs triggers nonsense mediated decay. **(G-H)** Luciferase activities were measured in HEK293T cells transfected with the indicated reporter gene constructs in presence (grey bars) or absence (black bars) of plasmid driving the expression of the long form of SZRD1 protein (+CAG) containing a lambdaN tag in the N-terminus (G) or the short form of SZRD1 protein (+CAG) containing a lambdaN tag in the N-terminus (H). In all panels, a firefly luciferase construct driven by the same promoter was used as a control for transfection efficiency and 'empty' indicates that an empty expression plasmid was transfected. Values are means +/-SEM of three independent experiments, where each condition was performed in triplicates, and asterisks indicate $p < 0.05$ in post-hoc testing after two-way ANOVA.
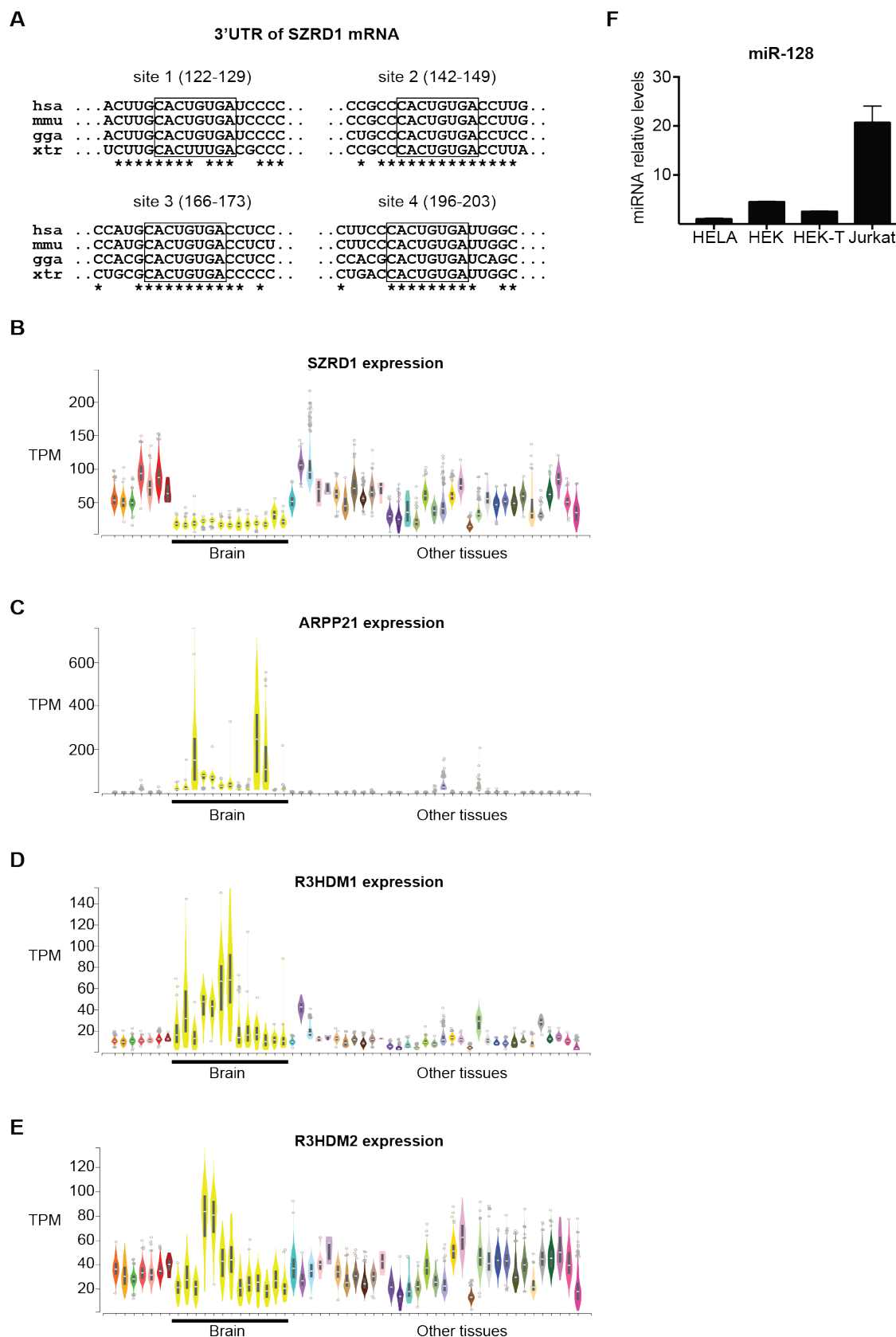
Figure S4.
(A) Protein lysates from HEK cells overexpressing the indicated SZRD1 proteins ('SZRD1') with a C-terminal SFB tag (consisting of a FLAG tag, a streptavidin binding peptide and S-peptide) or an empty vector ('control') were subjected to affinity purification using streptavidin-Sepharose beads. Input corresponds to 10% of the amount used in the pulldown. **(B)** Complement to figure 5E. Protein lysates from UPF1, eIF4A3 and RNPS1 knock-in HEK293 cells or a control cell line ('Ctrl') were subjected to affinity purification using magnetic streptavidin beads followed by western blot analysis with the indicated antibodies. Input corresponds to 10% of the amount used in the pulldown.

Figure S5.
(A-F) Quantification of UPF1 (A), SMG1 (C) and UPF2 (E) protein levels observed by western blot in figure 6F (and two additional independent experiments), and quantification of UPF1 (B), SMG1 (D), UPF2 (F) mRNA levels measured by RT-qPCR in HEK cells expressing a non-silencing control shRNA ('ctrl'), two different shRNAs ('1' and '2') targeting SZRD1 or one shRNA targeting UPF1. mRNA levels were normalized to the B2M and TBP levels, and protein levels were normalized to GAPDH levels. Values are means +/-SEM of three independent experiments, where each condition was performed in triplicates. Asterisks indicate p<0.05 in post-hoc testing after two way ANOVA. **(G-J)** Quantification of UPF1 (G), SMG1 (I) and UPF2 (J) protein levels observed by western blot shown in figure 6K (and two additional independent experiments) in U2OS cells, and quantification of UPF1 (H) mRNA by RT-qPCR in U2OS cells expressing a non-silencing control shRNA ('ctrl'), two different shRNAs ('1' and '2') targeting SZRD1 or one shRNA targeting UPF1. mRNA Expression levels were normalized to the B2-microglobulin and TBP mRNA levels, and protein levels were normalized to GAPDH levels. In all panels, values are means +/-SEM from three independent experiments, where each condition was performed in triplicates. Asterisks indicate p<0.05 in *post-hoc* testing after one-way ANOVA. **(K)** HEK293 cell lines expressing a non-silencing shRNA ('control') or shRNAs targeting only the exon 2 of SZRD1 ('#3','#4','#6') were analyzed by western blot using the indicated antibodies.

Figure S6.
(A-I) HEK293T cells were transfected with the indicated reporter gene constructs in presence (colored bars) or absence (black bars) of plasmids driving the expression of the N-terminally lambdaN-tagged R3HDM1 (A, D and G, brown bars), R3HMD2 (B, E and H, green bars) or ARPP21 (C, F and I, dark blue bars). Normalized (i.e. *Renilla*/firefly) luciferase activities (A - C) and mRNA levels (D - F) were measured as described in Figure 4F-K. Panels G, H and I represent the ratios of normalized luciferase activities relative to normalized mRNA levels. Values are means +/-SEM of three independent experiments, where each condition was performed in triplicates. Asterisks indicate p<0.05 in post-hoc testing after two-way ANOVA.

Figure S7.
(A) Alignment of the predicted miR-128 binding sites in the 3'UTR of *Homo sapiens ('hsa')*, *Mus musculus-* (*'mmu'*), *Gallus gallus ('gga')* and *Xenopus tropicalis ('xtr')* *SZRD1*. The position of the sequence predicted to interact with the miR-128 seed sequence is indicated relative to the stop codon of the human transcript and is highlighted with a box. **(B-E)** Expression levels of *SZRD1* (B), *ARPP21* (C), *R3HDM1* (D) and *R3HDM2* (E) in different human tissues obtained from The Genotype-Tissue Expression (GTEx) database (4). Expression values are presented in TPM (transcript per million). Yellow color highlights samples obtained from different brain regions. **(F)** Quantification of mature miR-128 levels by RT-qPCR in HeLa, HEK, HEK-T and Jurkat cells. miR-128 levels were normalized to U6 snRNA levels.
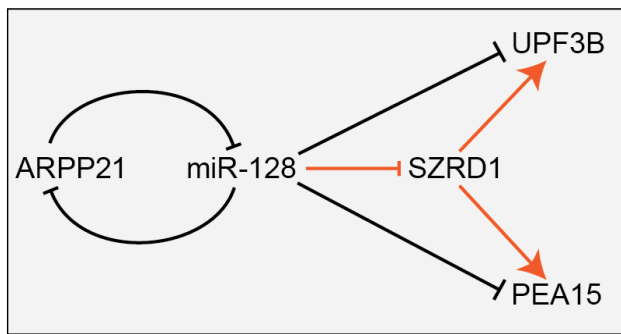
Figure S8.
Schematic representation of how downregulation of SZRD1 might be involved in the effects of miR-128.

REFERENCES

1.   Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.
2.   Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.
3.   la Cour, T., Kiemer, L., Molgaard, A., Gupta, R., Skriver, K. and Brunak, S. (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng Des Sel*, **17**, 527-536.
4.   Consortium, G.T. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet*, **45**, 580-585.