Teresa Szczepinska, Katarzyna Kalisiak, Rafal Tomecki, Anna Labno, Lukasz S. Borowski, Tomasz M. Kulinski, Dorota Adamska, Joanna Kosinska and Andrzej Dziembowski

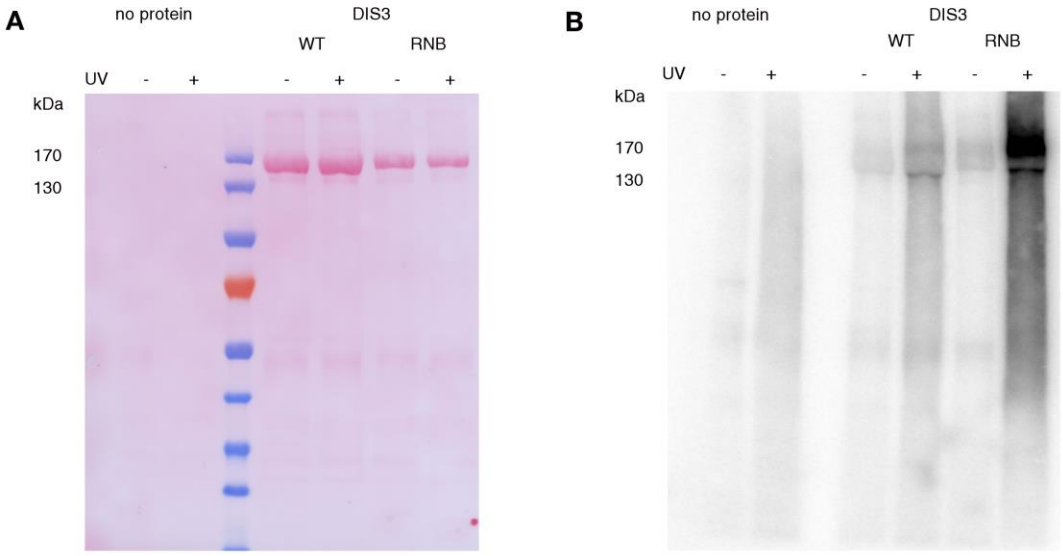
DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts

SUPPLEMENTARY DATA

Supplementary Table 1. The distribution of PAR-CLIP reads containing T to C transitions over different classes of transcripts. Reads were uniquely assigned to classes using a hierarchical procedure.

	(T-C transition)				All			
	PARCLIP 1	PARCLIP 2	**MEAN**	**RANGE**	PARCLIP 1	PARCLIP 2	**MEAN**	**RANGE**
protein coding	16.95%	17.50%	**17.23%**	**0.55%**	17.22%	19.04%	**18.13%**	**1.82%**
LINE	8.99%	8.52%	**8.75%**	**0.47%**	9.13%	9.32%	**9.22%**	**0.19%**
intron	17.30%	17.98%	**17.64%**	**0.68%**	17.42%	16.19%	**16.81%**	**1.23%**
SINE	24.09%	19.44%	**21.77%**	**4.66%**	22.59%	18.71%	**20.65%**	**3.89%**
no known feature	8.25%	10.03%	**9.14%**	**1.78%**	8.12%	8.55%	**8.34%**	**0.42%**
LTR	3.30%	3.17%	**3.23%**	**0.13%**	3.40%	3.23%	**3.31%**	**0.17%**
DNA repeat	2.01%	1.98%	**1.99%**	**0.03%**	2.23%	2.45%	**2.34%**	**0.22%**
miscRNA	0.38%	0.84%	**0.61%**	**0.46%**	0.33%	0.58%	**0.45%**	**0.25%**
snoRNA precursor	2.08%	2.00%	**2.04%**	**0.08%**	2.05%	2.02%	**2.03%**	**0.04%**
rRNA precursor	0.06%	0.06%	**0.06%**	**0.00%**	0.11%	0.08%	**0.09%**	**0.03%**
simple repeat	1.00%	0.88%	**0.94%**	**0.12%**	1.34%	1.49%	**1.41%**	**0.16%**
processed transcript	2.06%	2.24%	**2.15%**	**0.19%**	1.96%	2.19%	**2.08%**	**0.23%**
low complexity	0.81%	1.07%	**0.94%**	**0.26%**	1.03%	1.55%	**1.29%**	**0.52%**
PROMPT	7.51%	9.25%	**8.38%**	**1.74%**	7.59%	8.98%	**8.29%**	**1.38%**
lincRNA	1.21%	1.54%	**1.38%**	**0.33%**	1.93%	2.17%	**2.05%**	**0.24%**
satellite	0.05%	0.07%	**0.06%**	**0.02%**	0.06%	0.10%	**0.08%**	**0.04%**
pseudogene	0.18%	0.23%	**0.21%**	**0.06%**	0.16%	0.22%	**0.19%**	**0.06%**
snoRNA	0.18%	0.23%	**0.20%**	**0.05%**	0.17%	0.28%	**0.22%**	**0.11%**
tRNA	0.28%	0.34%	**0.31%**	**0.06%**	0.23%	0.33%	**0.28%**	**0.10%**
snRNA	0.80%	0.60%	**0.70%**	**0.20%**	0.71%	0.63%	**0.67%**	**0.09%**
other repeat	0.07%	0.09%	**0.08%**	**0.02%**	0.08%	0.09%	**0.09%**	**0.01%**
miRNA	0.08%	0.12%	**0.10%**	**0.03%**	0.10%	0.15%	**0.12%**	**0.06%**
antisense	0.16%	0.19%	**0.17%**	**0.03%**	0.14%	0.18%	**0.16%**	**0.04%**
sense overlapping	0.05%	0.05%	**0.05%**	**0.00%**	0.05%	0.04%	**0.04%**	**0.01%**
3prime overlapping ncRNA	0.01%	0.01%	**0.01%**	**0.01%**	0.01%	0.01%	**0.01%**	**0.00%**
rRNA	2.15%	1.59%	**1.87%**	**0.56%**	1.82%	1.43%	**1.63%**	**0.40%**

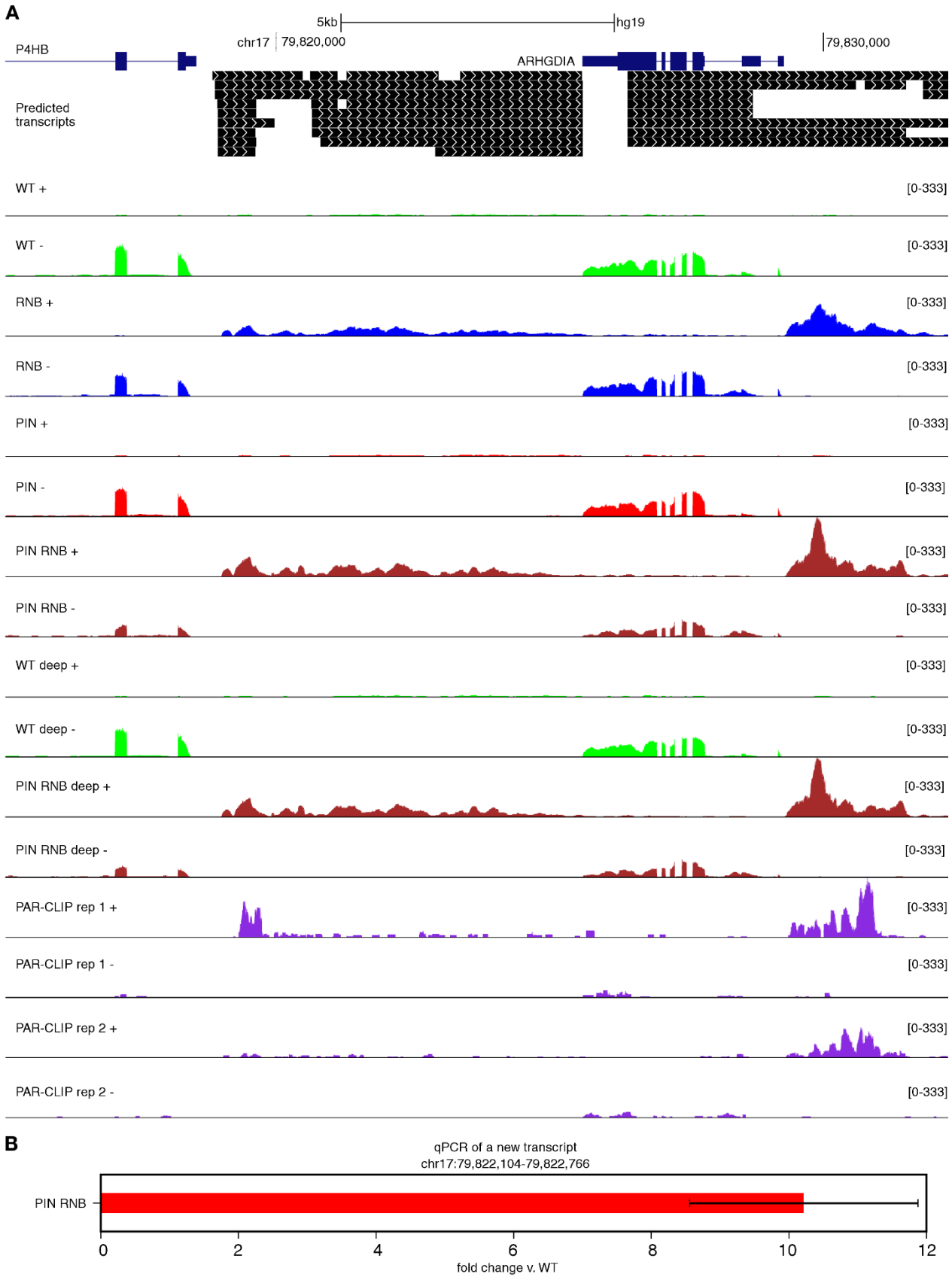
SUPPLEMENTARY FIGURES AND LEGENDS



Supplementary Figure 1. A strong radioactive signal for RNA cross-linked to DIS3 RNB mut protein variant.

(A) Ponceau S-Red staining of SDS-PAGE separated RNA-protein complexes transferred onto nitrocellulose membrane.

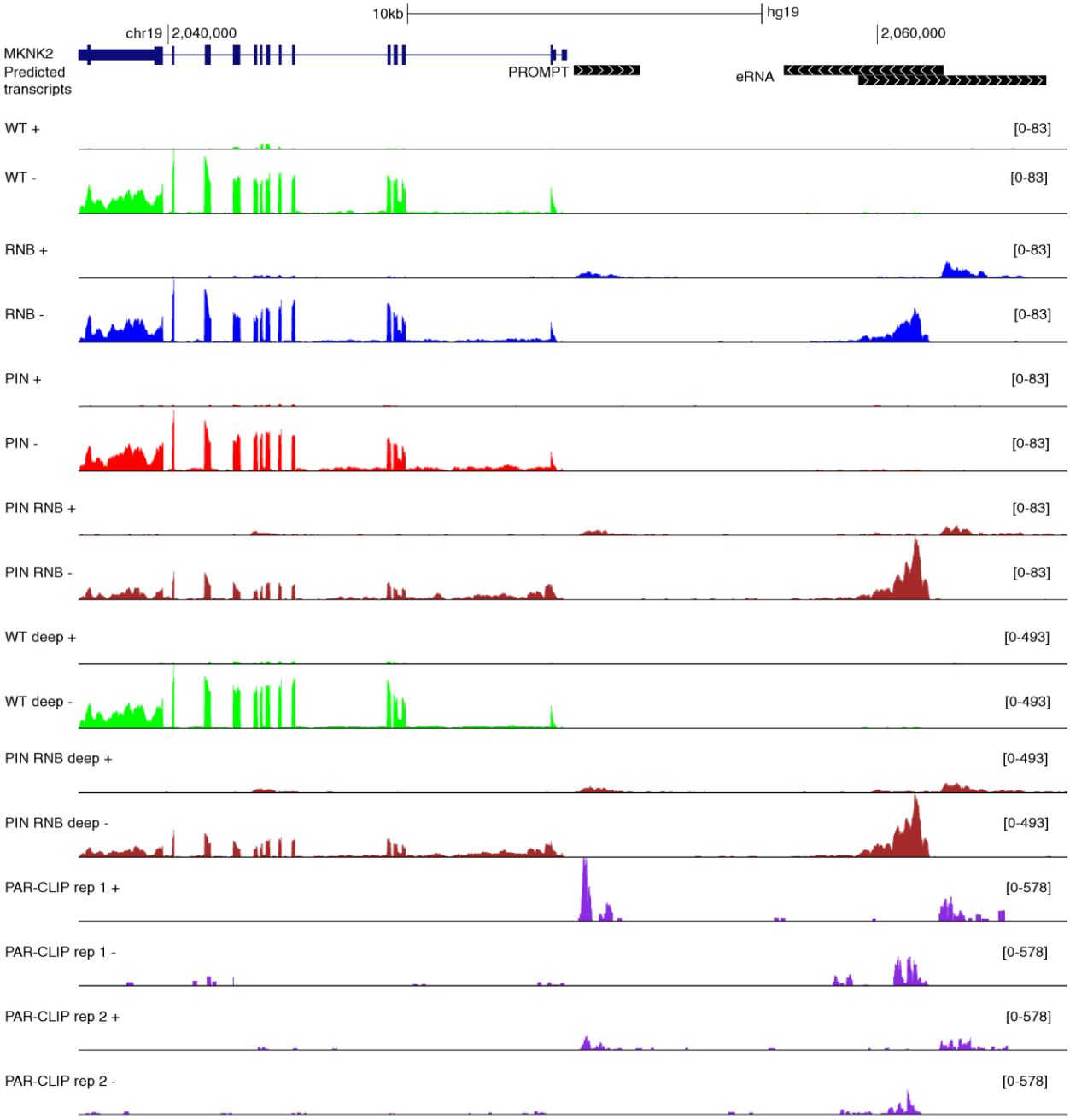
(B) An autoradiograph of the blot shown in A. Diffuse radioactivity of the protein bands results from the presence of cross-linked RNA. The strong signal in the DIS3 RNB mutant is a consequence of the large amount of RNAs.


A
5kb
hg19
chr17
79,820,000
79,830,000
P4HB
ARHGDIA
Predicted transcripts
WT +
WT -
RNB +
RNB -
PIN +
PIN -
PIN RNB +
PIN RNB -
WT deep +
WT deep -
PIN RNB deep +
PIN RNB deep -
PAR-CLIP rep 1 +
PAR-CLIP rep 1 -
PAR-CLIP rep 2 +
PAR-CLIP rep 2 -
[0-333]
B
qPCR of a new transcript
chr17:79,822,104-79,822,766
PIN RNB
0
2
4
6
8
10
12
fold change v. WT

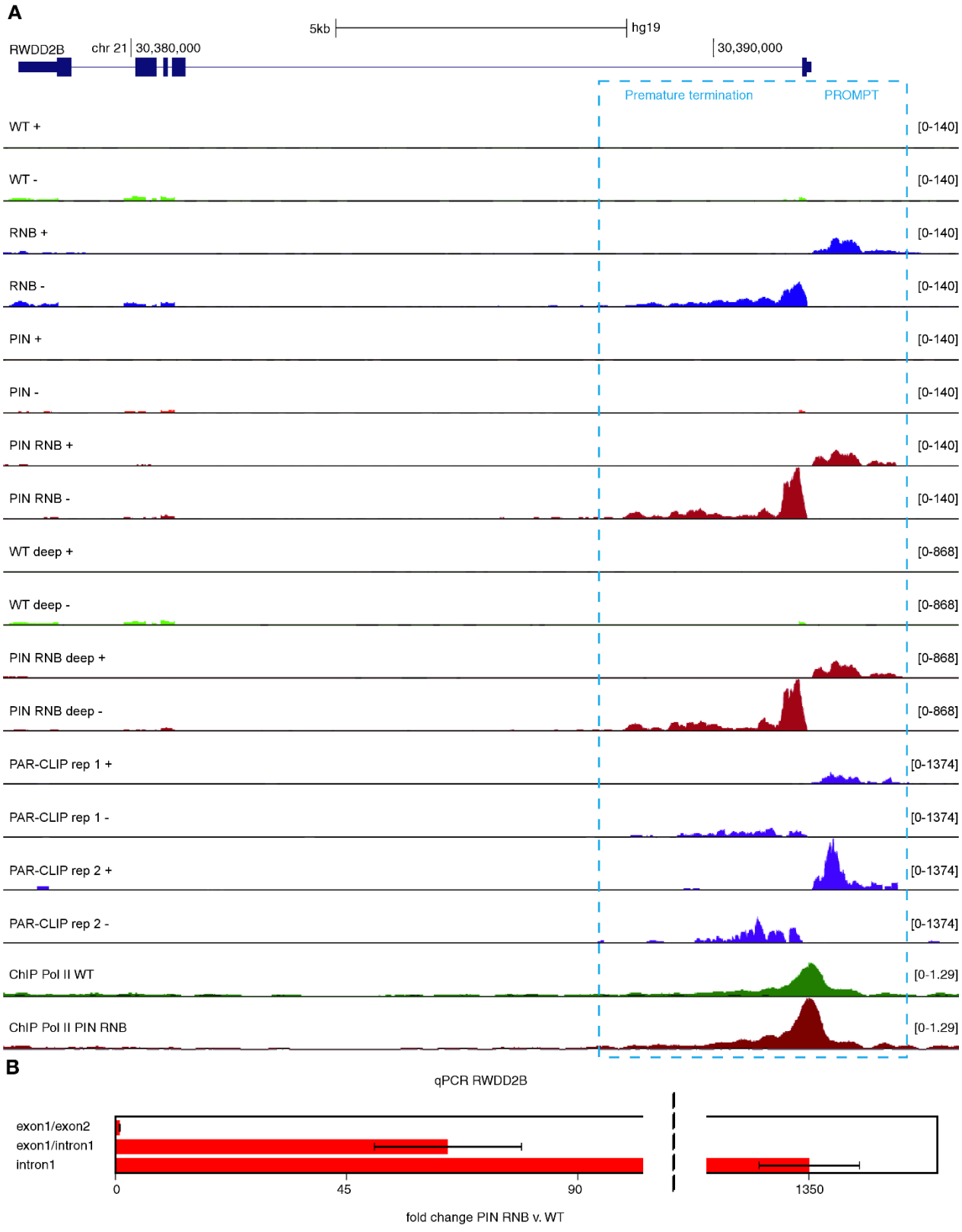
Supplementary Figure 2. An example of a previously unknown transcript.

(A) A genome browser screenshot of novel transcripts expressed in an antisense direction towards known transcripts, such as known PROMPTs. The RNA-seq signal supports our findings for the PAR-CLIP signal. RNA-seq reads mapped for WT, RNB, PIN, PIN RNB double-mutant, and DIS3 PAR-CLIP. The uniquely mapped reads are visualized separately for the plus and minus strands. Novel transcripts were predicted by Cufflinks software.

(B) Quantitative PCR validation of one transcript visualized in A. Error bar represents the standard deviation for three biological replicates.


10kb
hg19
chr19
2,040,000
2,060,000
MKNK2
Predicted
transcripts
PROMPT
eRNA
WT +
[0-83]
WT -
[0-83]
RNB +
[0-83]
RNB -
[0-83]
PIN +
[0-83]
PIN -
[0-83]
PIN RNB +
[0-83]
PIN RNB -
[0-83]
WT deep +
[0-493]
WT deep -
[0-493]
PIN RNB deep +
[0-493]
PIN RNB deep -
[0-493]
PAR-CLIP rep 1 +
[0-578]
PAR-CLIP rep 1 -
[0-578]
PAR-CLIP rep 2 +
[0-578]
PAR-CLIP rep 2 -
[0-578]

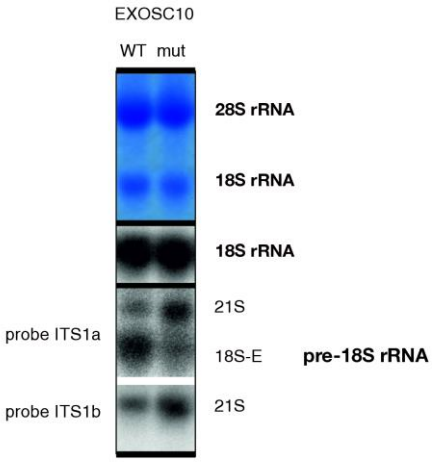
Supplementary Figure 3. A genome browser screenshot of an enhancer, PROMPT, and the MKNK2 gene, with altered expression in the PIN RNB double-mutant. RNA-seq reads mapped for WT, RNB, PIN, PIN RNB double-mutant, and DIS3 PAR-CLIP. The uniquely mapped reads are visualized separately for the plus and minus strands. Novel transcripts were predicted by Cufflinks software.


A
5kb
hg19
RWDD2B
chr 21
30,380,000
30,390,000
Premature termination
PROMPT
WT +
WT -
RNB +
RNB -
PIN +
PIN -
PIN RNB +
PIN RNB -
WT deep +
WT deep -
PIN RNB deep +
PIN RNB deep -
PAR-CLIP rep 1 +
PAR-CLIP rep 1 -
PAR-CLIP rep 2 +
PAR-CLIP rep 2 -
ChIP Pol II WT
ChIP Pol II PIN RNB
[0-140]
[0-868]
[0-1374]
[0-1.29]
B
qPCR RWDD2B
exon1/exon2
exon1/intron1
intron1
0
45
90
1350
fold change PIN RNB v. WT

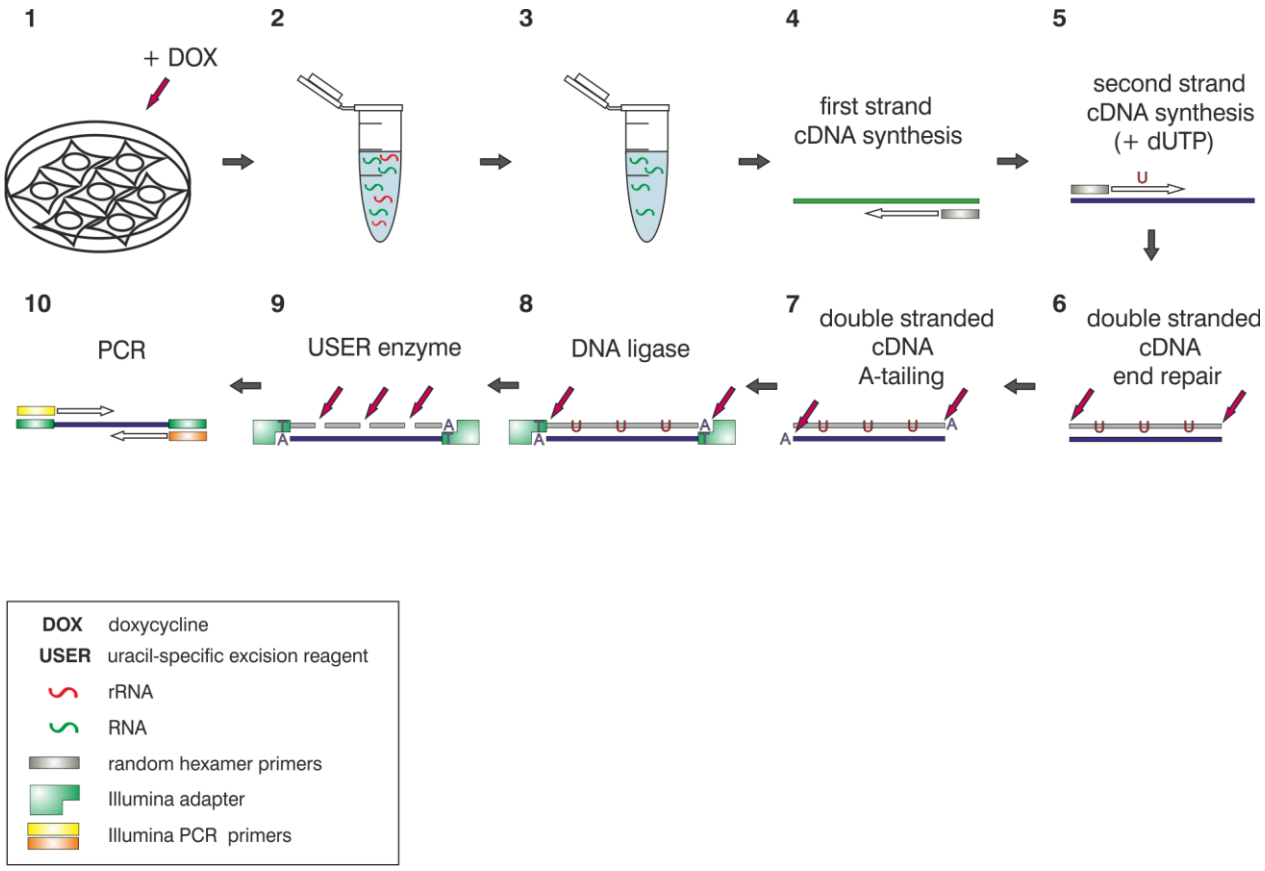
Supplementary Figure 4. The RWDD2B gene, an example of a transcript crossing the exon1/intron1 junction in DIS3 double-mutants.

(A) A genome browser screenshot. RNA-seq reads mapped for WT, RNB, PIN, PIN RNB double-mutant is supported by the PAR-CLIP and RNA polymerase II ChIP-seq signal. The uniquely mapped reads are visualized separately for the plus and minus strands. The mock control signals for Pol II ChIP are overlaid in scale on the respective tracks in grey colour.

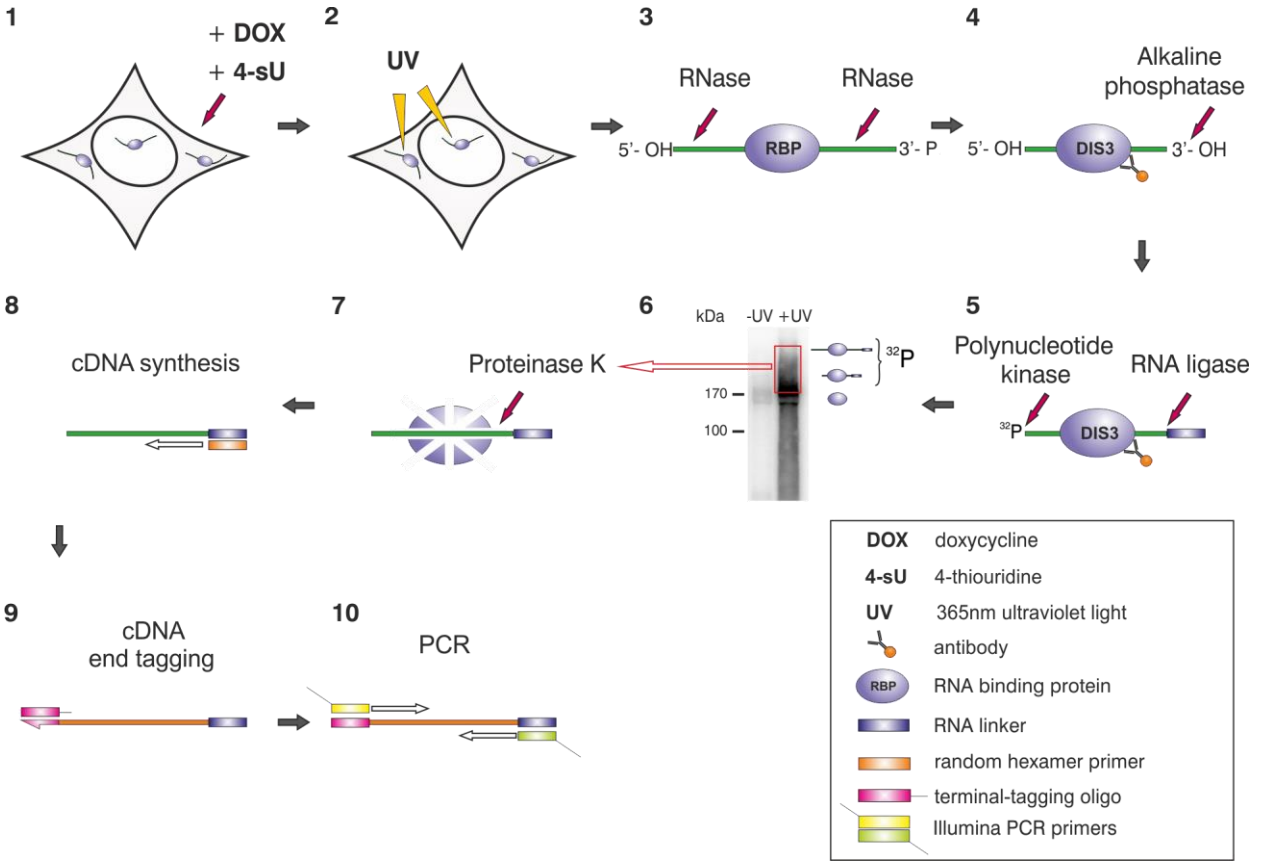
(B) Quantitative PCR validation of the increase of transcripts crossing exon1/intron1 junction. Error bars represent the range of two biological replicates.



Supplementary Figure 5. Analysis of 18S rRNA precursors in the model cell lines producing WT and mut variants of EXOSC10 (RRP6) upon sh-miRNA-mediated depletion of endogenous *EXOSC10*. Production of the EXOSC10 catalytic mutant results in the elevated levels of 21S pre-rRNA (detected by probes ITS1a and ITS1b) and concomitant decrease of the levels of shorter 18S-E precursor species (detected exclusively by probe ITS1a), which is in concordance with previous data obtained basing on siRNA silencing of *EXOSC10* gene in human cells (Sloan KE et al. 2013).



Supplementary Figure 6. The RNA-seq procedure. 1. Doxycycline induction; 2. RNA isolation; 3. rRNA removal; 4. First strand cDNA synthesis; 5. Second strand cDNA synthesis; 6. Double stranded cDNA end repair; 7. Double stranded cDNA 3′-end adenylation; 8. Adapter ligation; 9. Uracil-specific excision; 10. Amplification.



Supplementary Figure 7. The PAR-CLIP procedure. 1. Doxycycline and 4-thiouridine addition; 2. UV cross-linking; 3. RNase T1 digestion; 4. RNA 3′-end dephosphorylation; 5. Linker ligation and RNA radiolabeling; 6. SDS–PAGE and electrotransfer; 7. Proteinase K digestion and linker ligation; and 8. cDNA synthesis; 9. cDNA end tagging; 10. Amplification.

2. SUPPLEMENTARY FILE LEGENDS

Supplementary File 1. Differential expression of annotated transcripts in DIS3 mutants. A list of transcripts that were differentially expressed in at least one of the following mutants compared to WT levels: PIN, RNB, PIN RNB double-mutant (corrected *p*-value <0.05). Following columns with DESeq2 results are assigned to each comparison: 'baseMean' – average of the normalized count values, taken over mutant and wild type samples; 'log2FoldChange' – log2 fold change in the mutant comparing to wild type; 'lfcSE' – the standard error estimate for the log2 fold change estimate; 'stat' – Wald significance test statistics; 'pvalue' - significance test *p*-value; 'padj' – Benjamini-Hochberg adjusted *p*-value.

Supplementary File 2. Novel transcripts not overlapping with known transcripts identified in deeply sequenced libraries. BED file format. Transcripts were identified using Cufflinks program.

Supplementary File 3. Differential expression of annotated transcripts in deeply sequenced libraries. Transcripts that were differentially expressed in PIN RNB double-mutants compared with levels in WT (corrected *p*-value <0.05). Columns are analogous to those in Supplementary File 1.

Supplementary File 4. A list of differentially expressed enhancers along with the expression levels of neighboring genes. Low expression tag for closest transcripts means that the transcript was filter out from expression change test due to having expression below threshold. Two enhancers have two closest transcripts starting at the same position.

3. SUPPLEMENTARY METHODS

Buffers

RNA-seq alternative master mix for second strand cDNA synthesis components

0.06 × (final concentration) First-strand Buffer (Invitrogen)
0.9 × (final concentration) Second-strand Buffer (Invitrogen)
1.20 mM DTT
0.27 mM $MgCl_2$
0.2 mM dNTP Mix with dUTP
0.24 U/µl DNA Polymerase I (*E. coli*) (NEB)
0.01 U/µl RNase H (Invitrogen)
0.06 U/µl *E. coli* DNA Ligase (NEB)

RNA-seq PCR amplification mix components

1× HF Buffer
0.2 µM PP1 primer
5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA-3′
0.2 µM PP2 primer 5′-CAAGCAGAAGACGGCATACGAGAT-3′
0.2 mM dNTP mix
0.02 U/µl polymerase (Phusion; Thermo Scientific)

PAR-CLIP lysis buffer

1 mM EDTA
1 mM PMSF
50 mM Tris/HCl, pH 8.0
150 mM NaCl
0.5 M urea
0.5% (v/v) Triton X-100
Protease inhibitor mix (6 nM Leupeptin; 20 nM Pepstatin; Chymostatin 2 ng/ml)

PAR-CLIP high-salt (HS) wash buffer

1 mM PMSF
50 mM Tris-HCl, pH 8.0

1M NaCl
1M urea
0.1% (v/v) Triton X-100
Protease inhibitor mix

PAR-CLIP low-salt (LS) wash buffer

1 mM PMSF
50 mM Tris-HCl, pH 8.0
150 mM NaCl
0.1% (v/v) Triton X-100
Protease inhibitor mix

PAR-CLIP PNK buffer

1 mM PMSF
50 mM Tris-HCl, pH 7.5
50 mM NaCl
10 mM $MgCl_2$
Protease inhibitor mix

PAR-CLIP ligation buffer

50 mM Tris-HCl, pH 8.0
10 mM $MgCl_2$

PAR-CLIP PK buffer

150 mM Tris-HCl, pH 7.5
50 mM NaCl
10 mM EDTA
10 mM imidazol
5 mM β-mercaptoethanol
0.1% NP-40
1% SDS

Preparation of RNA-seq libraries (Supplementary Figure 6)

Previously described HEK293 Flp-In T-REx cell lines (Tomecki et al. 2014) were grown in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco) and penicillin/streptomycin (Sigma-Aldrich) at 37°C in a 5% CO_2 atmosphere. Cells were cultured in a petri dish (100 × 20 mm) until they reached 80% confluence. Expression of exogenous genes was induced by the addition of doxycycline at a final concentration of 100 ng/ml. RNA was isolated using TRI Reagent (Sigma-Aldrich), total RNA was treated with Turbo-DNase (Life Technologies). Then, rRNA was removed from samples using a Ribo-Zero Kit (Epicentre), according to the manufacturer's protocols with minor modifications. To control for experimental performance, external RNA (ERCC RNA Spike-In Mix; Life Technologies) was added to the samples. RNA libraries were prepared using a TruSeq RNA Sample Preparation Kit (Illumina) according to the manufacturer's manual, with the following changes: i) the RNA sequencing sample was obtained from total RNA, ii) SuperScript III Reverse Transcriptase was used instead of SuperScript II Reverse Transcriptase (Life Technologies) for first strand cDNA synthesis, iii) an alternative master mix that contained dUTP was added to the sample to synthesize the second strand, and dUTPs were incorporated into the emerging second strand to allow for the subsequent digestion of this strand and to obtain strand-specific libraries, iv) before PCR amplification of double stranded cDNA, the second strand was digested with 0.04 U/µl USER enzyme (Uracil-Specific Excision Reagent; BioLabs) for 30 min at 37°C; and v) a PCR amplification alternative mix was used.

PAR-CLIP (Supplementary Figure 7)

Stable inducible HEK293 Flp-In T-Rex cell lines were grown in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco) and penicillin/streptomycin (Sigma Aldrich) at 37°C in a 5% CO_2 atmosphere. Cells were cultured in a petri dish (145 × 20 mm) to ~80% confluence. Expression of exogenous genes was induced by the addition of doxycycline at a final concentration of 100 ng/ml. Simultaneously, 4-thiouridine (4-sU) was added to the culture medium at a final concentration of 100 µM. The 4-sU molecules incorporate into nascent RNAs allowing for subsequent photoreactive nucleoside-labeled RNA-protein cross-linking. After 24 h, cells were washed with ice-cold PBS, irradiated with 0.12 J/cm^2 of 365 nm UV light (CL-1000 UV Crosslinker; UVP), scraped off, and then centrifuged at 400×*g* for 5 min at 4°C. A total of five plates yielded 1 ml wet cell pellet, which was sufficient for one experiment.

Cell pellets were resuspended in lysis buffer (see above), sonicated, and centrifuged at 16,000×*g* for 15 min at 4°C. To obtain RNA of an appropriate size, cleared extracts were treated with 5 U/ml RNase T1 (Ambion) for 10 min at 22°C with shaking at 150 rpm, and then were cooled on ice for 5 min. Cell lysates containing RNA protein complexes were then mixed with urea at a final concentration of 0.5 M and added to 50 µl freshly washed magnetic beads coated with GFP-trap antibodies and were incubated for 1 h at 4°C with rotation. Co-immunoprecipitation was followed by three washes with high-salt wash buffer (HS buffer), one wash with low-salt wash buffer (LS buffer), and one wash with PNK buffer (see above). Beads were resuspended in 50 µl PNK buffer and treated with 35 U alkaline phosphatase (NEB) in the presence of an RNase inhibitor (Ribolock; Thermo Scientific) for 20 min at 37°C with shaking at 1100 rpm in

Thermomixer (Eppendorf). Next, the lysates were washed twice with PNK buffer and twice with ligation buffer (see above). The resulting short RNA fragments, dephosphorylated at their 3′-ends, and still cross-linked to proteins, were ligated with a pre-adenylated linker Ra3 (Illumina). The reaction was done by treating with 400 U T4 RNA Ligase 2 truncated (NEB) overnight at 16°C with shaking at 1100 rpm in 50 µl linker ligation buffer in the presence of 20% PEG 400 (Sigma) and 1 mM DTT (Sigma). This step was followed by washing once with HS buffer, twice with LS buffer, and once with PNK buffer.

RNA molecules cross-linked to immunoprecipitated proteins were then radiolabeled at their 5′-ends. Beads were resuspended in 50 µl PNK buffer with 15 U T4 polynucleotide kinase (NEB) and 5 µCi [^{32}P]-γ-ATP, and then were incubated at 37°C with shaking at 1100 rpm for 15 min. Unlabeled ATP was then added and the incubation continued for an additional 15 min. Beads with radiolabeled RNA were washed 4 times with PNK buffer, resuspended in NuPAGE LDS loading buffer (Life Technologies), and incubated for 10 min at 70°C with shaking at 1100 rpm to denature and release RNA–protein complexes. Then, DTT was added at a final concentration of 65 mM. RNA-protein complexes were resolved by size using 4–12% NuPAGE Bis-Tris gels (Life Technologies) in MOPS SDS running buffer (Invitrogen), and then transferred from the gel onto Protran Nitrocellulose membranes (Whatman) by wet-blotting (30 V, 1 h) in a Invitrogen XCell II Blot Module using NuPAGE transfer buffer (Invitrogen) according to the manufacturer's instructions.

Membranes were stained with Ponceau S-Red (Sigma-Aldrich; 0.1% in 3% acetic acid) and exposed to a phosphoscreen for 30 min. The RNA-protein complexes migrated above the expected MW of the protein, and appeared as a smeary signal (Supplementary Figure 1). The area of the membrane in which radioactivity appeared was cut into small slices and placed in 400 µl PK buffer

(see above). RNAs were released by proteinase K digestion as follows: Proteinase K (BioLine) was added to the samples at a final concentration of 500 μg/ml and was incubated at 37°C at 1100 rpm for 1 h. The solution containing RNAs was then mixed with phenol:chloroform and shaken at 1100 rpm. Phases were separated by spinning for 5 min at 16,000×*g* at room temperature. The aqueous layer was mixed with chloroform (1:1), shaken, and centrifuged again as described above. RNAs were precipitated overnight in the presence of Glycoblue (Ambion), NaAc, pH=5.2, and ethanol. Finally, RNA molecules were purified using magnetic beads (Agencourt AMPure XP; Beckman Coulter), according to the manufacturer's protocol. RNA libraries were prepared from the obtained RNA using a ScriptSeq™ v2 RNA-seq Library Preparation Kit (Epicentre), according to the manufacturer's instructions. A cDNA synthesis primer, reverse PCR primer containing a user-defined barcode, and a pre-adenylated linker Ra3 were from the TruSeq Small RNA Sample Preparation Kit (Illumina).

Quantitative PCR (qPCR) validation

RNA for qPCR was isolated from human cell lines induced with doxycycline (100 ng/ml) using the standard TRI Reagent (Sigma-Aldrich) protocol. 10 μg of total RNA was treated with TURBO^{TM} DNase (Ambion), according to manufacturer's guidelines. Following phenol: chloroform extraction and the precipitation of RNA with isopropanol, 2 μg DNase-treated RNA was used for cDNA synthesis; products of this reaction were then used for qPCR reactions, which were carried out in a Roche LightCycler® 480 system. Negative (- reverse transcriptase) controls were included for each experiment and showed a negligible background. The specificity of each reaction was verified by a melting curve analysis. Analyses were performed in triplicate with exception of the *RWDD2B* gene, which was assayed in duplicate. *GAPDH* mRNA was amplified as an internal control.

cDNA synthesis was conducted in a final volume of 20 µl in the following conditions:

50 pmol oligo(dT) primer
250 ng random hexamers (Invitrogen)
0.5 mM dNTP
40 U RiboLock™ RNase Inhibitor (Thermo Scientific)
200 U Superscript III™ reverse transcriptase (Invitrogen)

qPCR reactions were conducted in a final volume of 10 µl in the following conditions:

10-fold diluted cDNA template
5 µl Platinum® Quantitative PCR SuperMix-UDG (Invitrogen)
2.5 pmol each oligonucleotide (see below)
0.5 µg bovine serum albumin

Oligonucleotides used for qPCR validation

1. New transcript (chr21: 47873804–47878847), 5′-primer (GGGCCTGGTTCTTGTCTTTG), 3′-primer (GAAGGCAAGCAACAAACAGTTG), Figure 1B
2. New transcript (chr17: 79819010–79825596), 5′-primer (TAGGGTCCTCTCAAGGTGCC), 3′-primer (TCCTCCTCCAGCCTTACCAG), Supplementary Figure 2B
3. *TNFRSF9* gene, 5′-primer (ACGGGGCAGAAAGAAACTCC), 3′-primer (TCTGGAAATCGGCAGCTACA), Figure 3C
4. *TUT1* gene, exon–exon: 5′-primer (CTGCCACGTTACTACAGCCA), 3′-primer (CGTAGTTCTACCAGGTGCCG), exon–intron 1: 5′-primer (AAAAGACGAAAGCCCGCAAC), 3′-primer (TGCCTCTGCCACGTTACTAC), intron 1: 5′-primer (ACGGAGGTGAGAGACCCTAC), 3′-primer (TTGCGGGCTTTCGTCTTTTG), Figure 4C

5. *RWDD2B* gene, exon–exon: 5′-primer (GGGTTACAGCAGTGAGGGGG), 3′-primer (TAGCAGGTCTAACTCAGCAAGC), exon–intron 1: 5′-primer (GTTCCCAAAGGGTTGCACTG), 3′-primer (CGGGTTACAGCAGTGAGGG), intron 1: 5′-primer (AGAAGCCGGGAAGCAACG), 3′-primer (CAGTGCAACCCTTTGGGAAC), Supplementary Figure 4B

Paraspeckles analysis

HEK293 stable cell lines were plated on poly-L-lysine-coated glass cover slips 24 h prior to transfection. The expression of sh-miRNA and proteins was induced with tetracycline (50 ng/ml). The next day, cells were transiently transfected using the TransIT® LT2020 reagent (Mirus) according to the manufacturer's instructions; 450 ng DNA construct encoding Nono fused with mCherry was used per transfection. Following a 24 h incubation, cells were fixed with 3.7% formaldehyde + 5% sucrose in PBS, washed twice with PBS, stained with Hoechst 33342 (Molecular Probes) in PBS at concentration of 1 μg/ml, washed twice with PBS, and mounted with ProLong® Gold Antifade Mountant (Molecular Probes). Z-stack images were collected with a FluoView1000 confocal microscope (Olympus) using a PLANAPO 60.0 × 1.40 oil objective. Images were used for 3D rendering of paraspeckles. All measurements of paraspeckle numbers and volumes were performed using Imaris7.2.3 software (Bitplane).

ChIP analysis

Cells expressing WT and mutated variant of DIS3 were cross-linked with 1.1% formaldehyde in 5 mM Hepes pH 8.0, 10 mM NaCl, 0.1 mM EDTA for 15 min at room temperature. Cross-linking was stopped by addition of Glycine to 125 mM concentration. Cells were collected by

centrifugation at 1,000×*g* for 4 min at 4°C and washed twice with ice-cold PBS. Pellets were frozen and stored at -80°C. Nuclei were prepared by incubation of cells in 0.1% NP-40 in PBS (supplemented with 1 mM PMSF and protease inhibitors) for 10 min in ice and collected by centrifugation at 1,700×*g* for 4 min at 4°C. Chromatin was fragmented with MNase (600U per $15x10^7$ cells) in a Digestion Buffer (50 mM Tris-HCl pH 7.5, 4 mM $MgCl_2$, 2 mM $CaCl_2$, 320 mM sucrose, 1 mM PMSF and protease inhibitors) at 37°C for 15 min. Reaction was stopped by addition of EDTA to 6 mM concentration. Nuclei were collected by centrifugation at 1,700×*g* for 4 min at 4°C and washed with ice-cold PBS containing protease inhibitors. Nuclei were resuspended in Shearing Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1% SDS, 1 mM PMSF and protease inhibitors) at a concentration of $15x10^7$ per 1 ml and sonicated with S220 Focused-ultrasonicator (Covaris) at 7 Watts for 5 min at 4°C. The efficiency of chromatin shearing was analyzed by agarose gel electrophoresis. Bradford assay was performed to calculate protein concentration in the samples. Immunoprecipitation (IP) was carried out with rotation over night at 4°C in RIPA Buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS) with 25µl of Dynabeads® Protein A (LifeTechnologies) per IP and anti-RNA Polymerase II antibody (clone 8WG16, BioLegend) at 1:100 dilution. Chromatin was used at concentration of 500 µg of proteins per IP. As a mock control beads incubated only with chromatin (without antibodies) were used. Beads were washed for 5 min with rotation at 4°C twice with Wash Buffer 1 (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), twice with Wash Buffer 2 (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), and twice with TE Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Immunoprecipitated chromatin was eluted with TE Buffer supplemented with 0.5% SDS and incubated with mixing (900 rpm) at 65°C for 3 h (reverse cross-link). Chromatin was digested with

10 µg of RNase A (Fermentas) at 37 °C for 30 min, then with 20 µg of Proteinase K (Fermentas) at 55 °C with mixing (900 rpm) for 90 min. DNA was purified with a column based kit (DNA Clean & Concentrator™-25; Zymo Research; D4006) and eluted in 30 µl of H_2O.

The hierarchy of transcript classes (listed in decreasing order) used for unique assignments of sequenced reads for the classes

28S, 18S, ITS1, ITS2, 5′-ETSrDNA, 3′-ETSrDNA, 5.8S, mt_tRNA, tRNAscan, mt_rRNA, rRNA, miRNA, snRNA, snoRNA, misc_RNA, 3prime_overlapping_ncrna, lincRNA, sense_overlapping, protein_coding, IG_V_gene, TR_J_gene, TR_V_gene, processed_transcript, non_stop_decay, prompt, presnoRNA, processed_pseudogene, transcribed_processed_pseudogene, transcribed_unprocessed_pseudogene, unitary_pseudogene, polymorphic_pseudogene, pseudogene, retained_intron, nonsense_mediated_decay, sense_intronic, intron, antisense, no_feature

Northern blotting

RNA was isolated from human HEK293 Flp-In T-REx-derived stable cell lines using TRI Reagent (Sigma-Aldrich). 10 µg of total RNA was fractionated by electrophoresis either in denaturing 6% polyacrylamide-urea gel followed by electroblotting onto a Hybond N^+ membrane in 0.5 x TBE at 4^{o}C, or in a 1% formaldehyde-agarose gel [prepared using NBC buffer (50 mM boric acid, 1 mM sodium acetate, 5 mM NaOH)], followed by RNA immobilization on the same type of membrane by overnight capillary transfer in 20 x SSC (3 M NaCl, 0.3 M sodium citrate). RNA was fixed on membranes by UV cross-linking. Hybridizations were performed in PerfectHyb Plus hybridization buffer (Sigma-Aldrich). The blots were handled according to standard procedures and probed at

42°C. Between successive hybridizations, probes were stripped off the membranes at 65°C using boiling 0.1% SDS.

The following 5'-^{32}P-labeled oligonucleotides were used as probes:

SNORD13 (5'-GGTCAGACGGGTAATGTGCCCACGTCGTAA-3')

SNORD83 (5'-GCTGTTCTCAGAAGGAAGGCA-3')

SNORD118 (5'-AATCAGACAGGAGCAATCAGGGTGTTGCAA-3')

5.8S rRNA (5'-TCCTGCAATTCACATTAATTCTCGCAGCTAGC-3')

5S rRNA (5'-CATCCAAGTACTAACCAGGCCC-3')

18S rRNA (5'-TTTACTTCCTCTAGATAGTCAAGTTCGACC-3')

ITS1a (5'-CCTCGCCCTCCGGGCTCCGTTAATGATC-3')

ITS1b (5'-AGGGGTCTTTAAACCTCCGCGCCGGAACGCGCTAGGTAC-3').

After hybridization, membranes were washed with 2 x SSC, 0.1% SDS and eventually exposed to a PhosphorImager screen (FujiFilm), which was scanned following exposure using a FLA7000 scanner (FujiFilm).

Construction of the vectors for the production of exogenous EXOSC10 WT and mutant variants in the background of endogenous *EXOSC10* depletion

A multistep cloning procedure for generation of vectors for co-expression of recoded EXOSC10 with FLAG epitope at the C-terminus [wild type (WT) variant or its catalytic mutant counterpart (mut)] and sh-miRNAs directed against endogenous EXOSC10 mRNA was the following.

First, we performed a search for 3 miRNA sequences that should specifically and efficiently target endogenous EXOSC10 mRNA, using BLOCK-iT™ RNAi Designer tool from Invitrogen

(with "*miR RNAi"* option) (http://rnaidesigner.invitrogen.com/rnaiexpress). We chose candidate sequences starting at positions 165, 549 and 742 of *EXOSC10* ORF. Basing on the general idea of BLOCK-iT™ Pol II miR RNAi Expression Vector Kits from Invitrogen (http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/rnai/Vector-based-RNAi/Pol-II-miR-RNAi-Vectors.html), we then designed a synthetic DNA fragment, encompassing combination of miRNA sequences listed above. It encoded three tandemly positioned shRNAs corresponding to pre-designed miRNAs – so-called sh-miRs, where sense and antisense miRNA sequences were separated by the loop element enabling formation of the hairpin (Schematic 1). Each of the sh-miR sequences was flanked at both termini with motifs ensuring correct miRNA processing from the artificial pre-miRNA precursor, following a natural miRNA biogenesis pathway active in human cells (Schematic 1). In addition, polyadenylation signal derived from the gene encoding herpes simplex virus thymidylate kinase (HSV-TK-PA) was placed at the 3'-end of this synthetic cassette, allowing for correct termination of transcription in human cells (Schematic 1). The cassette contained *Eco*RI/*Sal*I and *Bsu*15I/*Hin*dIII restriction site combinations at the 5' and 3' extremities, respectively, which were used in subsequent cloning steps (Schematic 1). It was synthesized by BlueHeronBio company (http://blueheronbio.com) and inserted between *Eco*RI and *Hin*dIII sites of pUCAmpMinusMCS vector. Next, a sequence encoding eGFP (allowing for monitoring of expression of the cassette containing artificial pre-miRNA) was amplified in PCR using eGFPFor-eGFPRev primer pair and pEGFP-N1 plasmid (Clontech) as a template and inserted into *Eco*RI and *Sal*I sites of the provided [pUCAmpMinusMCS] *tri-miR* plasmid, thus giving [pUCAmpMinusMCS] *eGFP-tri-miR* transitory construct. Additionally, a site recognized by *Xma*JI restriction endonuclease was

introduced in eGFP or oligonucleotide upstream the 5'-end of *eGFP* ORF, which was used at further cloning stage.

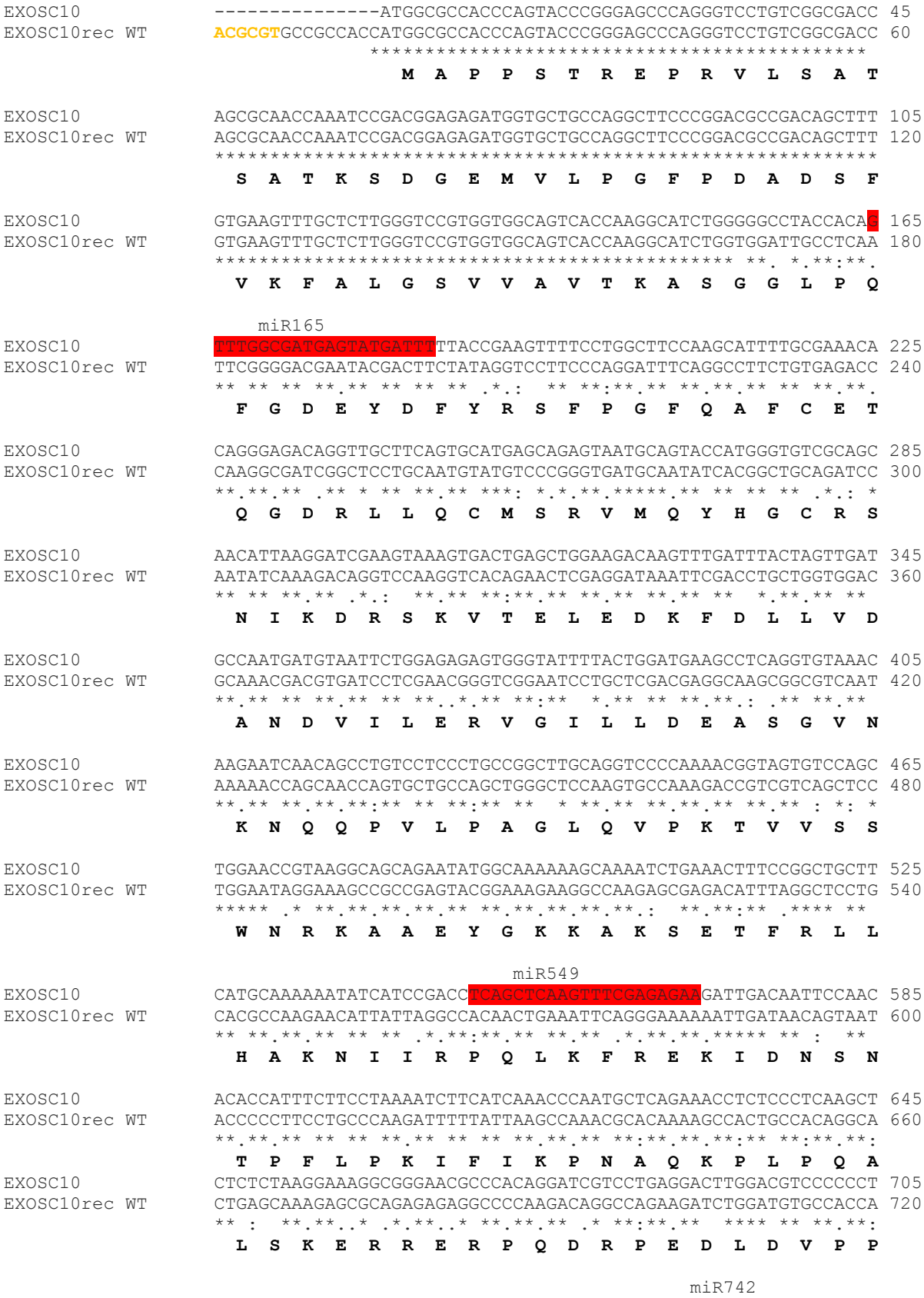
In the next phase of construct generation, it was necessary to modify the sequence of exogenous *EXOSC10* ORF in order to make it insensitive to miRNA action. To this end, we ordered synthesis of recoded *EXOSC10* open reading frame (see Schematic 2 for nucleotide sequence alignment of original and recoded *EXOSC10* ORF fragment). The idea of recoding was to introduce synonymous mutations into all possible codons (at those positions where degeneration of genetic code could be utilized and taking codon usage frequency into account), within the fragment containing sites recognized by miRNA, so that the sequence would be as much divergent from the initial one as possible (Schematic 2). We ordered WT variant of recoded ORF, introducing additionally: restriction site (*Mlu*I) and Kozak sequence at the 5'-end; and sequence coding for FLAG-tag, encompassing STOP codon and restriction site (*Apa*I) at 3'-end; recoded ORF was provided by BlueHeronBio as insert cloned into pUCAmpMinusMCS. Subsequently, it was transferred into *Mlu*I and *Apa*I sites of BI-16 vector, thus replacing *RLUC* ORF present therein, with the use of *E. coli* MH1 strain. This way, [BI-16'] EXOSC10*rec WT* transitory vector was constructed. The presence of recoded fragment was checked by digestion with *Nde*I restriction enzyme and sequencing.

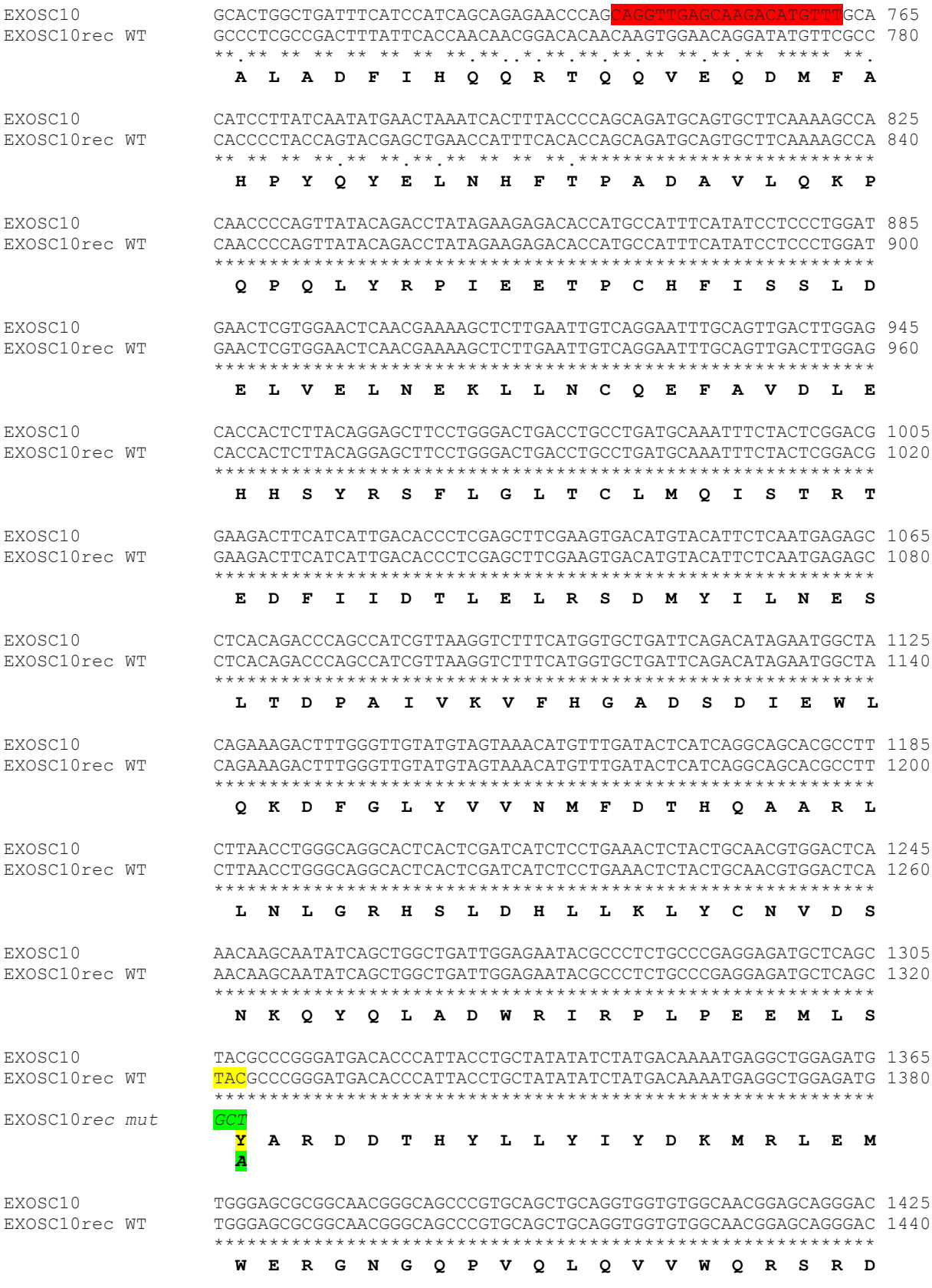
The aim of the ultimate cloning stage was to transfer a DNA fragment containing a co-cistron of eGFP coding sequence and pre-miRNA/HSV-TK-PA from [pUCAmpMinusMCS] *eGFP-tri-miR* construct to the BI-16 vector derivative from the previous step, through replacement of the *FLUC* ORF present in the latter. To this end, all plasmids were propagated in *E. coli* dam-/dcm- strain prior to the standard cloning procedure, utilizing *Xma*JI and *Bsu15*I restriction sites, followed by transformation of the ligation products into *E. coli* MH1 strain. This eventually led to

the generation of final construct: pEXOSC10-WT. Both *EXOSC10* and *eGFP-tri-miR* inserts were sequenced. Finally, construct encoding catalytic mutant counterpart of WT EXOSC10 was prepared (pEXOSC10 -mut) by *in vitro* mutagenesis. Presence of catalytic mutation was confirmed by digestion with *Afe*I and sequencing.

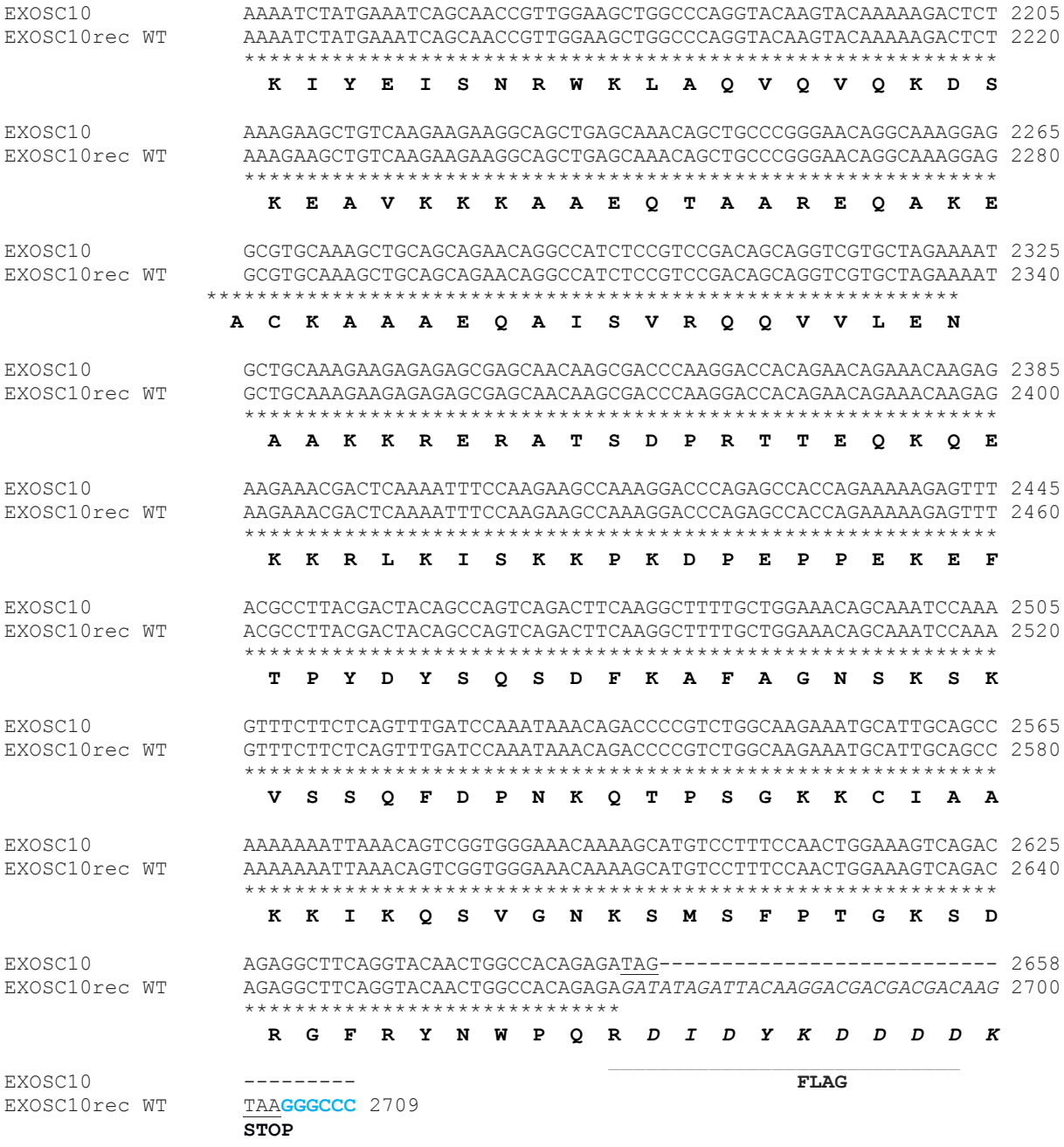
GAATTCATATAGTCGACCAGTGGATCCTGGAGGCTTGCTGAAGGCTGTATGCTGAAATCATACTCATCGCCAAACGTTTTGGC
CACTGACTGACGTTTGGCGGAGTATGATTTCAGGACACAAGGCCTGTTACTAGCACTCACATGGAACAAATGGCCCAGATCCT
GGAGGCTTGCTGAAGGCTGTATGCTGTTCTCTCGAAACTTGAGCTGAGTTTTGGCCACTGACTGACTCAGCTCATTTCGAGAG
AACAGGACACAAGGCCTGTTACTAGCACTCACATGGAACAAATGGCCCAGATCCTGGAGGCTTGCTGAAGGCTGTATGCTGAA
ACATGTCTTGCTCAACCTGGTTTTGGCCACTGACTGACCAGGTTGAAAGACATGTTTCAGGACACAAGGCCTGTTACTAGCAC
TCACATGGAACAAATGGCCCAGATCTGGCCGCACTCGAGATATCTAGTGATCTAGAGGGCCCGCGGTTCGCTGATGGGGGAGG
CTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGGTG
TTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGTGACCCCATTGGGGCCAA
TACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGC
AGGCCCTGCCATAGCATCGATCGCAAGCTT

Schematic 1. Sequence of the *tri-miR* insert. Red, blue, green and violet letters indicate *Eco*RI, *Sal*I, *Bsu*15I and *Hin*dIII restriction sites, utilized in the cloning procedure; grey and black background indicate 5' and 3' miR flanking regions, respectively; red backgrounds correspond to 21 nt-long antisense target sequences (mature miRNA sequences), beginning in positions 165, 549 and 742 of *EXOSC10* ORF (from 5' to 3'); violet backgrounds correspond to nucleotides 1-8 and 11-21 of the respective sense target sequences; green backgrounds represent a 19 nt-long sequence derived from endogenous murine miR-155, with underlined 13 nt-long segment able to form a loop within sh-miRNA structure; yellow background corresponds to HSV TK polyadenylation signal.


EXOSC10 ----------ATGGCGCCACCCAGTACCCGGGAGCCCAGGGTCCTGTCGGCGACC 45
EXOSC10rec WT ACGCGTGCCGCCACCATGGCGCCACCCAGTACCCGGGAGCCCAGGGTCCTGTCGGCGACC 60
M A P P S T R E P R V L S A T
EXOSC10 AGCGCAACCAAATCCGACGGAGAGATGGTGCTGCCAGGCTTCCCGGACGCCGACAGCTTT 105
EXOSC10rec WT AGCGCAACCAAATCCGACGGAGAGATGGTGCTGCCAGGCTTCCCGGACGCCGACAGCTTT 120
S A T K S D G E M V L P G F P D A D S F
EXOSC10 GTGAAGTTTGCTCTTGGGTCCGTGGTGGCAGTCACCAAGGCATCTGGGGGCCTACCACAG 165
EXOSC10rec WT GTGAAGTTTGCTCTTGGGTCCGTGGTGGCAGTCACCAAGGCATCTGGTGGATTGCCTCAA 180
V K F A L G S V V A V T K A S G G L P Q
miR165
EXOSC10 TTTGGCGATGAGTATGATTTTTACCGAAGTTTTCCTGGCTTCCAAGCATTTTGCGAAACA 225
EXOSC10rec WT TTCGGGGACGAATACGACTTCTATAGGTCCTTCCCAGGATTTCAGGCCTTCTGTGAGACC 240
F G D E Y D F Y R S F P G F Q A F C E T
EXOSC10 CAGGGAGACAGGTTGCTTCAGTGCATGAGCAGAGTAATGCAGTACCATGGGTGTCGCAGC 285
EXOSC10rec WT CAAGGCGATCGGCTCCTGCAATGTATGTCCCGGGTGATGCAATATCACGGCTGCAGATCC 300
Q G D R L L Q C M S R V M Q Y H G C R S
EXOSC10 AACATTAAGGATCGAAGTAAAGTGACTGAGCTGGAAGACAAGTTTGATTTACTAGTTGAT 345
EXOSC10rec WT AATATCAAAGACAGGTCCAAGGTCACAGAACTCGAGGATAAATTCGACCTGCTGGTGGAC 360
N I K D R S K V T E L E D K F D L L V D
EXOSC10 GCCAATGATGTAATTCTGGAGAGAGTGGGTATTTTACTGGATGAAGCCTCAGGTGTAAAC 405
EXOSC10rec WT GCAAACGACGTGATCCTCGAACGGGTCGGAATCCTGCTCGACGAGGCAAGCGGCGTCAAT 420
A N D V I L E R V G I L L D E A S G V N
EXOSC10 AAGAATCAACAGCCTGTCCTCCCTGCCGGCTTGCAGGTCCCCAAAACGGTAGTGTCCAGC 465
EXOSC10rec WT AAAAACCAGCAACCAGTGCTGCCAGCTGGGCTCCAAGTGCCAAAGACCGTCGTCAGCTCC 480
K N Q Q P V L P A G L Q V P K T V V S S
EXOSC10 TGGAACCGTAAGGCAGCAGAATATGGCAAAAAAGCAAAATCTGAAACTTTCCGGCTGCTT 525
EXOSC10rec WT TGGAATAGGAAAGCCGCCGAGTACGGAAAGAAGGCCAAGAGCGAGACATTTAGGCTCCTG 540
W N R K A A E Y G K K A K S E T F R L L
miR549
EXOSC10 CATGCAAAAAATATCATCCGACCTCAGCTCAAGTTTCGAGAGAAGATTGACAATTCCAAC 585
EXOSC10rec WT CACGCCAAGAACATTATTAGGCCACAACTGAAATTCAGGGAAAAAATTGATAACAGTAAT 600
H A K N I I R P Q L K F R E K I D N S N
EXOSC10 ACACCATTTCTTCCTAAAATCTTCATCAAACCCAATGCTCAGAAACCTCTCCCTCAAGCT 645
EXOSC10rec WT ACCCCCTTCCTGCCCAAGATTTTTATTAAGCCAAACGCACAAAAGCCACTGCCACAGGCA 660
T P F L P K I F I K P N A Q K P L P Q A
EXOSC10 CTCTCTAAGGAAAGGCGGGAACGCCCACAGGATCGTCCTGAGGACTTGGACGTCCCCCCT 705
EXOSC10rec WT CTGAGCAAAGAGCGCAGAGAGAGGCCCCAAGACAGGCCAGAAGATCTGGATGTGCCACCA 720
L S K E R R E R P Q D R P E D L D V P P
miR742


EXOSC10 GCACTGGCTGATTTCATCCATCAGCAGAGAACCCAGCAGGTTGAGCAAGACATGTTTGCA 765
EXOSC10rec WT GCCCTCGCCGACTTTATTCACCAACAACGGACACAACAAGTGGAACAGGATATGTTCGCC 780
A L A D F I H Q Q R T Q Q V E Q D M F A
EXOSC10 CATCCTTATCAATATGAACTAAATCACTTTACCCCAGCAGATGCAGTGCTTCAAAAGCCA 825
EXOSC10rec WT CACCCCTACCAGTACGAGCTGAACCATTTCACACCAGCAGATGCAGTGCTTCAAAAGCCA 840
H P Y Q Y E L N H F T P A D A V L Q K P
EXOSC10 CAACCCCAGTTATACAGACCTATAGAAGAGACACCATGCCATTTCATATCCTCCCTGGAT 885
EXOSC10rec WT CAACCCCAGTTATACAGACCTATAGAAGAGACACCATGCCATTTCATATCCTCCCTGGAT 900
Q P Q L Y R P I E E T P C H F I S S L D
EXOSC10 GAACTCGTGGAACTCAACGAAAAGCTCTTGAATTGTCAGGAATTTGCAGTTGACTTGGAG 945
EXOSC10rec WT GAACTCGTGGAACTCAACGAAAAGCTCTTGAATTGTCAGGAATTTGCAGTTGACTTGGAG 960
E L V E L N E K L L N C Q E F A V D L E
EXOSC10 CACCACTCTTACAGGAGCTTCCTGGGACTGACCTGCCTGATGCAAATTTCTACTCGGACG 1005
EXOSC10rec WT CACCACTCTTACAGGAGCTTCCTGGGACTGACCTGCCTGATGCAAATTTCTACTCGGACG 1020
H H S Y R S F L G L T C L M Q I S T R T
EXOSC10 GAAGACTTCATCATTGACACCCTCGAGCTTCGAAGTGACATGTACATTCTCAATGAGAGC 1065
EXOSC10rec WT GAAGACTTCATCATTGACACCCTCGAGCTTCGAAGTGACATGTACATTCTCAATGAGAGC 1080
E D F I I D T L E L R S D M Y I L N E S
EXOSC10 CTCACAGACCCAGCCATCGTTAAGGTCTTTCATGGTGCTGATTCAGACATAGAATGGCTA 1125
EXOSC10rec WT CTCACAGACCCAGCCATCGTTAAGGTCTTTCATGGTGCTGATTCAGACATAGAATGGCTA 1140
L T D P A I V K V F H G A D S D I E W L
EXOSC10 CAGAAAGACTTTGGGTTGTATGTAGTAAACATGTTTGATACTCATCAGGCAGCACGCCTT 1185
EXOSC10rec WT CAGAAAGACTTTGGGTTGTATGTAGTAAACATGTTTGATACTCATCAGGCAGCACGCCTT 1200
Q K D F G L Y V V N M F D T H Q A A R L
EXOSC10 CTTAACCTGGGCAGGCACTCACTCGATCATCTCCTGAAACTCTACTGCAACGTGGACTCA 1245
EXOSC10rec WT CTTAACCTGGGCAGGCACTCACTCGATCATCTCCTGAAACTCTACTGCAACGTGGACTCA 1260
L N L G R H S L D H L L K L Y C N V D S
EXOSC10 AACAAGCAATATCAGCTGGCTGATTGGAGAATACGCCCTCTGCCCGAGGAGATGCTCAGC 1305
EXOSC10rec WT AACAAGCAATATCAGCTGGCTGATTGGAGAATACGCCCTCTGCCCGAGGAGATGCTCAGC 1320
N K Q Y Q L A D W R I R P L P E E M L S
EXOSC10 TACGCCCGGGATGACACCCATTACCTGCTATATATCTATGACAAAATGAGGCTGGAGATG 1365
EXOSC10rec WT TACGCCCGGGATGACACCCATTACCTGCTATATATCTATGACAAAATGAGGCTGGAGATG 1380
EXOSC10rec mut GCT
Y A R D D T H Y L L Y I Y D K M R L E M
A
EXOSC10 TGGGAGCGCGGCAACGGGCAGCCCGTGCAGCTGCAGGTGGTGTGGCAACGGAGCAGGGAC 1425
EXOSC10rec WT TGGGAGCGCGGCAACGGGCAGCCCGTGCAGCTGCAGGTGGTGTGGCAACGGAGCAGGGAC 1440
W E R G N G Q P V Q L Q V V W Q R S R D

```
EXOSC10          ATCTGCCTCAAGAAATTCATCAAACCTATCTTCACGGATGAGTCCTACCTTGAACTCTAT 1485
EXOSC10rec WT    ATCTGCCTCAAGAAATTCATCAAACCTATCTTCACGGATGAGTCCTACCTTGAACTCTAT 1500
                 ************************************************************
                   I  C  L  K  K  F  I  K  P  I  F  T  D  E  S  Y  L  E  L  Y

EXOSC10          AGGAAGCAGAAGAAGCACCTTAACACACAGCAGTTGACAGCCTTTCAGCTGCTGTTTGCC 1545
EXOSC10rec WT    AGGAAGCAGAAGAAGCACCTTAACACACAGCAGTTGACAGCCTTTCAGCTGCTGTTTGCC 1560
                 ************************************************************
                   R  K  Q  K  K  H  L  N  T  Q  Q  L  T  A  F  Q  L  L  F  A

EXOSC10          TGGAGGGATAAAACAGCTCGCAGGGAAGATGAAAGTTACGGATATGTACTGCCAAACCAC 1605
EXOSC10rec WT    TGGAGGGATAAAACAGCTCGCAGGGAAGATGAAAGTTACGGATATGTACTGCCAAACCAC 1620
                 ************************************************************
                   W  R  D  K  T  A  R  R  E  D  E  S  Y  G  Y  V  L  P  N  H

EXOSC10          ATGATGCTGAAAATAGCTGAAGAACTGCCTAAGGAACCTCAGGGCATCATAGCTTGCTGC 1665
EXOSC10rec WT    ATGATGCTGAAAATAGCTGAAGAACTGCCTAAGGAACCTCAGGGCATCATAGCTTGCTGC 1680
                 ************************************************************
                   M  M  L  K  I  A  E  E  L  P  K  E  P  Q  G  I  I  A  C  C

EXOSC10          AACCCAGTACCGCCCCTTGTGCGGCAGCAGATCAACGAAATGCACCTTTTAATCCAGCAG 1725
EXOSC10rec WT    AACCCAGTACCGCCCCTTGTGCGGCAGCAGATCAACGAAATGCACCTTTTAATCCAGCAG 1740
                 ************************************************************
                   N  P  V  P  P  L  V  R  Q  Q  I  N  E  M  H  L  L  I  Q  Q

EXOSC10          GCCCGAGAGATGCCCCTGCTCAAGTCTGAAGTTGCAGCCGGAGTGAAGAAGAGCGGACCG 1785
EXOSC10rec WT    GCCCGAGAGATGCCCCTGCTCAAGTCTGAAGTTGCAGCCGGAGTGAAGAAGAGCGGACCG 1800
                 ************************************************************
                   A  R  E  M  P  L  L  K  S  E  V  A  A  G  V  K  K  S  G  P

EXOSC10          CTGCCCAGTGCTGAGAGATTGGAGAATGTTCTCTTTGGACCTCACGACTGCTCCCATGCC 1845
EXOSC10rec WT    CTGCCCAGTGCTGAGAGATTGGAGAATGTTCTCTTTGGACCTCACGACTGCTCCCATGCC 1860
                 ************************************************************
                   L  P  S  A  E  R  L  E  N  V  L  F  G  P  H  D  C  S  H  A

EXOSC10          CCTCCGGATGGCTATCCAATCATCCCAACCAGTGGATCTGTGCCAGTTCAGAAGCAGGCG 1905
EXOSC10rec WT    CCTCCGGATGGCTATCCAATCATCCCAACCAGTGGATCTGTGCCAGTTCAGAAGCAGGCG 1920
                 ************************************************************
                   P  P  D  G  Y  P  I  I  P  T  S  G  S  V  P  V  Q  K  Q  A

EXOSC10          AGCCTCTTCCCTGATGAAAAAGAAGATAACTTGCTGGGTACCACATGCCTGATTGCCACA 1965
EXOSC10rec WT    AGCCTCTTCCCTGATGAAAAAGAAGATAACTTGCTGGGTACCACATGCCTGATTGCCACA 1980
                 ************************************************************
                   S  L  F  P  D  E  K  E  D  N  L  L  G  T  T  C  L  I  A  T

EXOSC10          GCTGTCATCACGTTATTTAATGAACCTAGTGCTGAAGACAGTAAAAAGGGTCCATTGACA 2025
EXOSC10rec WT    GCTGTCATCACGTTATTTAATGAACCTAGTGCTGAAGACAGTAAAAAGGGTCCATTGACA 2040
                 ************************************************************
                   A  V  I  T  L  F  N  E  P  S  A  E  D  S  K  K  G  P  L  T

EXOSC10          GTTGCACAGAAAAAAGCCCAGAACATCATGGAGTCCTTTGAAAATCCATTTAGGATGTTT 2085
EXOSC10rec WT    GTTGCACAGAAAAAAGCCCAGAACATCATGGAGTCCTTTGAAAATCCATTTAGGATGTTT 2100
                 ************************************************************
                   V  A  Q  K  K  A  Q  N  I  M  E  S  F  E  N  P  F  R  M  F

EXOSC10          CTGCCCTCACTGGGACACCGTGCTCCCGTCTCTCAGGCAGCGAAGTTCGATCCATCAACC 2145
EXOSC10rec WT    CTGCCCTCACTGGGACACCGTGCTCCCGTCTCTCAGGCAGCGAAGTTCGATCCATCAACC 2160
                 ************************************************************
                   L  P  S  L  G  H  R  A  P  V  S  Q  A  A  K  F  D  P  S  T
```



Schematic 2. Alignment of the native (EXOSC10) and recoded (EXOSC10rec) *EXOSC10* ORF. Nucleotides unchanged during recoding are marked with asterisks; orange and blue letters indicate *Mlu*I and *Apa*I restriction sites, respectively; red backgrounds highlight positions of sequences targeted by pre-designed sh-miRs; green backgrounds indicate nucleotide (and resulting amino acid) difference in the EXOSC10rec mutant variant (*EXOSC10rec mut*); sequence coding for the

FLAG epitope, present at the 3’-end of EXOSC10rec insert is marked with italics; positions of translation termination codons are underlined.