

Supplementary material for

**Identification of factors involved in target RNA-directed microRNA
degradation**

Gabrielle Haas[#], Semih Cetin[#], Mélanie Messmer, Béatrice Chane-Woon-Ming, Olivier Terenzi, Johana Chicher, Lauriane Kuhn, Philippe Hammann and Sébastien Pfeffer*

Correspondence: s.pfeffer@ibmc-cnrs.unistra.fr

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Characterization of m169-miR-27 features triggering TDMD, related to Figure 1.

(A) Representation of m169 pairing with miR-27. Three mismatches were introduced in the seed region and are represented in red. (B) HeLa cells were transfected with increasing amounts of F-Luc, F-Luc-m169 or F-Luc-6xSH reporters and together with a plasmid coding for the *Renilla* luciferase (R-Luc) as transfection control, and subjected to luciferase assay. The F-Luc values were normalized to those of *Renilla* and arbitrarily set at 100 in cells expressing the F-Luc reporter. The graph represents mean values +/- standard deviations from three independent experiments. (C) Description of the miRNA isoforms. miR-27 and miR-16 serve as examples. The major isoforms (red) correspond to the isomiRs always detectable in absence of specific antimiRNAs. When TDMD is induced by transfection of antimiRNAs, tailed isoforms correspond to the bands of higher sizes than the major isoforms (purple) and the trimmed isomiRs to the shorter bands (blue).

Figure S2. Identification of the cellular factors involved in TDMD, related to Figure 2.

(A-C) Proteins identified by mass spectrometry using antimiR-27 (A), m169-like (B) and antimiR-16 (C) are represented in Venn diagrams including proteins found with the control-oligoribonucleotide or bound to the beads alone. For each section, the corresponding number of proteins validated with FDR<1% and the percentage (in parentheses) are mentioned.

Figure S3. Identification of the cellular factors involved in TDMD, related to Figure 3F.

AntimiR-27 and antimiR-16 were tested for their ability to pull-down several known exonucleases or the Terminal-Uridylyl-transferase (TUT)-7 by western blot. AntimiRNAs were transfected at 10 nM. A dashed line indicates discontinuous lanes but Input and IP fractions were systematically analyzed on a same gel.

Figure S4. TDMD is induced in the RISC complex by recruitment of TUT1 and DIS3L2, related to Figure 4.

(A) AGO2-associated tailed and trimmed isoforms of miR-27 were analyzed by northern blot. The Ctrl IP (with IgG) serves as negative control. A ladder made of 4 RNA oligos with sizes

of 16, 18, 21 and 24 nucleotides in length was loaded to evaluate the sizes of miR-27 isoforms.

(B) AGO2-associated tailed and trimmed isoforms of miR-16 by RNA-immunoprecipitation. (C) AGO2-associated tailed and trimmed isoforms of miR-27 after induction of TDMD by the F-Luc-3xSH reporter. The 3xSH m169mut version serves as negative control for tailing-trimming. Asterisks highlight tailed isoforms. (D) Comparative immunoprecipitation of miR-27 isoforms bound to DIS3L2mut and DIS3L2. MBP serves as negative control. A dashed line indicates discontinuous lanes but Input and IP fractions were systematically analyzed on a same gel (B, C). (E) Immunoprecipitation (anti-GFP) of miR-27 from cells expressing GFP-TUT2, GFP-TUT4 or GFP-TUT7 in presence or absence of anti-miR-27. tRNA is used as loading control.

Figure S5. Effect of TUT1 and DIS3L2 knockdowns on TDMD, related to Figure 5.

(A) The efficiency of miR-27 TDMD in cells depleted from TUT1 (siTUT1), DIS3L2 (siDIS3L2) or both was quantified by northern blot analysis. A representative experiment is depicted on the left, while the right panel depicts the quantification of miR-27 accumulation after normalization to tRNA. It was arbitrarily set at 100 in absence of anti-miR-27 for each condition. Error bars indicate SDs (n = 3 independent experiments). (B) Left: Analysis of miR-16 TDMD by northern blot as described in (A). Right: the graph represents miR-16 levels from three independent experiments. miR-16 accumulation is represented normalized to tRNA and set at 100 in absence of anti-miR-16 for each condition. Error bars indicate SDs. (C) Distribution in percentage of miR-27a tailed and trimmed isoforms as assessed by small RNA deep-sequencing of cells treated with the indicated siRNA and anti-miRNA oligonucleotides. The “0” indicates the wild-type (WT) isoform of 21 nt. The sum of all trimmed (≤ -2) or tailed ($\geq +2$) sequences is indicated.

Figure S6. Functional implication of DIS3L2 in miRNA degradation, related to Figure 5.

(A) Representative examples of the most abundant miR-27b isoforms detected in one of the libraries treated with the control siRNA and anti-miR-67 or anti-miR-27. Tailed nucleotides are highlighted in purple. The relative abundance of each isoform (%) is given before (anti-miR-67) or after (anti-miR-27) TDMD induction by the transfection of the corresponding anti-miRNAs. The red color indicates a decrease and the green an increase in abundance of the respective sequence. (B) and (C) Distribution in percentage of miR-27b tailed and trimmed

isoforms as assessed by small RNAs deep-sequencing of cells treated with the indicated siRNA and antimRNA oligonucleotides. The “0” indicates the wild-type (WT) isoform of 21 nt. The sum of all trimmed (≤ -2) or tailed ($\geq +2$) sequences is indicated.

Figure S7. DIS3L2 is involved in TDMD induced during MCMV infection, related to Figure 6.

(A) Representative northern blot of miR-27 levels in Hepa 1.6 cells overexpressing HA-mDIS3L2, HA-mDIS3L2mut or an empty plasmid (pcDNA) during a time course of MCMV infection. Samples are the same as in Figure 6D and 6E. tRNA serves as loading control. Time points are indicated in hours post-infection (hpi).

Table S1. Short list of the most frequent proteins associated with anti-miRNAs and identified by nanoLC-MS/MS, related to Figure 2.

This table summarizes the proteins found in at least 3 independent experiments for anti-miR-27 (5 experiments in total), anti-miR-16 (4 experiments) and the m169-like oligoribonucleotides (4 experiments) and absent in all their respective negative controls (beads alone and anti-miR-67). For each protein, both their gene name and accession number are mentioned.

* An exception is made for DIS3L2 in the m169 samples as only 3 of the 4 experiments allowed to trigger TDMD and the protein was found twice in these 3 samples.

Gene names	Accession number	anti-miR-27	anti-miR-16	m169-like
AGO1	Q9UL18	<i>yes</i>	<i>no</i>	<i>yes</i>
AGO2	Q9UKV8	<i>yes</i>	<i>yes</i>	<i>yes</i>
AGO3	Q9H9G7	<i>no</i>	<i>no</i>	<i>yes</i>
DIS3L2	Q8IYB7	<i>yes</i>	<i>yes</i>	<i>yes*</i>
FUBP3	Q96I24	<i>yes</i>	<i>yes</i>	<i>yes</i>
SSB	P05455	<i>yes</i>	<i>no</i>	<i>no</i>
MEPCE	Q7L2J0	<i>yes</i>	<i>no</i>	<i>no</i>
MYH10	P35580	<i>yes</i>	<i>no</i>	<i>no</i>
NONO	Q15233	<i>yes</i>	<i>no</i>	<i>no</i>
RBM4	Q9BWF3	<i>yes</i>	<i>no</i>	<i>yes</i>
SART3	Q15020	<i>yes</i>	<i>no</i>	<i>no</i>
SND1	Q7KZF4	<i>no</i>	<i>no</i>	<i>yes</i>
TUT1	Q9H6E5	<i>yes</i>	<i>no</i>	<i>no</i>
ZFP36L2	P47974	<i>no</i>	<i>no</i>	<i>yes</i>
TNRC6B	Q9UPQ9	<i>no</i>	<i>no</i>	<i>yes</i>
XRN2	Q9H0D6	<i>no</i>	<i>no</i>	<i>yes</i>
U2AF1	Q01081	<i>yes</i>	<i>no</i>	<i>no</i>
U2AF2	P26368	<i>yes</i>	<i>no</i>	<i>no</i>

Table S2. List of the oligonucleotides used in this study.

Nucleotides generating a mutation or modifying a pairing are in bold; Locked Nucleic Acids are underlined; 2'O-methyl nucleotides are preceded by a "m". F, forward; R, reverse.

Cloning primers (provided by Sigma-Aldrich)

Name	Type	Sequence (5' to 3')
HA-TUT1-F	DNA	GGAATTCAATGGCGGCGGTGGATTCCGGATG
HA-TUT1-R	DNA	GAAAAAAGCGGCCGCTCACTTGAGATGTCTGAATTGCTTG
HA-DIS3L2-F	DNA	CGGGATCCATGAGCCATCCTGACTACAGAATG
HA-DIS3L2-R	DNA	GAAAAAAGCGGCCGCTCAGCTGGTGCTTGAGTCCTCG
HA-AGO2-F	DNA	GGAATTCAATGTACTCGGGAGCCGGCC
HA-AGO2-R	DNA	GAAAAAAGCGGCCGCTCAAGCAAAGTACATGGTGCG
GFP-TUT1-F	DNA	GAAGATCTTAATGGCGGCGGTGGATTCCGGATG
GFP-TUT1-R	DNA	GGAATTCTCACTTGAGATGTCTGAATTGCTTG
GFP-DIS3L2-F	DNA	GAAGATCTTAATGAGCCATCCTGACTACAGAATG
GFP-DIS3L2-R	DNA	CGGGATCCTCAGCTGGTGCTTGAGTCCTCG
HA-mTUT1-F	DNA	GGAATTCAATGGCGGCGGTGGATTCCGG
HA-mTUT1-R	DNA	GAAAAAAGCGGCCGCTCACTTGAGGAGATTTTAAAGTG
HA-mDIS3L2-F	DNA	GGAATTCAATGAACCATCCTGACTACAAGC
HA-mDIS3L2-R	DNA	GAAAAAAGCGGCCGCTCAGTCCTCAGGCTCCTCATC
GFP-mTUT1-F	DNA	GAAGATCTTAATGGCGGCGGTGGATTCCGG
GFP-mTUT1-R	DNA	GGAATTCTCACTTGAGGAGATTTTAAAGTG
GFP-mDIS3L2-F	DNA	GGAATTCATGAACCATCCTGACTACAAGC
GFP-mDIS3L2-R	DNA	CGGGGTACCTCAGTCCTCAGGCTCCTCATC
HA-mAGO2-F	DNA	GGAATTCAATGTACTCGGGAGCCGGCCC
HA-mAGO2-R	DNA	GAAAAAAGCGGCCGCTCAAGCAAAGTACATGGTGCG
GFP-TUT2-F	DNA	GGAATTCATGTTCCCAAACCTCAATTTTGG
GFP-TUT2-R	DNA	CGGGATCCTTATCTTTTCAGGACAGCAGC
GFP-TUT4-F	DNA	GCGTCGACCATGGAAGAGTCTAAAACCTTAAAAAG
GFP-TUT4-R	DNA	CCATTTAAATTTACTCCGACACGTTTCCTCTTG

GFP-TUT7-F	DNA	GCTCGAGAATGGGAGATACAGCAAACC
GFP-TUT7-R	DNA	CGGGATCCTCATGATTCTGCTGGGTCC

Mutagenesis primers (provided by Sigma-Aldrich)

Name	Type	Sequence (5' to 3')
DIS3L2mut-F	DNA	ACCCATCAACCGCCGAGACCTCAATAATGCCCTCTCCTGCAAGCCAC
DIS3L2mut-R	DNA	GTGGCTTGCAGGAGAGGGCATTATTGAGGTCTCGGGCGGTTGATGGGT
mDIS3L2mut-F	DNA	GTAATAGCTTTGATGTTTCATGGCTGTGCCCTCGCCCTCTTCTTGGACATGGGTGATATGG
mDIS3L2mut-R	DNA	CCATATCACCCATGTCCAAGAAGAGGGCGAGGGCACAGCCATGAACATCAAAGCTATTTAC
pcDNA3.1-HA-F	DNA	GTTTAAACTTAAGCTTGGTACCGAGGCCACCATGTACCCATACGATGTTCCAGATTACGCTCTCGGATCCACTAGTCCAGTGTG
pcDNA3.1-HA-R	DNA	CACACTGGACTAGTGGATCCGAGAGCGTAATCTGGAACATCGTATGGGTACATGGTGGCCTCGGTACCAAGCTTAAGTTTAAAC

Northern probes (provided by Eurogentec)

Name	Type	Sequence (5' to 3')
miR-27	DNA (LNA)	GCGGA <u>ACTTAGCCACTGTGAA</u>
miR-16	DNA (LNA)	CGCCA <u>ATATTACGTGCTGCTA</u>
tRNA Asp	DNA	CCGGTCTCCCGCGTGACAGCGGGGATACTA
U6	DNA	GCAGGGGCCATGCTAATCTTCTCTGTATCG

Biotinylated oligoribonucleotides (provided by IDT)

Name	Type	Sequence (5' to 3')
ASmir27-space3Bio	RNA	mGmCmGmGmAmAmCmUmUmAmGmCmCmAmCmUmGmUmGmA mA/iSp9/3Bio
ASmir67Ce-space3Bio	RNA	mUmCmAmCmAmAmCmCmUmCmCmUmAmGmAmAmAmGmAmG mUmAmGmA/iSp9/3Bio/
m169-space3Bio	RNA	mGmCmGmGmAmAmUmAmAmUmAmAmGmCmUmGmUmGmA /iSp9/3Bio/
ASmir16-space3Bio	RNA	mCmGmCmCmAmAmUmAmUmUmUmAmCmGmUmGmCmUmGmC mUmA/iSp9/3Bio/

AntimiRNAs (provided by IDT)

Name	Type	Sequence (5' to 3')
antimiR-27	RNA	mGmCmGmGmAmAmCmUmUmAmGmCmCmAmCmUmGmUmGmA mA
antimiR-16	RNA	mCmGmCmCmAmAmUmAmUmUmUmAmCmGmUmGmCmUmGmC mUmA
antimiR-67Ce	RNA	mUmCmAmCmAmAmCmCmUmCmCmUmAmGmAmAmAmGmAmG mUmAmGmA
m169-like	RNA	mGmCmGmGmAmAmUmAmAmUmAmAmGmCmUmGmUmGmA A
Mutant 1	RNA	m UmCmGmGmAmAmUmAmAmUmAmAmGmCmUmGmUmGmA A

Mutant 2	RNA	mUmAmGmGmAmAmUmAmAmUmAmAmGmCmUmGmUmGmAm A
Mutant 3	RNA	mUmAmUmGmAmAmUmAmAmUmAmAmGmCmUmGmUmGmAm A
Mutant 4	RNA	mUmAmUmUmAmAmUmAmAmUmAmAmGmCmUmGmUmGmAm A
Bulge +	RNA	mGmCmGmGmAmAm UmUmAmUmAmAmGmCmUmGmUmGmAm A

siRNA sequences (provided by Sigma-Aldrich)

Name	Type	Sequence (5' to 3') sense strand
siRLuc (siCtrl)	siRNA	CACAUCGAGCCCGUGGCUA
siTUT1-1	siRNA	GUGUGUUUGUCAGUGGCUU
siDIS3L2-1	siRNA	GGGGAUCUGGUGGUCGUGAA

qRT-PCR primers (provided by Sigma-Aldrich)

Name	Type	Sequence (5' to 3')
miR-27	DNA	TTCACAGTGGCTAAGTTCCGC
miR-24	DNA	TGGCTCAGTTCAGCAGGAACAG
PPIA-F	DNA	GCGGCAGGTCCATCTACG
PPIA-R	DNA	GCCATCCAGCCATTCAGTC
m169-F	DNA	ATCTTCTTCGGCGTTAGCGA
m169-R	DNA	TGAGGTCCAGGTCGTGTGA

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Mass spectrometry procedure and analysis

In-solution digestion was carried out directly on the magnetic beads. Straight before digestion, the beads were washed twice in 25 mM ammonium bicarbonate using a magnetic stand to retain the beads. Out of the stand, proteins were reduced with 5 mM DTT for 6 min at 95°C, alkylated with 10 mM iodoacetamide for 30 min in the dark and digested overnight with 100 ng of modified sequencing-grade trypsin in 25 mM ammonium bicarbonate.

Dried tryptic peptides were re-suspended in 15 μ L of water containing 0.1% FA (solvent A) before analysis on a NanoLC-2DPlus system (with nanoFlex ChiP module; Eksigent, ABSciex, Concord, Ontario, Canada) coupled to a TripleTOF 5600 mass spectrometer (ABSciex) operating in high-sensitivity positive mode. Sample acquisitions were performed using a "trap and elute" configuration on the nanoFlex System using C-18 precolumn (300 μ m ID x 5 mm ChromXP; Eksigent) and column (75 μ m ID x 15 cm ChromXP; Eksigent) Peptides were eluted with a 60 min 5%-40% gradient of solvent B (0.1% formic acid in ACN). The TripleTOF 5600 was operated with Analyst software (version 1.5, ABSciex) in data-dependent acquisition mode with survey MS scans acquired in the 400-1250 m/z range and up to 20 of the most intense multiply charged ions (2+ to 5+) selected for CID fragmentation. To prevent carry-over due to stationary phase memory, 2 consecutive washing runs were performed after each sample injection. Data were searched against the complete Human proteome set from the UniProt database (released 2013/01/09; 141168 sequences) with an added decoy database. Mascot algorithm (version 2.2, Matrix Science, London, UK) was used through the ProteinScape package (Bruker, v3.1) which validated protein identifications with a FDR < 1%.

Small RNA libraries and sequencing

RNA quality was assessed by Agilent 2100 Bioanalyzer (Agilent), and samples with an RNA integrity number (a measure of RNA quality) higher than 8 were used for the study. We followed Illumina's protocol (Truseq small RNA, # 15004197 Rev. C) to generate small RNA libraries directly from total RNAs, suitable for subsequent high throughput sequencing. The protocol takes advantage of the natural structure common to most microRNA molecules that have a 3' hydroxyl group resulting from enzymatic cleavage by Dicer or other RNA processing enzymes. Briefly, in the first step, RNA adapters were sequentially ligated to each

end of the RNA, first the 3' RNA adapter that is specifically modified to target microRNAs and other small RNAs, then the 5' RNA adapter. Small RNA ligated with 3' and 5' adapters were reverse transcribed and PCR amplified (30 sec at 98°C; [10 sec at 98°C, 30 sec at 60°C, 15 sec at 72°C] x 13 cycles; 10 min at 72°C) to create cDNA constructs. This process selectively enriched those fragments that have adapter molecules on both ends. The last step was an acrylamide gel purification of the 140-150 nt amplified cDNA constructs (corresponding to cDNA inserts from small RNA + 120 nt from the adapters). Small RNA libraries were checked for quality and quantified using 2100 Bioanalyzer (Agilent). The libraries were loaded in the flowcell at 8 pM concentration and clusters were generated using the Cbot and sequenced on Hiseq 2500 as single-end 50 bases reads following Illumina's instructions.

Bioinformatics analysis of deep sequencing data

Short sequences generated by the Illumina instrument were first preprocessed using the Dustmasker program (1) and FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit) to filter out low complexity reads and remove instances of the 3' adapter, respectively. Before further analysis, reads corresponding to the exogenous siRNA and anti-miR sequences were also excluded. Remaining reads of at least 15 nt in length were then mapped to the human genome (assembly version hg19 – UCSC repository), using Bowtie 1.0.0 (2), by permitting up to 1 mismatch in the first 15 nucleotides of each read. Only alignments from the lowest mismatch-stratum were recorded provided they didn't exceed a total number of 9 mismatches and reads that could map to more than 50 loci were discarded. From there, all known human miRNAs (miRBase v.20 (3)) were annotated using BEDTools 2.16.2 (4) by comparing their genomic coordinates to those of the aligned reads, and by keeping reads with at least 80 % of their length inside the genomic feature. By doing so, we were able to inventory and quantify all miRNA isoforms. During the quantification process, multiple mapped reads were weighted by the number of mapping sites in miRNAs and the annotations of miR-27a-3p and miR-27b-3p were manually curated to redistribute the few sequences that were initially mis-attributed between the two (most probably due to the loose constraints applied on the 3' read ends during the mapping step).

To further characterize the TDMD mechanism of miR-27a and b, only isoforms corresponding to the wild-type (WT) mature sequence (as defined in miRBase) of these miRNAs, as well as the ones presenting with shifted 3' extremities were taken into account,

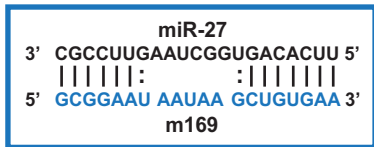
providing that they showed a 5' genome matched component and an optional 3' tail (addition of non-templated nucleotides). These sequences were classified as: WT, trimming, tailing, trimming+tailing, cleavageshift, and cleavageshift+tailing. For the representation of histograms shown in Figure 5C and S5C, we determined the number of sequences presenting with a tail of +1, ..., +4 nucleotides as well as the sequences presenting with a trimming of -1, -2, ..., -6 nucleotides. We then calculated the percentage of each category. This allowed us to see that the predominant isoforms in all libraries were the WT (i.e. 0), and the -1, and +1 sequences. We therefore took into account for our further calculation only sequences that were tailed or trimmed by more than 2 nucleotides. We then computed a tailing ratio by dividing all tailed sequences (tailing, trimming+tailing, cleavageshift+tailing) by the number of WT sequences, and a trimming ratio by dividing all trimmed sequences (trimming and trimming+tailing) by the number of WT sequences for each isoform. We then calculated the fold change of the tailing ratio of cells transfected with the anti-miR-27 to the one of cells transfected with the control anti-miR-67. This fold change was then normalized to the control siRNA (against R-Luc) condition, which was set to 1. The same was done for the trimming ratio.

Supplementary references

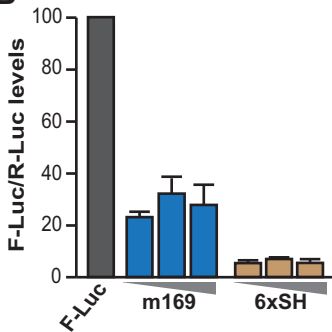
1. Morgulis,A., Gertz,E.M., Schäffer,A.A. and Agarwala,R. (2006) A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol. J. Comput. Mol. Cell Biol.*, 13, 1028–1040.
2. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, 10, R25.
3. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, 42, D68–73.
4. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.*, 26, 841–842.

Figure S1

A



B



C

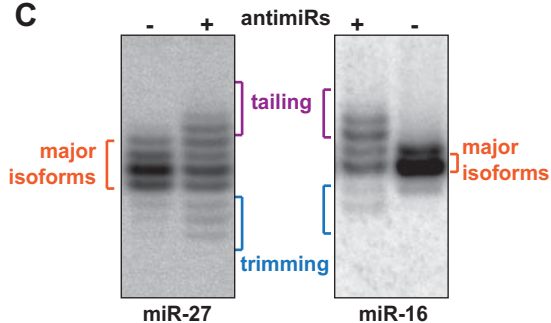
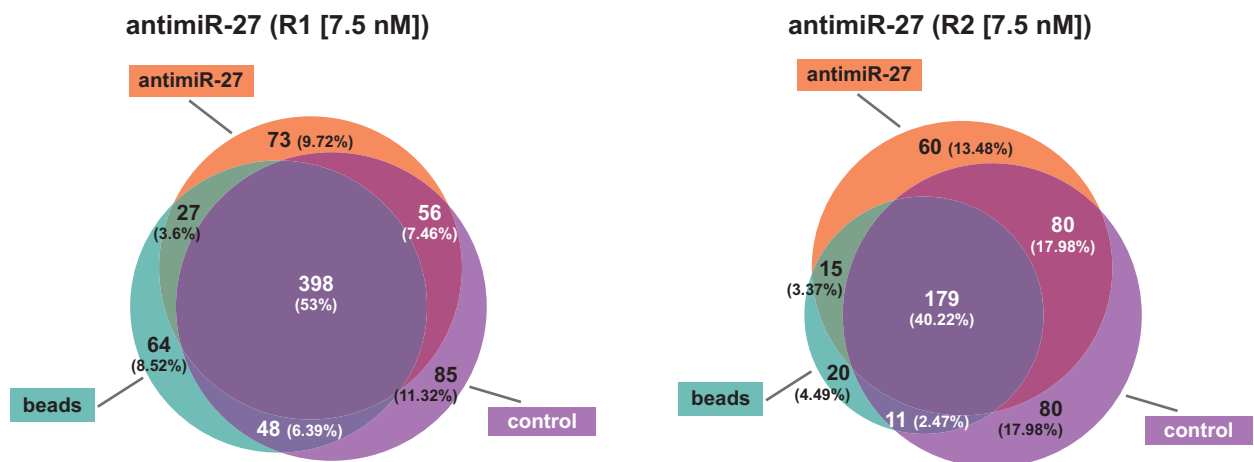
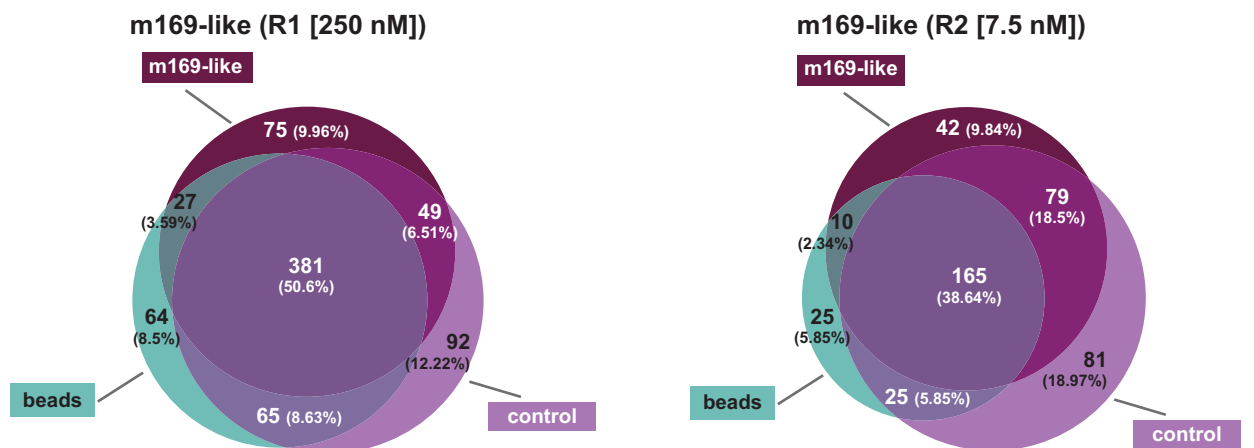


Figure S2

A



B



C

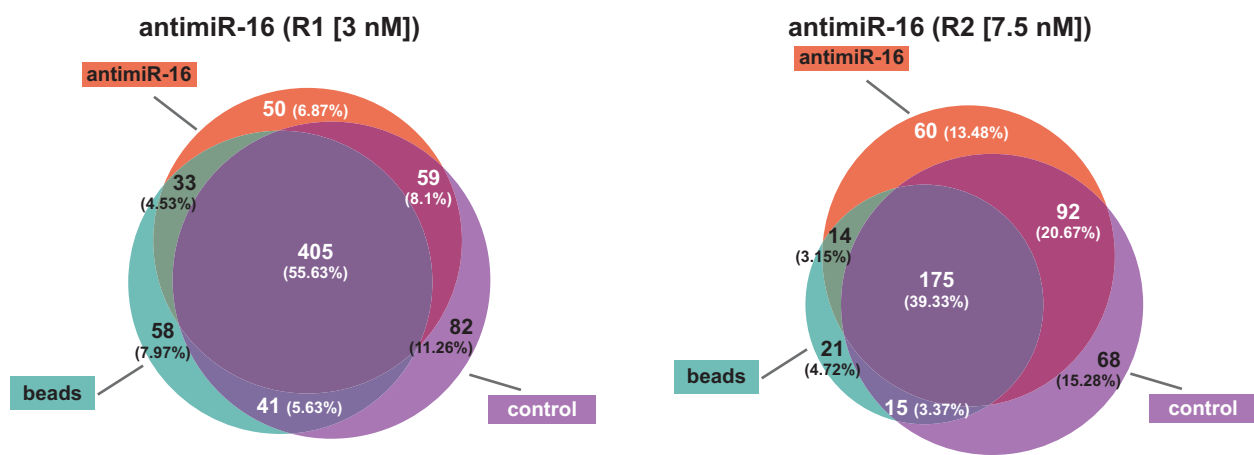


Figure S3

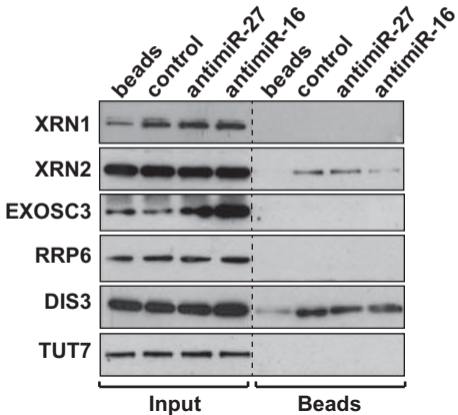


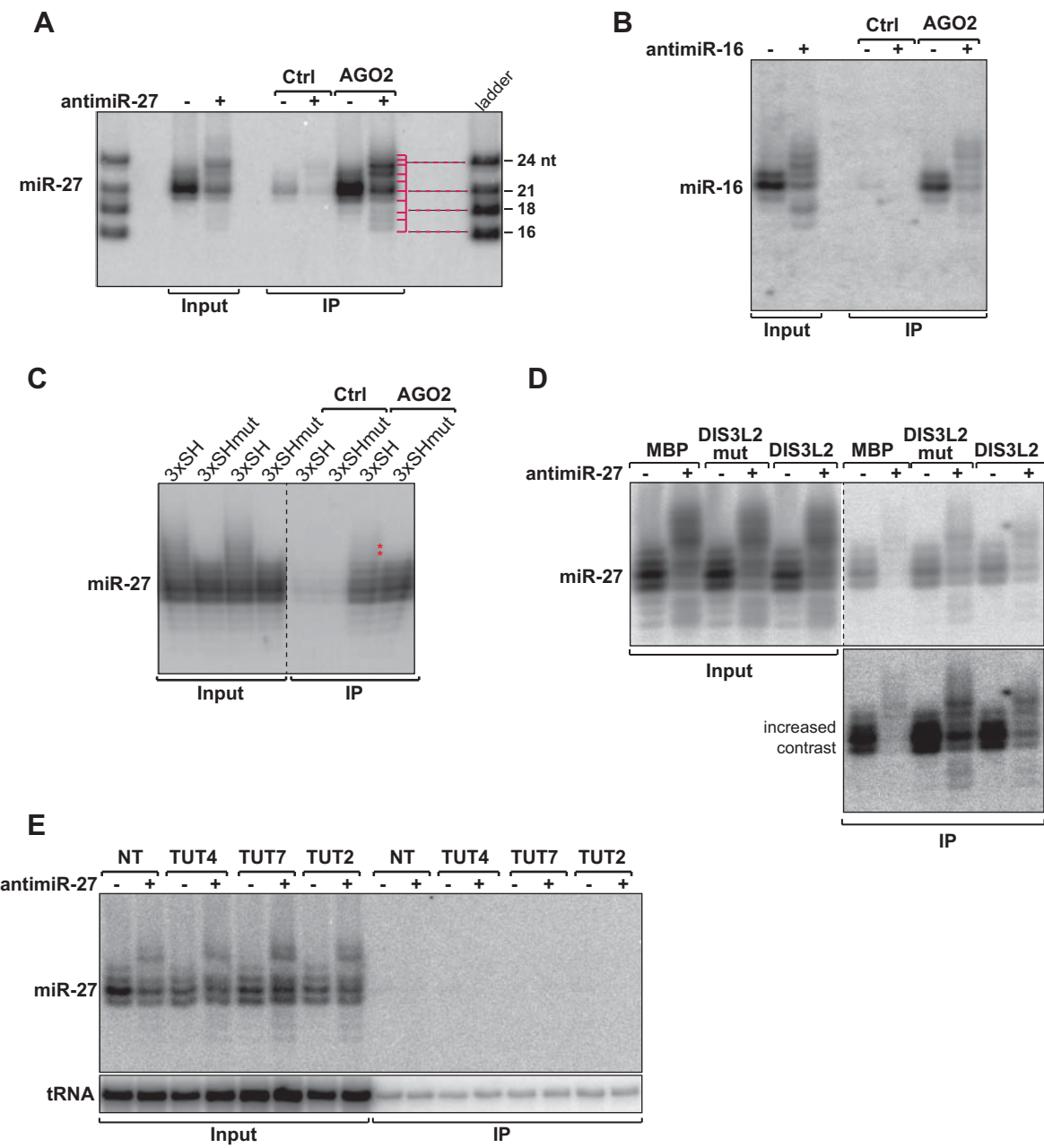
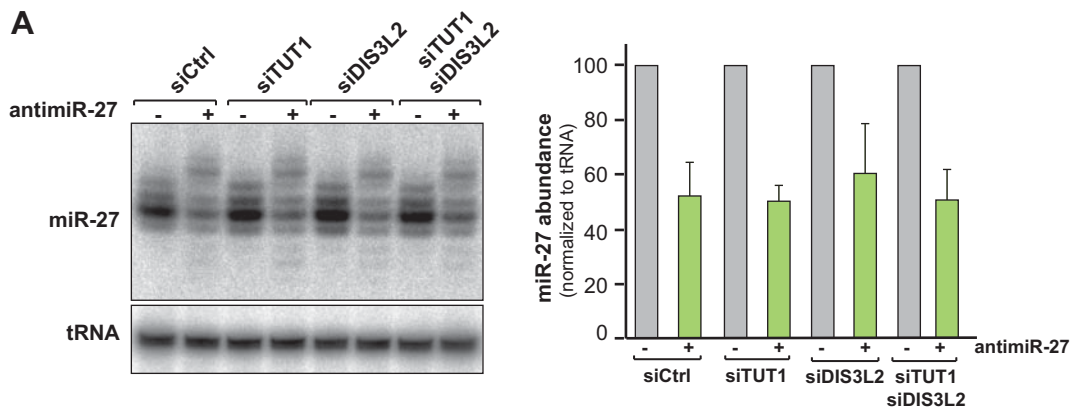
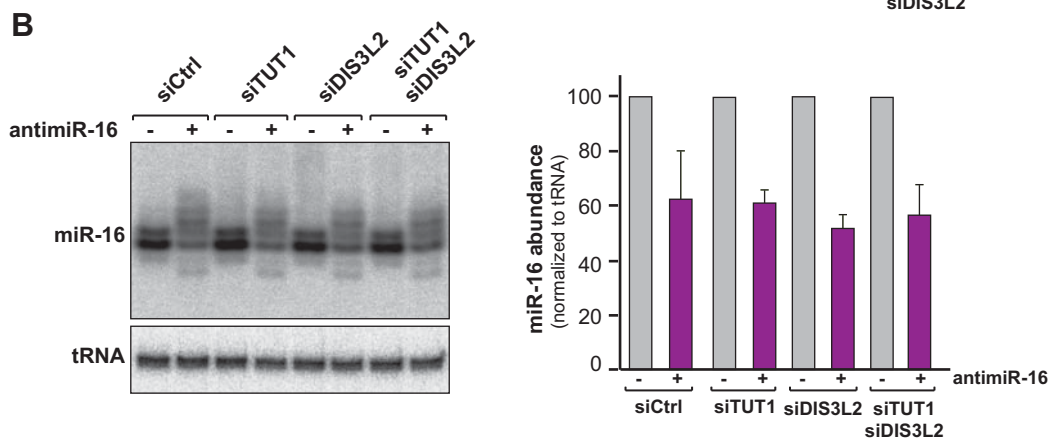
Figure S4

Figure S5

A



B



C miR-27a

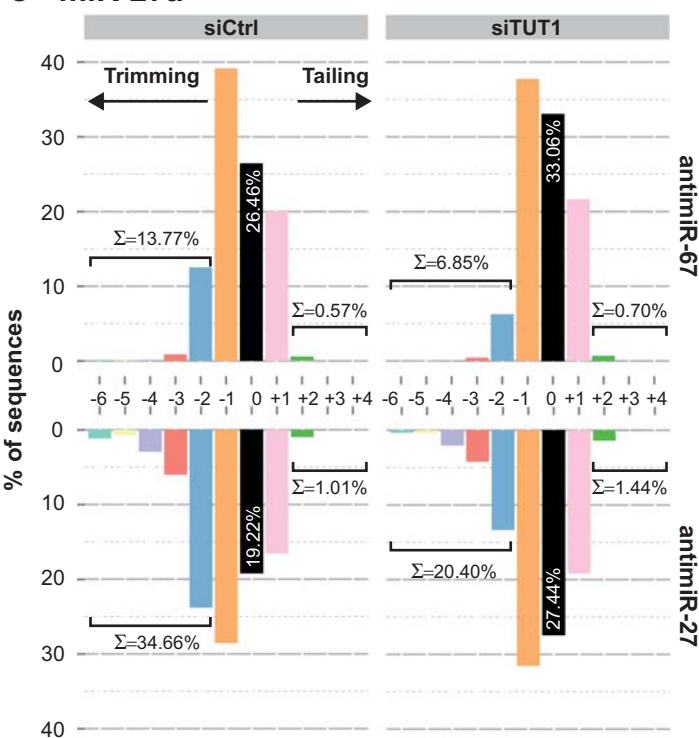


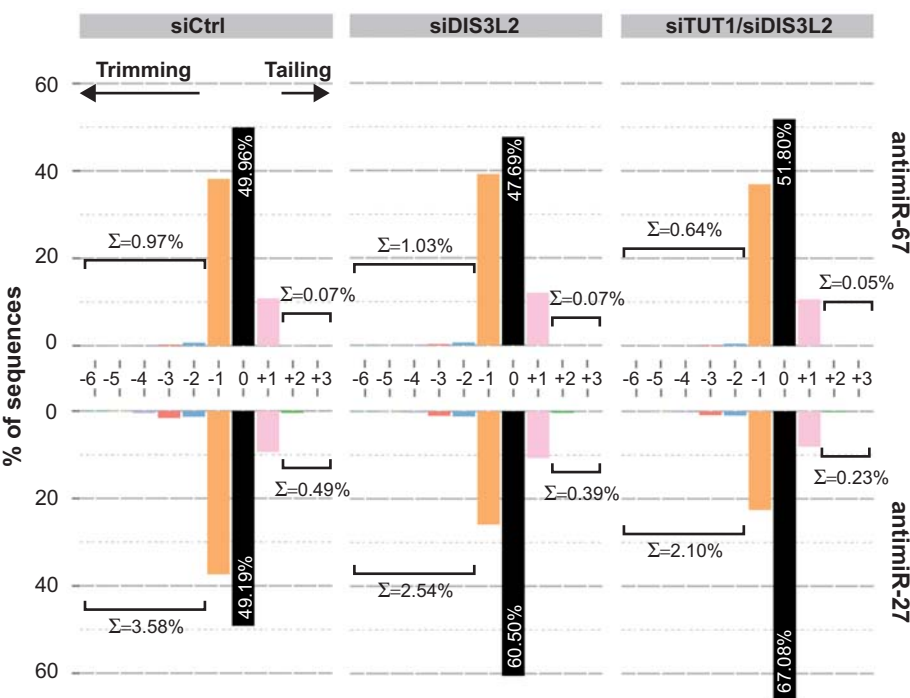
Figure S6

A

miR-27b (5'-3')	antimiR-67 (%)	antimiR-27 (%)
UUCACAGUGGCUAAGUUCUGCA	1.68	0.96
UUCACAGUGGCUAAGUUCUGC	46.18	44.18
UUCACAGUGGCUAAGUUCUGC U	8.09	6.56
UUCACAGUGGCUAAGUUCUG	34.49	33.38
UUCACAGUGGCUAAGUUCUG U	0.64	0.54
UUCACAGUGGCUAAGUUC	0.21	1.43
UUCACAGUGGCUAAGUUC U	0.60	1.14

B

miR-27b



C

miR-27b

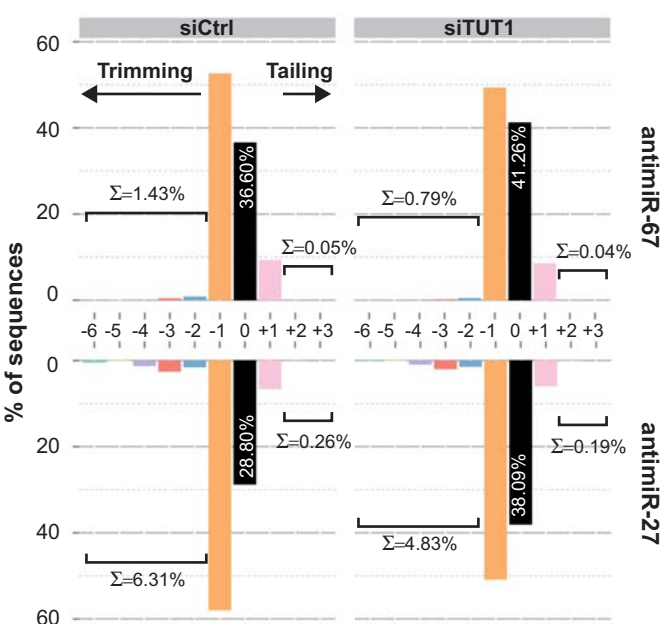


Figure S7

