Supplemental Material (online)

Title: The RNA-binding protein AATF coordinates rRNA maturation

Authors: Rainer W. Kaiser, Michael Ignarski, Eric L. Van Nostrand, Christian Frese, Manaswita Jain, Sadrija Cukoski, Heide Heinen, Melanie Schaechter, Konstantin Bunte, Peter Frommolt, Patrick Keller, Mark Helm, Katrin Bohl, Martin Höhne, Bernhard Schermer, Thomas Benzing, Katja Höpker, Christoph Dieterich, Gene W. Yeo, Roman-Ulrich Müller and Francesca Fabretti

Supplemental Tables and Figures

Supplemental Figure 1. (related to Fig. 4)
Nucleolar localization of AATF and re-distribution upon RNAPI and RNAPII inhibition

Supplemental Figure 2. (related to Fig. 5)
Impact of AATF depletion on non-ribosomal non-coding RNA species

Supplemental Figure 3. (related to Fig. 6)
Both R-proteins and RNAPI RNA interactors are part of the AATF protein interactome

Supplemental Table 1.
The AATF RNA interactome (after rRNA removal) as identified by eCLIP

Supplemental Table 2.
The AATF protein interactome as identified by MS/MS

Supplemental methods

Primary Antibodies

targeted protein	manufacturer	order ID
AATF	Abnova	H00026574-A01
AATF	Sigma-Aldrich	HPA004940
FLAG (M2)	Sigma-Aldrich	F1804-1MG
GFP (B2)	SantaCruz BioTech	sc-9996
TUBB/ß-Tubulin (E7)	Develop. Studies Hybridoma Bank	E7

DNA/RNA oligonucleotides

Genotyping	Sequence 5´-3´
AAVS1 Integration PCR HA-R 2R	FP GTGAGTTTGCCAAGCAGTCA
AAVS1 Integration PCR HA-R 2F	RP TATCCGCTCACAATTCCACA
AAVS 1 Empty locus	RP CGGAACTCTGCCCTCTAACG

qPCR	
Target	**Sequence 5´-3´**
18S	FP CTCAACACGGGAAACCTCAC RP CGCTCCACCAACTAAGAACG
AATF	FP CTTGGACACGGACAAAAGGT RP CTCCAGACCCTTCCTCATCA
ACTB	FP GGACTTCGAGACAAGAGATGG RP AGCACTGTGTTGGCGTACAG
HPRT	FP TGACACTGGCAAAACAATGCA RP GGTCCTTTTCACCAGCAAGCT
45S	FP ACCCACCCTCGGTGAGA RP CAAGGCACGCCTCTCAGAT

RNAi			
ON-TARGETplus Human AATF siRNA SMARTpool	CAAGCGCUCUGUCUAUCGA, GUGAUGACCUUCUCUAGUG, GCACUUAAAGCAUUGUUGA, CCAGGGUGAUUGACAGGUU	Dharmacon	L-004373-00-0005
ON-TARGETplus Mouse AATF siRNA SMARTpool	GACACGAGACAUAUAGUAA, GUGAGUAGCAUUAGAAAGU, CUAUAGGAAUCACACACUA, GGACGGAGUUGUUUCGAUC	Dharmacon	L-050461-00-0005
ON-TARGETplus Non-targeting pool	GUUUACAUGUCGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGUUUUCUGA, UGGUUUACAUGUUUUCCUA	Dharmacon	D-001810-10-05
Human AATF Custom siRNA (3' UTR specific)	GAGCAUUGUUACCGCCAAAUU, UUUGGCGGUAACAAUGCUCUU	Dharmacon	-

Statistical analysis

Unless stated otherwise, a two-tailed, unpaired student's t-test was used for statistical analysis of e.g. immunofluorescence data using GraphPad Prism v5. Statistical significance is indicated by asterisks as follows: *p<0,05, **p<0,01, ***p<0.,05, ****p<0,001, ns = not significant. Depicted results show the mean of at least three independent experiments/ biological replicates. Unless stated otherwise, error bars depict the standard deviation of the mean.

GO Term analysis

Gene ontology analyses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/)(1). Online results were downloaded as .txt- files and depicted using GraphPad Prism v5 or Microsoft Excel.

LC-MS analysis (protein interactome)

All samples were analyzed on a Q Exactive Plus (Thermo Scientific) that was coupled to an EASY nLC 1200 (Thermo Scientific). Peptides were loaded with solvent A (0.1% formic acid in water) onto an in-house packed analytical column (50 cm, 75 µm I.D., filled with 2.7 µm Poroshell EC120 C18, Agilent). Peptides were chromatographically separated at a constant flow rate of 250 nL/min using the following gradient: 10-23% solvent B (0.1% formic acid in 80 % acetonitrile) within 75.0 min, 23-39% solvent B within 5.0 min, 39-95% solvent B within 5.0 min, followed by washing and column equilibration. The mass spectrometer was operated in data-dependent acquisition mode. The MS1 survey scan was acquired from 300-1750 m/z at a resolution of 70,000. The top 10 most abundant peptides were isolated within a 1.8 Th window and subjected to HCD fragmentation at a normalized collision energy of 27%. The AGC target was set to 5e5 charges, allowing a maximum injection time of 120 ms. Product ions were detected in the Orbitrap at a resolution of 35,000. Precursors were dynamically excluded for 20.0s.

MS data analysis (protein interactome)

All mass spectrometric raw data were processed with Maxquant (version 1.5.3.8) using default parameters. Briefly, MS2 spectra were searched against the Uniprot HUMAN.fasta database, including a list of common contaminants. False discovery rates on protein and PSM level were estimated by the target-decoy approach to 1% (Protein FDR) and 1% (PSM FDR) respectively. The minimal peptide length was set to 7 amino acids and carbamidomethylation at cysteine residues was considered as a fixed modification. Oxidation (M) and Acetyl (Protein N-term) were included as variable modifications. The match-between runs option was disabled. Dimethyl triplex labeling quantification was used, and the re-quantify option was enabled. Maxquant output files were further processed using Perseus (version 1.5.5.3, standard settings, s0=0.1) and R/Bioconductor. Obtained protein ratios were $\log_2$ transformed. Proteins flagged as “only identified by site”, “reverse” and “potential contaminant” were removed from the data set. For comparison of AATF pulldowns with and without RNase the data were normalized such that the ratio of AATF in those two conditions equals zero.

Analysis of RNA modifications: LC-MS/MS

RNA hydrolysis to nucleoside level

The RNA samples were digested into nucleosides as it was described before(2): 500 ng of RNA were treated with 0.3 U nuclease P1 (Roche Diagnostics, Germany), 0.1 U snake venom phosphodiesterase (Worthington, USA), 200 ng Pentostatin (Adenosine deaminase inhibitor, Sigma-Aldrich, Germany) and 500 ng Tetrahydrouridine (Cytidine deaminase inhibitor, Merck-Millipore, Germany) in 1/10 vol. of 10x nuclease P1 buffer (10x NP1 Buffer: 9 vol 250 mM NH_4OAc, pH 5.0; 1 vol 2 mM $ZnCl_2$) for 2 h at 37 °C. Next, 1/10 vol. of 10x fast alkaline phosphatase buffer (10x FAP Buffer: 100 mM M NH_4OAc, pH 9.0)

and 1 U fast alkaline phosphatase (Fermentas, Germany) were added and the mixture was incubated at 37 °C for another 60 min.
For LC-MS/MS analysis 3x 25 ng of each RNA sample were employed (technical triplicates).

Relative quantification of modified nucleosides *via* LC-MS/MS

For RNA analysis, an Agilent1260 series equipped with a diode array detector (DAD) and a Triple Quadrupole mass spectrometer (Agilent 6460) were utilized. In addition to that, a Synergy Fusion RP18 column (4 µm particle size, 80 Å pore size, 250 mm length, 2 mm inner diameter) from Phenomenex (Germany) was used at 35°C and separation of nucleosides was performed using a flow rate of 0.35 mL/min. 5 mM ammonium acetate buffer (pH 5.3) served as solvent A and acetonitrile (LCMS grade, Sigma-Aldrich, Germany) as solvent B. The respective LC-gradient is shown in table M1 below.

Table M1: Gradient for nucleoside separation prior to MS analysis.

Time [min]	Solvent A [%]	Solvent B[%]
0	100	0
10	92	8
20	60	40
23	100	0
30	100	0

The main nucleosides (cytidine, uridine, guanosine, adenosine) were measured photometrically at 254 nm by the DAD and modified nucleosides were analyzed *via* the mass spectrometer (operated in the positive ion mode) equipped with an electrospray ion source (Agilent Jet Stream, for settings see table M2).

Table M2: Electrospray ionization settings.

Parameter	
Gas temperature	350 °C
Gas flow	8 L/min
Nebulizer pressure	50 psi
Sheath gas temperature	350 °C
Sheath gas flow	12 L/min
Capillary voltage	3000 V

To monitor the mass transitions of the modified nucleosides, Agilent Mass Hunter software was used in the Dynamic Multiple Reaction Monitoring (DMRM) mode. Further details are displayed in Table M3.

Table M3: MS-parameters of the Dynamic Multiple Reaction Monitoring mode.

Modified nucleoside	Molecular weight [Da]	Precursor ion [m/z]	Product ion [m/z]	Retention time [min]	Fragmentor voltage [V]	Collision energy [eV]	Cell accelerator voltage [V]
A_m	281	282	136	16	92	13	2
$^{13}C/^{15}N$-A_m	292	293	141	16	92	13	2
C_m	257	258	112	9.5	60	9	2
$^{13}C/^{15}N$-C_m	270	271	119	9.5	60	9	2
G_m	297	298	152	12.6	72	5	2
$^{13}C/^{15}N$-G_m	313	314	162	12.6	72	5	2
m^1A	281	282	150	5.6	92	17	2
$^{13}C/^{15}N$-m^1A	297	298	161	5.6	92	17	2
$m^2{}_2G$	311	312	180	14.9	82	9	2
$^{13}C/^{15}N$-$m^2{}_2G$	328	329	192	14.9	82	9	2
$m^6{}_2A$	295	296	164	18.6	102	17	2
$^{13}C/^{15}N$-$m^6{}_2A$	312	313	176	18.6	102	17	2
m^5C	257	258	126	8.8	40	9	2
$^{13}C/^{15}N$-m^5C	270	271	134	8.8	40	9	2
m^6A	281	282	150	16.7	92	17	2
$^{13}C/^{15}N$-m^6A	297	298	161	16.7	92	17	2
m^7G	297	298	166	8.7	82	9	2
$^{13}C/^{15}N$-m^7G	313	314	177	8.7	82	9	2
Ψ	244	245	209	4	81	5	2
$^{13}C/^{15}N$- Ψ	255	256	220	4	81	5	2
U_m	258	259	113	11.1	66	5	2
$^{13}C/^{15}N$- U_m	270	271	119	11.1	66	5	2

Resulting spectra were processed using Agilent MassHunter Qualitative Analysis Software: In a first step, the recorded UV chromatogram of the main nucleoside guanosine was extracted to receive the 'area under the curve' (AUC) and then, the UV-signal derived from the SIL-IS (see below) was subtracted. After that, calibration measurements of guanosine dilutions (5-500 pmol) were applied for exact quantification and the amount of injected guanosine (in pmol) of each RNA sample was calculated by using the resulting guanosine calibration factor.

Quantification of modified nucleosides was achieved by utilizing ^{13}C- and ^{15}N-labeled total RNA from *C. elegans* as a Stable Isotope-Labeled Internal Standard (SIL-IS)(3): For each investigated modification, 10 calibration solutions (0.1-5000 fmol) consisting of the modified reference nucleoside (all Sigma-Aldrich, Germany) including 20 ng SIL-IS were prepared and subjected to LC-MS measurement. The

resulting mass spectra of each solution were then processed by integrating the MS/MS peaks to receive the AUC values. In the next step, the ratios of the extracted peak areas of the modified nucleosides and their corresponding ^{13}C-labeled isotopes were calculated and plotted using Microsoft Excel to receive calibration curves. This enabled the determination of the slopes of these curves which corresponds to the nucleoside-isotope response factors for each modification. To obtain the modification content in the RNA samples, the same amount of SIL-IS (= 20 ng) was added and again the resulting ratio of modified nucleoside and its ^{13}C-labeled isotope was calculated. Afterwards, the latter was divided by the corresponding modification response factor to receive the modification amount in fmol. Lastly, the ratios of modified nucleosides and the injected guanosine amounts were calculated. Further and more detailed information can be found in Thüring et al.(3) and Kellner et al.(2).

Table M4: List of modifications

Abbreviation	**Name**
A_m	2′-O-methyladenosine
C_m	2′-O-methylcytidine
G_m	2′-O-methylguanosine
m^1A	1-methyladenosine
$m^2{}_2G$	*N2*,*N2*-dimethylguanosine
$m^6{}_2A$	*N6*,*N6*-dimethyladenosine
m^5C	5-methylcytidine
m^6A	*N6*-methyladenosine
m^7G	7-methylguanosine
Ψ	pseudouridine
U_m	2′-O-methyluridine

Supplemental Figure legends

Suppl. Figure 1: Nucleolar localization of AATF and re-distribution upon RNAP inhibition

A The amino acid sequence of AATF was analyzed for nucleolar localization sequences using the Nucleolar localization sequence Detector tool (NoD)(4), where the AATF sequence (UNIPROT ID Q9NY61) was plotted against a NoLS likelihood score. Two NoLSs (as assessed by a score of > 0.8) were identified in the C-terminal part of the protein: AA 326-345 (NoLS 1) and AA 494-522 (NoLS 2).

B/C Immunofluorescence microscopy showing GFP::AATF expressing transgenic U2OS cells co-localizing with the nucleolar marker Nucleolin (NCL::mCherry). The transgenic protein is also recognized by an AATF-specific antibody, showing co-localization with the endogenous protein.

D Immunofluorescence microscopy of U2OS GFP::AATF-expressing cells that were cultured on microscopy plates and treated with either DMSO (control), 2µM CX-5461 (RNAPI inhibitor), 2µg/ml Flavopiridol (RNAPII inhibitor) or 0,1µg/ml Actinomycin D (inhibits both polymerases) for two hours. The lower panel shows the analysis of the inhibitor-specific response regarding e.g. nucleoli size, nucleolar diameter and nucleoplasmic fluorescence using a custom-made macro integrated into ImageJ Fiji v1 (5). At least 100 cells were imaged and analyzed per replicate.

Suppl. Figure 2: Impact of AATF depletion on non-ribosomal non-coding RNA species

A Box plots showing the relative expression of microRNAs isolated from AATF-depleted and control U2OS cells. After loss of AATF, there was no difference in global miRNA expression.

B Analysis of the most significantly dysregulated miRNA transcripts shows that these miRNAs are strongly affected regarding their abundance upon loss of AATF.

C Analysis of rRNA modifications of whole RNA isolated from murine IMCD cells after transfection with AATF-targeted siRNA or a control using LC-MS/MS (see suppl. methods for details)**.** Here, no differences between control cells and cells depleted of AATF were detected for the most common modifications.

Suppl. Figure 3: Both R-proteins and RNAPI RNA interactors are part of the AATF protein interactome

A Bar chart depicting R-proteins that are part of the (partially) non-RNA- and RNA-dependent protein interactome of AATF. The grey lower part of the bars depict the percentage of known R-proteins(6) in the analyzed subgroups.

B Venn diagram depicting the high level of overlap of the AATF protein interactome as well as the RNAPI RNA interactome as published by Piñeiro et al.(7).

Supplemental Table legends

Suppl. Table 1: The AATF RNA interactome (after rRNA removal) as identified by eCLIP

A: HepG2, replicate 1, annotated .bed files

B: HepG2, replicate 2, annotated .bed files

C: K562, replicate 1, annotated .bed files

D: K562, replicate 2, annotated .bed files

Each dataset was filtered for significant peaks FC>=3 and p-value>=5, and pooled.

To define a list of bound RNAs, replicates ENSG identifiers were removed

The ENSG IDs were then annotated for Gene Stable ID, Gene name and Gene Type using biomart on the ENSEMBL website (ensembl genes 93) 17 IDs were not annotated (obsolete)

E: "bound RNAs", 665 targets with at least one significant peak in 1 experiment (FC≥3 and p-value≥5) in 1 out of the 4 experiments (2 replicates K562 cells, 2 replicates HepG2 cells)

F: "at least in 2 experiments", 292 targets all sequences containing a significant peak (FC>=3 and p-value>=5) in at least 2 out of the 4 experiments (2 replicates K562 cells, 2 replicates HepG2 cells)

Suppl. Table 2: The AATF protein interactome as identified by MS/MS

"data": all proteins identified**"A-AATF bona fide interactors":** only proteins that show log_2FC >=2 and $-log_{10}$ pval >=1.3 in AATF vs. GFP**"B-RNA dependent":** all proteins among bona fide interactors that show a log_2FC <=-2 and $-log_{10}$ pval >=1.3 in AATF+RNAse vs AATF normalized

"C-partially RNA dependent": all proteins among bona fide interactors that show a log_2FC <=-1 and $-log_{10}$ pval >=1.0 in AATF+RNAse vs AATF normalized

"D-partially non-RNA dependent": all proteins among bona fide interactors that show a log_2FC >=1 and $-log_{10}$ pval >=1.0 in AATF RNAse vs. GFP

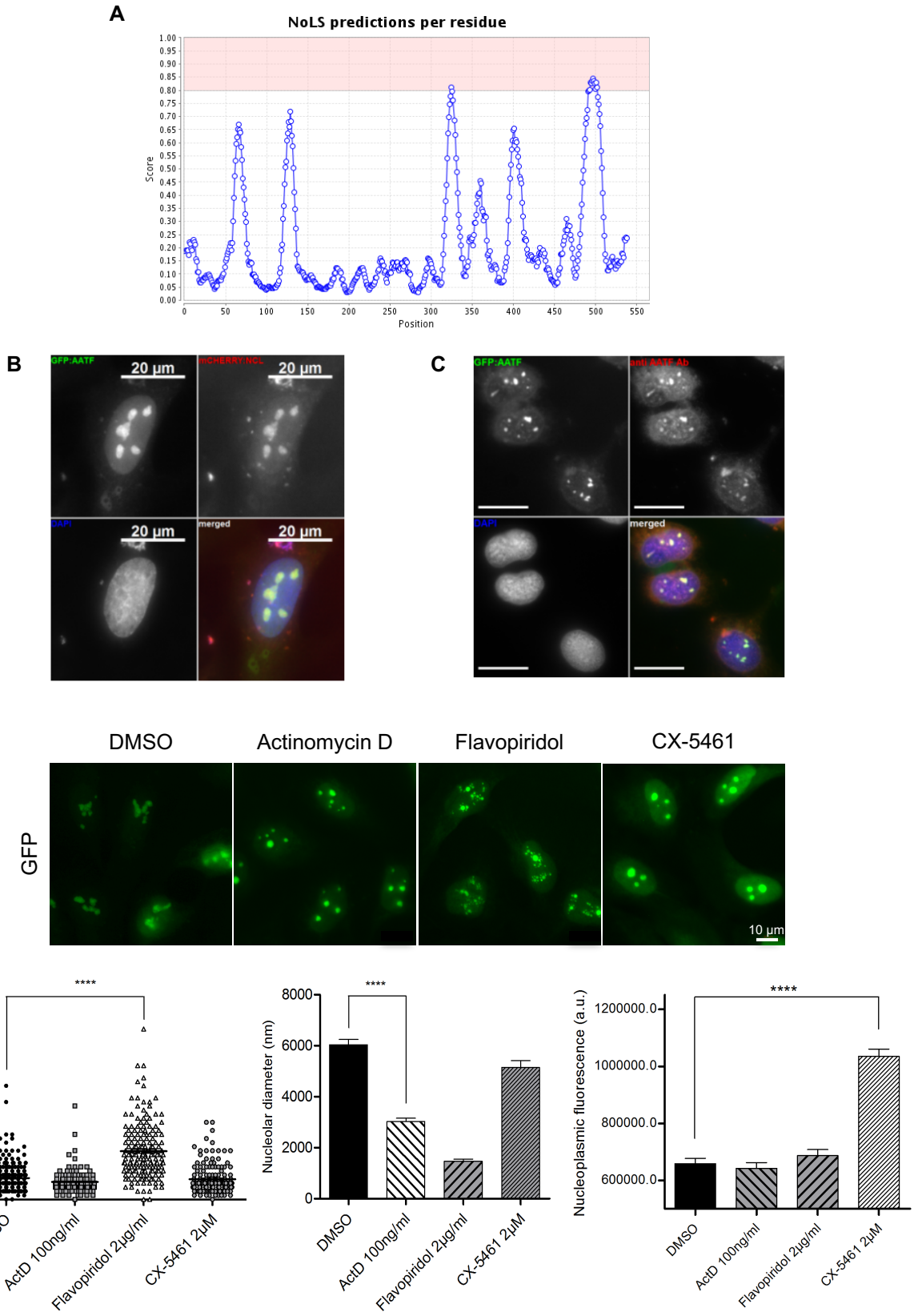
"E-non RNA dependent": all proteins among bona fide interactors that show a log_2FC >=2.0 and $-log_{10}$ pval >=1.3 in AATF RNAse vs. GFP

"F-AATF bona fide vs Pineiro": all proteins in A matched versus the RNApol I RNA interactome (3)

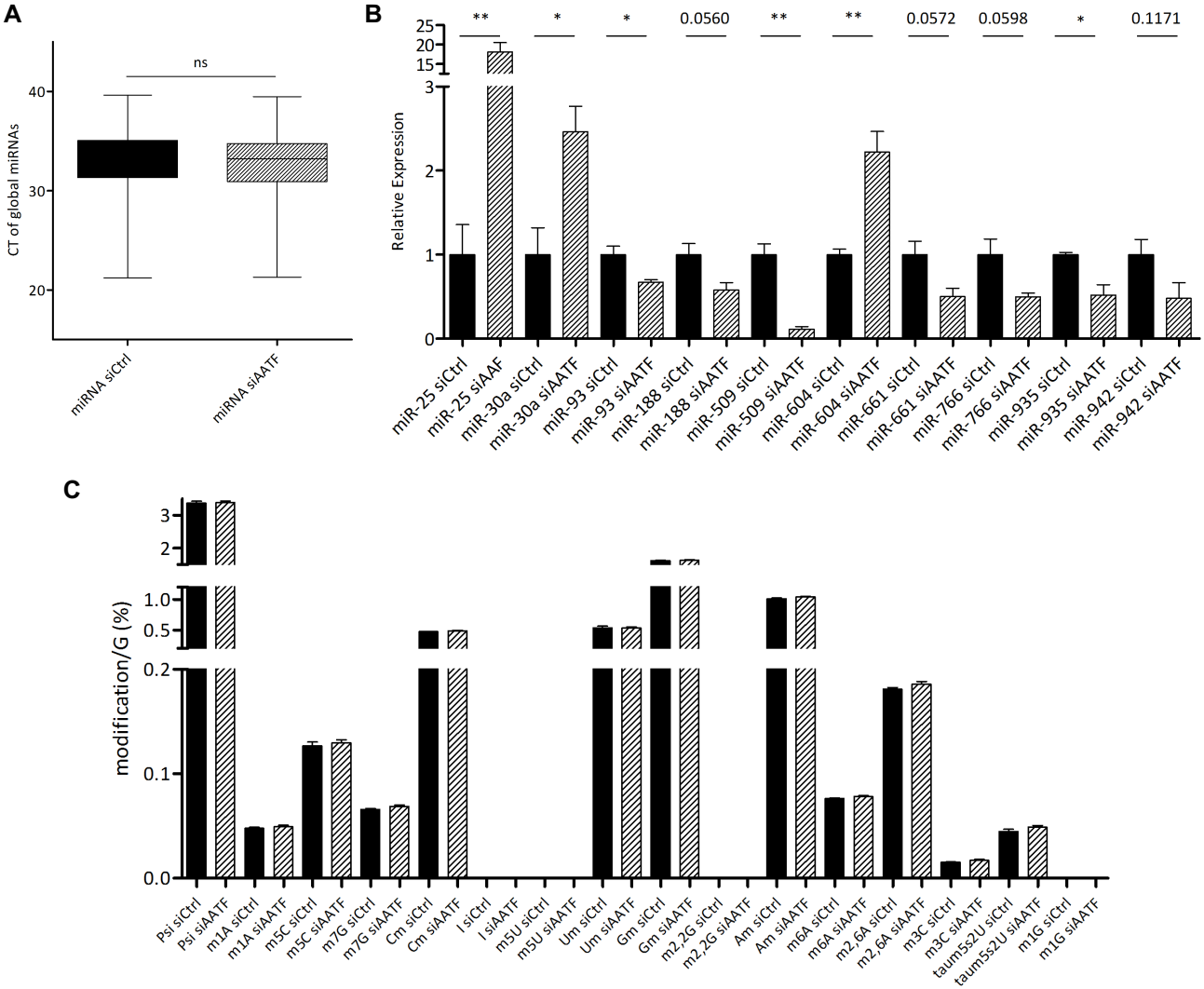
References

1. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, **4**, 44-57.
2. Kellner, S., Ochel, A., Thuring, K., Spenkuch, F., Neumann, J., Sharma, S., Entian, K.D., Schneider, D. and Helm, M. (2014) Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res*, **42**, e142.
3. Thüring, K., Schmid, K., Keller, P. and Helm, M. (2016) Analysis of RNA modifications by liquid chromatography-tandem mass spectrometry. *Methods*, **107**, 48-56.
4. Scott, M.S., Troshin, P.V. and Barton, G.J. (2011) NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. *BMC Bioinformatics*, **12**, 317.
5. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods*, **9**, 676-682.
6. Nicolas, E., Parisot, P., Pinto-Monteiro, C., de Walque, R., De Vleeschouwer, C. and Lafontaine, D.L. (2016) Involvement of human ribosomal proteins in nucleolar structure and p53-dependent nucleolar stress. *Nat Commun*, **7**, 11390.
7. Pineiro, D., Stoneley, M., Ramakrishna, M., Alexandrova, J., Dezi, V., Juke-Jones, R., Lilley, K.S., Cain, K. and Willis, A.E. (2018) Identification of the RNA polymerase I-RNA interactome. *Nucleic Acids Res*.

Supplemental figure 1 (relates to figure 4)

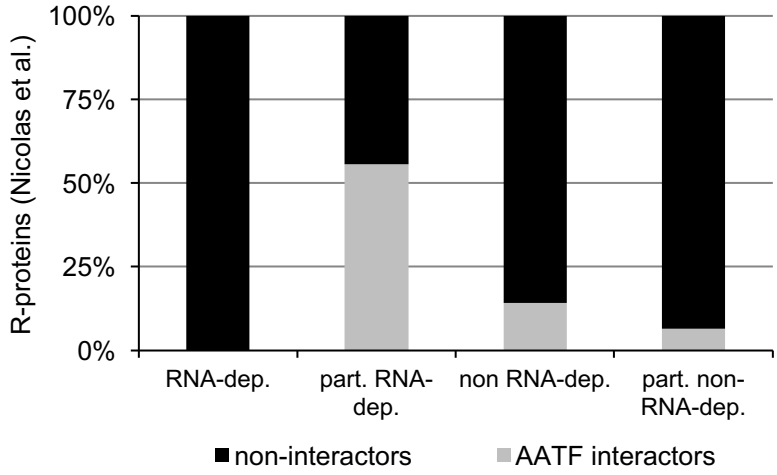


Supplemental figure 2 (relates to figure 5)



Supplemental figure 3 (relates to figure 6)

A



B