PUMILIO, but not RBMX, binding is required for regulation of genomic stability by noncoding RNA *NORAD*

Mahmoud M. Elguindy[1,2], Florian Kopp[1], Mohammad Goodarzi[3], Frederick Rehfeld[1], Anu Thomas[1], Tsung-Cheng Chang[1], Joshua T. Mendell[1,4,5,6*]

[1]Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[2]Medical Scientist Training Program, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[3]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[4]Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas TX 75390, USA
[5]Hamon Center for Regenerative Science and Medicine, University of Texas Southwestern Medical Center, Dallas TX 75390, USA
[6]Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

* Correspondence: Joshua.Mendell@UTSouthwestern.edu

ABSTRACT

NORAD is a highly-conserved and abundant long noncoding RNA (lncRNA) that is required for maintenance of genomic stability in mammals. Although initial characterization of *NORAD* established it as a negative regulator of PUMILIO (PUM) proteins in the cytoplasm, a nuclear role for *NORAD* in genome maintenance through an interaction with the RNA binding protein RBMX has also been reported. Here we addressed the relative contributions of *NORAD*:PUM and *NORAD*:RBMX interactions to the regulation of genomic stability by this lncRNA. Extensive RNA FISH and fractionation experiments established that *NORAD* localizes predominantly to the cytoplasm with or without DNA damage. Moreover, genetic rescue experiments demonstrated that PUM binding is required for maintenance of genomic stability by *NORAD* whereas binding of RBMX is dispensable for this function. These data therefore establish an essential role for the *NORAD*:PUM interaction in genome maintenance and provide a foundation for further mechanistic dissection of this pathway.

INTRODUCTION

Long noncoding RNAs (lncRNAs) have emerged as important regulators of diverse biological processes. Among these transcripts, *Noncoding RNA activated by DNA damage* (*NORAD*) is particularly noteworthy, due to its unusually abundant expression in mammalian cells and tissues and strong evolutionary conservation across mammalian species. Initial studies of *NORAD* revealed that this lncRNA is required to maintain genomic stability in mammalian cells (Lee et al., 2016), and provided strong evidence that this function is mediated through the ability of *NORAD* to bind to and negatively regulate PUMILIO RNA binding proteins (PUM1 and PUM2) in the cytoplasm (Lee et al., 2016; Tichon et al., 2016). PUM proteins bind with high specificity to the 8 nucleotide (nt) PUMILIO response element (PRE) (UGUANAUA or UGUANAUN) on target messenger RNAs (mRNAs), triggering their deadenylation, decapping, and eventual degradation (Miller and Olivas, 2011; Quenault et al., 2011). Notably, *NORAD* contains approximately 18 conserved PREs and has the capacity to bind a large fraction of PUM1/2 within the cell, thereby limiting the availability of these proteins to repress target mRNAs. Consequently, loss of *NORAD* results in PUM hyperactivity and increased repression of PUM targets that include important regulators of mitosis, DNA repair, and DNA replication, resulting in a dramatic genomic instability phenotype in *NORAD*-deficient cells and mouse tissues (Kopp et al., 2019; Lee et al., 2016). Accordingly, PUM1/2 overexpression is sufficient to phenocopy loss of *NORAD* in both human cells and mice (Kopp et al., 2019; Lee et al., 2016), while PUM1/2 loss-of-function suppresses the genomic instability phenotype in *NORAD* knockout cells (Lee et al., 2016).

Recently, an alternative mechanism for the regulation of genomic stability by *NORAD* was proposed (Munschauer et al., 2018). Proteomic analysis of the *NORAD* interactome revealed an interaction with RBMX, an RNA binding protein that contributes to the DNA damage response

(Adamson et al., 2012). Subsequent experiments suggested that the *NORAD*:RBMX interaction facilitates the assembly of a ribonucleoprotein (RNP) complex in the nucleus that includes Topoisomerase I (TOP1) and other proteins that are critical for genome maintenance. Importantly, PUM and RBMX interact with different sites on *NORAD* and function in distinct subcellular compartments. Thus, while it remains to be determined whether the *NORAD*:RBMX interaction is necessary for regulation of genomic stability, both PUM and RBMX may play important, non-mutually exclusive roles in the genome maintenance functions of *NORAD*.

Here we further examined the mechanism by which *NORAD* functions to maintain genome stability in human cells and directly tested the requirement for PUM and RBMX interactions in this activity. RNA fluorescent in situ hybridization (FISH) using a panel of probes spanning the entire length of *NORAD*, as well as cellular fractionation studies, definitively demonstrated that this lncRNA localizes predominantly to the cytoplasm and does not traffic to the nucleus upon induction of DNA damage. Genetic rescue experiments in *NORAD* knockdown cells established that PUM binding is essential for maintenance of genomic stability whereas interaction with RBMX is completely dispensable for this function. Further experiments demonstrated that RBMX is not required for induction of *NORAD* following DNA damage nor its cytoplasmic localization. Together, these studies establish the importance of the *NORAD*:PUM axis in regulating genomic stability in mammalian cells and provide a foundation for further dissection of the mechanism and physiologic role of this pathway.

RESULTS AND DISCUSSION

NORAD localizes predominantly to the cytoplasm with or without DNA damage

Initial studies of *NORAD* reported a predominantly cytoplasmic localization of this lncRNA in human cell lines, based on RNA FISH using pools of fluorescently-labeled oligonucleotide

probes and subcellular fractionation experiments (Lee et al., 2016; Tichon et al., 2016). Recently, however, a distinct localization pattern was reported based upon RNA FISH performed using a commercially-available kit with a proprietary set of oligonucleotide probes that hybridize to an unknown segment of *NORAD* (Munschauer et al., 2018). In these more recent experiments, *NORAD* was reported to localize equally between the nucleus and cytoplasm and appeared to redistribute almost entirely to the nuclear compartment upon treatment of cells with the DNA damaging agents camptothecin and doxorubicin. Importantly, a common cell line (human colon cancer cell line HCT116) was used in the different studies, arguing against a cell-type specific difference in *NORAD* trafficking as the cause of these discordant results.

We considered the possibility that the disparate observed localization patterns could be due to unrecognized processing of the *NORAD* transcript, such that different segments of the RNA that are recognized by the different FISH probes accumulate in distinct subcellular compartments. To investigate this possibility, and to reliably establish the localization of the entire *NORAD* transcript, we generated a panel of 11 *in vitro* transcribed RNA FISH probes spanning the complete *NORAD* sequence (**Figure 1A**) (Mito et al., 2016). One probe, that recognized a segment of *NORAD* containing an Alu repeat element (probe 7), gave rise to a nonspecific nuclear signal that was present in both $NORAD^{+/+}$ and $NORAD^{-/-}$ HCT116 cells. The remaining 10 probes produced a highly consistent, predominantly cytoplasmic, punctate localization pattern in wild-type cells that was absent in $NORAD^{-/-}$ cells (**Figure 1A-B**). These results were confirmed using subcellular fractionation followed by quantitative reverse transcriptase-PCR (qRT-PCR) using primers located at the 3′ or 5′ ends of *NORAD*, which revealed that 80-90% of the transcript is localized to the cytoplasmic compartment (**Figure 1C**).

Next, we examined *NORAD* localization following treatment of cells with agents that induce DNA damage. RNA FISH using the panel of probes spanning *NORAD* revealed clear cytoplasmic localization after treatment with doxorubicin or camptothecin, without a significant increase in nuclear signal compared to untreated cells (**Figure 2A-B**). Consistent with the previously reported induction of *NORAD* after DNA damage (Lee et al., 2016), a clear increase in cytoplasmic *NORAD* signal was apparent in treated cells (**Figure 2B**). These findings were further corroborated by subcellular fractionation experiments following treatment with DNA damaging agents, which confirmed that *NORAD* remained predominantly in the cytoplasmic compartment at all time points (**Figure 2C and Figure 2 – figure supplement 1A**). Interestingly, we observed a modest increase in nuclear *NORAD* levels that peaked after 12 hours of camptothecin or doxorubicin treatment. We speculated that this might represent a burst of *NORAD* transcription in response to accumulating DNA damage. To test this hypothesis, we co-treated cells with DNA damaging agents and the transcriptional inhibitor actinomycin D. As expected, this completely abrogated any detectable increase in nuclear *NORAD* abundance in treated cells (**Figure 2D and Figure 2 – figure supplement 1B**).

These comprehensive RNA FISH and subcellular fractionation experiments provide definitive evidence that *NORAD* is a cytoplasmic RNA in HCT116 cells and does not redistribute to the nucleus upon DNA damage. These findings are consistent with the reported localization of *NORAD* in other human and mouse cell lines (Kopp et al., 2019; Lee et al., 2016; Tichon et al., 2016). We conclude that the disparate localization pattern observed using a commercially-available RNA FISH probe set (Munschauer et al., 2018) most likely represented a non-specific signal.

PUM1, PUM2, and RBMX are components of the *NORAD* interactome

Previous crosslinking-immunoprecipitation coupled with high throughput sequencing (CLIP-seq) studies demonstrated that *NORAD* is the preferred binding partner of PUM2 in both human cells (Lee et al., 2016) and mouse brain (Kopp et al., 2019). In light of these findings, it was surprising that PUM1/2 were not reported among the most enriched *NORAD*-bound proteins in the recent RNA antisense purification with quantitative mass spectrometry (RAP-MS) experiments that identified the *NORAD*:RBMX interaction (Munschauer et al., 2018). Since these RAP-MS experiments utilized pulse labeling with 4-thiouridine to crosslink *NORAD* to protein interactors, a bias towards detection of proteins that bind to newly synthesized *NORAD* would be expected, likely explaining the enrichment of nuclear interactors observed. Nevertheless, we reanalyzed the published RAP-MS dataset to determine whether PUM1 or PUM2 were enriched in *NORAD* pull-downs compared to control *RMRP* pull-downs. Peptides were identified and scored using a combined algorithm that employed three search engines (Sequest HT, Mascot, and MS Amanda). Isoforms of PUM1 and PUM2, along with RBMX, were indeed identified as significantly-enriched interacting partners of *NORAD* compared to *RMRP* (**Figure 3 – figure supplement 1**). Notably, PUM1 was more enriched than PUM2 in our analysis, which may reflect its higher abundance in HCT116 cells (Lee et al., 2016). These results confirmed that both PUM proteins and RBMX are identified by RAP-MS as significant *NORAD*-interacting partners.

Binding of PUMILIO, but not RBMX, to *NORAD* is necessary for genome stability

Genetic epistasis experiments have strongly implicated a role for PUM1/2 in the regulation of genomic stability by *NORAD*, with PUM2 overexpression phenocopying, and PUM1/2 knockdown suppressing, the effects of *NORAD* deletion (Kopp et al., 2019; Lee et al., 2016). Nevertheless, it has not yet been directly tested whether binding of PUM1/2 is required for *NORAD* function. Similarly, a requirement for RBMX binding in genome maintenance by

175 *NORAD* has not been established. Therefore, to directly interrogate the importance of PUM and
176 RBMX binding for *NORAD* function, we generated a series of mutant *NORAD* constructs lacking
177 either PUM or RBMX binding sites (**Figure 3A**). For each of the 18 PREs within *NORAD*, the
178 UGU sequence, which is essential for PUM binding (Bohn et al., 2018; Van Etten et al., 2012),
179 was mutated to ACA to abolish PUM interaction (PREmut construct). To remove the RBMX
180 binding site, the first 898 nucleotides (nt) of *NORAD* were deleted (Munschauer et al., 2018) (5′
181 deletion construct). We also sought to determine whether PUM or RBMX binding regions of
182 *NORAD* could represent minimal functional domains that are sufficient for maintaining genomic
183 stability. To this end, we generated a fragment comprising *NORAD* domain 4 (ND4), which
184 represents the most conserved of five repeated segments within this lncRNA termed *NORAD*
185 domains (Lee et al., 2016) and contains 4 PREs (nt 2494-3156). An RBMX binding site
186 fragment, representing the 5′ end of *NORAD* (nt 33-898) was also generated (5′ fragment).
187
188 Wild-type or mutant *NORAD* constructs, as well as a control GFP sequence, under the control
189 of a constitutive promoter were introduced into the *AAVS1/PPP1R12C* locus of HCT116 cells
190 using a previously published TALEN pair (Sanjana et al., 2012) (**Figure 3B**). Endogenous
191 *NORAD* was then depleted using CRISPR interference (CRISPRi) with a single-guide RNA
192 (sgRNA) targeting the endogenous *NORAD* promoter. *NORAD* transcripts produced from the
193 *AAVS1* locus were expressed at near physiologic levels in this system.
194
195 We next used UV crosslinking and RNA immunoprecipitation (RIP) to assess binding of wild-
196 type and mutant *NORAD* transcripts to endogenous PUM1, PUM2, and RBMX. Pull-downs of
197 each of these proteins resulted in the expected enrichment of wild-type *NORAD*, but not
198 *GAPDH*, relative to immunoprecipitation with control IgG (**Figure 4A and Figure 4 – figure**
199 **supplement 1A**). The PREmut transcript as well as the 5′ fragment did not bind to PUM1/2 but

were recovered in RBMX RIP samples as efficiently as wild-type *NORAD*. In contrast, the 5′ deletion construct and ND4 fragment retained PUM1 and PUM2 binding activity, but interaction with RBMX was abolished. Furthermore, RNA FISH documented a predominantly cytoplasmic localization pattern of each construct (**Figure 4B**).

Genome stability was assessed in cell populations expressing wild-type or mutant *NORAD* constructs by quantifying the number of aneuploid cells in each cell population using DNA FISH for marker chromosomes 7 and 20, as described previously (Lee et al., 2016). Knockdown of endogenous *NORAD* in GFP-control cells resulted in a significant accumulation of aneuploid cells (**Figure 4C**). The frequency of aneuploidy observed under these conditions was very similar to that observed previously in *NORAD*$^{-/-}$ HCT116 cells (Lee et al., 2016). Expression of wild-type *NORAD* in this system was sufficient to fully suppress the accumulation of aneuploid cells. Cells expressing the PREmut transcript, however, exhibited high levels of aneuploidy, demonstrating that loss of PUM binding abrogated the ability of *NORAD* to maintain genomic stability. Furthermore, the 5′ deletion construct that lacks the RBMX binding site, but preserves the PUM interaction, was fully functional in this assay and completely prevented the accumulation of aneuploid cells. Thus, RBMX binding to *NORAD* is dispensable for genome maintenance. Remarkably, we observed a strong suppression of aneuploidy in cells expressing the minimal ND4 fragment, further supporting the centrality of the PUM interaction for *NORAD* function, while the 5′ fragment of *NORAD* had no activity in this assay. Overall, these data provide compelling evidence that PUM, but not RBMX, binding to *NORAD* is necessary for the maintenance of genomic stability by this lncRNA.

RBMX is not required for *NORAD* expression or localization

Although RBMX is not required for maintenance of genomic stability by *NORAD*, we were able to confirm binding of this protein to the 5′ end of *NORAD*, as reported (Munschauer et al., 2018) (**Figure 3 and Figure 4A**). Thus, we investigated whether RBMX functions as an upstream regulator of *NORAD* expression or localization. Depletion of RBMX using CRISPRi resulted in an increase in *NORAD* expression that was further augmented by doxorubicin treatment (**Figure 5A**). We speculate that this increase in *NORAD* levels may be an indirect effect of the previously reported accumulation of DNA damage caused by RBMX loss of function (Adamson et al., 2012). Additionally, RBMX knockdown did not alter the predominantly cytoplasmic localization of *NORAD*, as indicated by subcellular fractionation experiments (**Figure 5B**). We conclude that RBMX is not an essential co-factor for *NORAD* expression or localization.

In sum, these results establish the essential role of PUM binding for the regulation of genomic stability by *NORAD*. A systematic examination of the subcellular localization of this lncRNA unequivocally established its predominantly cytoplasmic localization under baseline conditions as well as after treatment with DNA-damaging agents. Moreover, genetic complementation experiments demonstrated that PUM binding is essential, whereas RBMX interaction is dispensable, for the genome maintenance function of *NORAD*. These results further define and clarify the *NORAD* molecular mechanism of action and direct future investigation towards elucidation of the regulation and physiologic roles of the *NORAD*:PUM axis.

METHODS

Cell culture and generation of HCT116 CRISPRi cell line

HCT116 cells (ATCC) were cultured in McCoy's 5a media (Thermo Fisher Scientific) supplemented with 10% FBS (Gibco, Sigma-Aldrich) and 1X AA (Gibco). All cell lines were confirmed to be free of mycoplasma contamination. HCT116 *NORAD*$^{-/-}$ cells were generated previously (Lee et al., 2016).

To generate the HCT116 CRISPRi cell line, lentivirus expressing dCas9/BFP/KRAB was produced by first seeding 6×10^5 HEK293T cells per well in a six-well plate. The following day, cells were transfected with 1.4 µg of pHR-SFFV-dCas9-BFP-KRAB (Addgene plasmid #46911), 0.84 µg of psPAX2 (Addgene plasmid #12260), 0.56 µg of pMD2.G (Addgene plasmid #12259), 8.4 µl of FuGENE HD (Promega), and 165 µl Opti-MEM (Thermo Fisher) according to the manufacturer's instructions. Medium was changed the next day. Two days after transfection, medium was collected and passed through a 0.45 µm SFCA sterile filter. Recipient HCT116 cells were transduced overnight using medium supplemented with 8 µg/ml polybrene (EMD Millipore). Cells expressing BFP were enriched by FACS and single-cell clonal lines were derived.

RNA Fluorescent *in situ* hybridization (RNA FISH)

RNA FISH was performed as described previously (Mito et al., 2016) with the following modifications. DIG-labeled RNA probes for human *NORAD* were synthesized by in vitro transcription using a DIG-labeling mix (Roche). Primers used for amplification of the DNA template for each probe are provided in Supplementary File 1. 2 x 10^5 cells were grown on poly-L-lysine coated coverslips for 24 to 36 hours. For RNA FISH experiments with DNA damage treatment, cells were grown for 24 hours and treated with either 1 µM doxorubicin or 200 nM

camptothecin for an additional 12 hours. Samples were rinsed twice in phosphate buffer saline (PBS), fixed in 4% paraformaldehyde for 10 minutes, washed again in PBS, and permeabilized in 0.5% Triton X-100 for 10 minutes. Samples were then washed twice with PBS and rinsed with DEPC-treated water prior to incubation in prehybridization buffer (50% formamide, 2X SSC, 1X Denhardt's solution, 10 mM EDTA, 0.1 mg/ml yeast tRNA, 0.01% Tween-20) for 1 hour. 10 ng/μl DIG-labeled RNA probe was diluted in hybridization buffer (prehybridization buffer with 5% dextran sulfate) and used for hybridization at 55°C for 16 to 20 hours. Following hybridization, samples were washed, treated with RNase A, and blocked using Blocking Reagent (Roche). DIG-labeled probes were detected using mouse monoclonal anti-DIG primary antibody (Roche; 1:100 dilution) and a Cy3-labeled goat anti-mouse IgG secondary antibody (Roche; 1:100 dilution). Samples were mounted using SlowFade Diamond Antifade with DAPI mounting media (Invitrogen) and imaging was performed using a Zeiss LSM700 confocal microscope. ImageJ was used for further image analysis.

Subcellular fractionation

Cells were seeded in triplicate and 1 x 10^6 cells were collected for subcellular fractionation, which was performed as previously described (Kopp et al., 2019; Lee et al., 2016). Briefly, cell pellets were lysed in RLN1 buffer (50 mM Tris-HCl pH 8.0, 140 mM NaCl, 1.5 mM $MgCl_2$, 0.5% NP-40, RNAse inhibitor) on ice for 5 minutes and centrifuged at 500g x 2 min. The supernatant containing the cytoplasmic fraction was separated from the pelleted nuclear fraction. RNA was then isolated from both fractions using the Qiagen RNAeasy kit and equal cell equivalents of nuclear and cytoplasmic RNA were used in subsequent qRT-PCR reactions. All samples were tested for *NORAD* as well as *NEAT1* (nuclear control) and *GAPDH* (cytoplasmic control). The sum of the nuclear and cytoplasmic expression level of each transcript was set to 100%, and the percentage of each transcript localized to each compartment was determined. *NEAT1* and

295 *GAPDH*, respectively, showed the expected nuclear and cytoplasmic localization in each
296 experiment, confirming successful fractionation.
297
298 **Reanalysis of *NORAD* RAP-MS data**
299 The raw mass spectra files from iTRAQ-labeled *NORAD* and *RMRP* RAP-MS experiments
300 (Munschauer et al., 2018) were downloaded from MassIVE (https://massive.ucsd.edu) using the
301 identifier: MSV000082561. Peptide identification and quantification was performed using
302 Proteome Discoverer (Thermo Fisher) with three search engines combined (Sequest HT,
303 Mascot, and MS Amanda). MS/MS spectra were searched against the human Uniprot database.
304 Search parameters included: trypsin enzyme specificity with a maximum of 2 missed cleavages
305 tolerated, False Discovery Rate (FDR) set to 0.01 at both peptide and protein level, ± 10 ppm
306 for precursor mass tolerance with a shorter window for fragment mass tolerance for the first
307 search, and carbamidomethylation of cysteine modification and iTRAQ labels on N-termini and
308 lysine residues as fixed modifications and oxidation of methionine and N-termini acetylation as
309 variable modifications. All peptide and protein identifications had scores surpassing the
310 combined search engine significance threshold for identification. Protein abundance was
311 calculated as the intensity given from precursor quantification and was then normalized to the
312 total peptide amount. To correct for total abundance differences between samples, protein and
313 peptide abundance values in each sample were corrected by a constant factor such that the end
314 total abundance was equivalent across all samples. Fold change was calculated as the $\log_2$
315 difference of average scaled protein abundance in *NORAD* samples and *RMRP* sample. For
316 statistical analysis, we used the limma package (Smyth, 2004) in R (https://www.r-project.org/)
317 to calculate the adjusted p-value using a moderated *t*-test and Benjamini Hochberg method to
318 control the FDR.
319

320 **RNA isolation and quantitative reverse transcription PCR (qRT-PCR)**

321 RNA was isolated from cells using the RNeasy Mini Kit (Qiagen), or, for RIP experiments, Trizol
322 (Invitrogen), and treated with RNase-free DNase (Qiagen). RNA was reverse transcribed with
323 PrimeScript RT-PCR mix (Clonetech), and Power SYBR Green PCR Master Mix (Applied
324 Biosystems) was used for qPCR. Biological replicates represent independently grown and
325 processed cells. Technical replicates represent multiple measurements of the same biological
326 sample. Primer sequences are provided in Supplementary file 1.

327

328 **Generation and *AAVS1* knock-in of *NORAD* constructs**

329 Full-length wild-type *NORAD* was amplified from a modified pcDNA3.1 vector containing the
330 *NORAD* cDNA (Lee et al., 2016), along with an additional 115 base pairs downstream of the
331 endogenous *NORAD* polyadenylation site. The PRE-mutant (PREmut) construct containing 18
332 PRE mutations (TGT to ACA) was synthesized by GENEWIZ. The 5′ deletion construct (Δ1-
333 898), 5′ fragment (nt 33-898), and ND4 and were amplified from the full-length *NORAD*
334 construct using primers provided in Supplementary file 1. Constructs were cloned into an
335 *AAVS1/PPP1R12C* targeting vector (*AAVS1* hPGK-PuroR-pA donor, Addgene plasmid #22072)
336 digested with KpnI and MfeI to remove the GFP cassette. These vectors, as well as a control
337 GFP vector, were then inserted into the *AAVS1* locus of HCT116 CRISPRi cells using a
338 previously described TALEN pair targeting the *AAVS1/PPP1R12C* locus (Sanjana et al., 2012)
339 (hAAVS1 1L TALEN, Addgene plasmid #35431; hAAVS1 1R TALEN, Addgene plasmid
340 #35432). Transfection of these plasmids was performed using FugeneHD (Promega) at a 1:1:8
341 ratio of L-TALEN:R-TALEN:Donor as previously described (Lee et al., 2016). 48 hours after
342 transfection, cells were selected with hygromycin (500 μg/ml) for at least 10 days prior to
343 introducing sgRNAs for CRISPRi-mediated knockdown.

344

345 **CRISPRi-mediated knockdown**

346 Single guide RNAs (sgRNAs) targeting a sequence upstream of the endogenous *NORAD*
347 transcription start site or targeting RBMX were cloned into a pU6-sgRNA EF1a-PuroR-T2A-BFP
348 vector (Addgene plasmid #60955). sgRNA sequences are provided in Supplementary File 1.
349 pU6-sgRNA vectors were then packaged into lentivirus by transfecting HEK293T cells using a
350 4:2:1 ratio of pU6-sgRNA:psPAX2:pMD2.G with FuGENE HD. Medium was changed the next
351 day. Media containing the virus was collected and filtered at 48 hours and 72 hours after
352 transfection. Virus was then diluted 1:3 with fresh media and used to transduce HCT116
353 CRISPRi cell lines overnight in a final polybrene concentration of 8 μg/ml. 48 hours after
354 transduction, selection with 1 μg/ml puromycin was initiated. For HCT116 CRISPRi cells with
355 *AAVS1/NORAD* construct insertion and sgRNA expression, cells were grown in 1 μg/ml
356 puromycin and 500 μg/ml hygromycin.

357

358 **UV crosslinking and RNA immunoprecipitation (RIP)**

359 PUM1, PUM2, and RBMX RIP experiments were performed in HCT116 CRISPRi cells stably
360 expressing *AAVS1/NORAD* constructs and depleted of endogenous *NORAD* with CRISPRi as
361 described above. 20 x 10^6 cells were washed in cold PBS and UV crosslinked on ice in a
362 Spectrolinker XL-1500 (Spectronics) at 254 nm (400 mJ/cm^2). Cells were then scraped,
363 centrifuged, snap-frozen in liquid nitrogen, and stored at -80°C. RIP was performed following a
364 modified eCLIP protocol (Van Nostrand et al., 2016) as follows: Cells were lysed in 1 mL cold
365 iCLIP lysis buffer (50mM Tris-HCl, 100mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium
366 deoxycholate, 1:200 Protease Inhibitor Cocktail III, RNAse inhibitor) for 25 minutes on ice.
367 Lysed cells were then centrifuged at 14,000 g for 15 min at 4°C and the supernatant was added

368 to pre-washed and antibody-coupled Protein G Dynabeads (Invitrogen). For each RIP, 5 μg of
369 antibody (anti-PUM1, Santa Cruz sc-135049; anti-PUM2, Santa Cruz sc-31535; anti-RBMX Cell
370 Signaling #14794; Goat IgG control, Santa Cruz sc-2028; Rabbit IgG control Cell Signaling
371 #2729) was coupled to 3.75 mg of beads at room temperature for 45 min, after which unbound
372 antibody was removed. Sample and beads were incubated at 4°C overnight. The next day,
373 beads were washed three times with 900 μL cold High Salt Wash Buffer #1 (50mM Tris-HCl, 1M
374 NaCl, 1mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate) and three times with 500
375 μL Wash Buffer #2 (20mM Tris-HCl, 10mM $MgCl_2$, 0.2% Tween-20). Beads were then
376 resuspended in 100 μL Wash Buffer #2, and 70 μL was used for RNA extraction and the
377 remainder for western blotting. Proteins were extracted by incubation in Laemmli buffer for 10
378 min at 70ºC. Antibodies used for western blotting were anti-PUM1 (ab92545, Abcam), anti-
379 PUM2 (ab92390, Abcam), and anti-RBMX (14794, Cell Signaling).
380
381 **DNA fluorescence in situ hybridization (DNA FISH)**
382 Aneuploidy in *NORAD* construct rescue experiments was assessed 18 to 21 days after
383 knockdown of endogenous *NORAD*. DNA FISH was performed as described previously (Kopp
384 et al., 2019; Lee et al., 2016). Chromosome enumeration probes for chromosome 7 (CHR7-10-
385 GR) and chromosome 20 (CHR20-10-RE) were purchased from Empire Genomics. Cells were
386 trypsinized, washed in PBS, and incubated in hypotonic 0.4% KCl solution for 5 minutes at room
387 temperature. Cells were then fixed in 3:1 methanol:glacial acetic acid and dropped onto slides.
388 DNA FISH hybridizations were performed by the Veripath Cytogenetics laboratory at UT
389 Southwestern. Slides were analyzed using an AxioObserver Z1 microscope (Zeiss). For each
390 sample, 200 nuclei were counted and aneuploidy was defined as a chromosome count that
391 differed from 2n for at least one of the two tested chromosomes. Samples were prepared and

counted in an experimenter-blinded manner. Two independent HCT116 CRISPRi cell lines stably expressing each *AAVS1* knock-in construct were generated, and each was independently tested for aneuploidy using this method.

ACKNOWLEDGEMENTS

We thank Rudolf Jaenisch, Didier Trono, Stanley Qi, Jonathan Weissman, and Feng Zhang for plasmids; Shinichi Nakagawa for technical assistance with RNA FISH; Sangeeta Patel in the Veripath Cytogenetics laboratory at UT Southwestern for DNA FISH; and Sungyul Lee, Kathryn O'Donnell, and members of the Mendell laboratory for helpful discussions and comments on the manuscript. This work was supported by grants from CPRIT (RP160249 to J.T.M.), NIH (R35CA197311 to J.T.M.; P30CA142543 to J.T.M.; and P50CA196516 to J.T.M.), and the Welch Foundation (I-1961-20180324 to J.T.M.). J.T.M. is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

M.M.E., F.K., F.R., A.T., and T.C.C. performed experiments. M.G. performed bioinformatic analyses. M.M.E. and J.T.M. wrote the manuscript.

AUTHOR INFORMATION

The authors declare no competing interests. Correspondence and requests for materials should be addressed to Joshua.Mendell@UTSouthwestern.edu.

REFERENCES

Adamson, B., Smogorzewska, A., Sigoillot, F. D., King, R. W., & Elledge, S. J. (2012). A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat Cell Biol,* 14(3):318-328. doi:10.1038/ncb2426

Bohn, J. A., Van Etten, J. L., Schagat, T. L., Bowman, B. M., McEachin, R. C., Freddolino, P. L., et al. (2018). Identification of diverse target RNAs that are functionally regulated by human Pumilio proteins. *Nucleic Acids Res,* 46(1):362-386. doi:10.1093/nar/gkx1120

Kopp, F., Elguindy, M. M., Yalvac, M. E., Zhang, H., Chen, B., Gillett, F. A., et al. (2019). PUMILIO hyperactivity drives premature aging of Norad-deficient mice. *Elife,* 8:e42650. doi:10.7554/eLife.42650

Lee, S., Kopp, F., Chang, T. C., Sataluri, A., Chen, B., Sivakumar, S., et al. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell,* 164(1-2):69-80. doi:10.1016/j.cell.2015.12.017

Miller, M. A., & Olivas, W. M. (2011). Roles of Puf proteins in mRNA degradation and translation. *Wiley Interdiscip Rev RNA,* 2(4):471-492. doi:10.1002/wrna.69

Mito, M., Kawaguchi, T., Hirose, T., & Nakagawa, S. (2016). Simultaneous multicolor detection of RNA and proteins using super-resolution microscopy. *Methods,* 98:158-165. doi:10.1016/j.ymeth.2015.11.007

Munschauer, M., Nguyen, C. T., Sirokman, K., Hartigan, C. R., Hogstrom, L., Engreitz, J. M., et al. (2018). The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature,* 561(7721):132-136. doi:10.1038/s41586-018-0453-z

Quenault, T., Lithgow, T., & Traven, A. (2011). PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol,* 21(2):104-112. doi:10.1016/j.tcb.2010.09.013

Sanjana, N. E., Cong, L., Zhou, Y., Cunniff, M. M., Feng, G., & Zhang, F. (2012). A transcription activator-like effector toolbox for genome engineering. *Nat Protoc,* 7(1):171-192. doi:10.1038/nprot.2011.431

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol,* 3:Article3. doi:10.2202/1544-6115.1027

Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., et al. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat Commun,* 7:12209. doi:10.1038/ncomms12209

Van Etten, J., Schagat, T. L., Hrit, J., Weidmann, C. A., Brumbaugh, J., Coon, J. J., et al. (2012). Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. *J Biol Chem,* 287(43):36370-36383. doi:10.1074/jbc.M112.373522

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein

452 binding sites with enhanced CLIP (eCLIP). *Nat Methods,* 13(6):508-514.
453 doi:10.1038/nmeth.3810
454

FIGURE LEGENDS

Figure 1. *NORAD* localizes predominantly to the cytoplasm. (A) RNA FISH in HCT116 cells using a panel of 11 probes tiling the entire *NORAD* transcript reveals a predominantly cytoplasmic signal that is absent in *NORAD*$^{-/-}$ cells with all probes except probe 7, which produces a nonspecific signal likely due to the presence of an Alu repeat element. *NORAD* FISH signal in red, DAPI counterstain in blue. Locations of PREs indicated by arrowheads. ND1-ND5 represent repetitive *NORAD* domains, as previously described (Lee et al., 2016). **(B)** RNA FISH image using probe 3 showing a wider field of cells. **(C)** Subcellular fractionation followed by qRT-PCR in HCT116 cells using primers located in the 3′ or 5′ end of *NORAD*, *GAPDH* (cytoplasmic control), or *NEAT1* (nuclear control). n = 3 biological replicates each with 3 technical replicates.

Figure 2. *NORAD* remains predominantly in the cytoplasm after treatment with DNA damaging agents. (A) RNA FISH in HCT116 cells using the indicated *NORAD* probes following a 12 hour treatment with doxorubicin or camptothecin. **(B)** *NORAD* RNA FISH (probe 5) after the indicated drug treatments. Images captured with identical microscope settings. **(C)** Subcellular fractionation followed by qRT-PCR in HCT116 cells after treatment with camptothecin for the indicated number of hours. n = 3 biological replicates each with 3 technical replicates. **(D)** As in (C) except cells were treated with camptothecin plus actinomycin D. n = 3 technical replicates.

Figure 3. Generation and stable expression of *NORAD* constructs. (A) Schematic depicting wild-type or mutant *NORAD* constructs. *NORAD* sequence conservation in mammals (UCSC Genome Browser Hg38 PhastCons track) highlights the strong conservation of the region of *NORAD* harboring PREs (arrowheads). PREmut contains 18 UGU to ACA mutations in PREs

(grey arrowheads); 5′ deletion (5′ del) lacks the RBMX binding site (nt 1-898) (Munschauer et al., 2018); 5′ fragment (5′ frag) spans the RBMX binding site (nt 33-898); ND4 construct represents the most conserved segment of *NORAD* (nt 2494-3156). **(B)** (Upper) Schematic depicting insertion of constructs into the *AAVS1/PPP1R12C* locus using TALENs. (Lower) qRT-PCR analysis of expression of each *NORAD* construct in HCT116 CRISPRi cells after infection with control or endogenous *NORAD*-targeting sgRNAs. Expression was normalized to endogenous *NORAD* level (represented by expression in *AAVS1*-GFP cells infected with sgControl). The data in the left graph was generated with a primer pair in ND4 that does not amplify the 5′ fragment, while the right graph used primers at the *NORAD* 5′ end. Values normalized to *GAPDH* expression. n = 3 technical replicates.

Figure 4. PUMILIO, but not RBMX, binding to *NORAD* is necessary for genome stability. **(A)** UV crosslinking and RNA immunoprecipitation (RIP) was used to assess PUM1, PUM2, and RBMX interactions with GFP mRNA or the indicated *NORAD* constructs. After knock-in of the indicated constructs to the *AAVS1* locus in HCT116 CRISPRi cells, endogenous *NORAD* was silenced with a lentivirally-expressed sgRNA. qRT-PCR was used to assess *NORAD* or *GAPDH* recovery in each RIP sample, expressed as fold-enrichment over pull-down with IgG. The data in the left graphs were generated with a primer pair in ND4 that does not amplify the 5′ fragment, while the right graphs used primers at the *NORAD* 5′ end. n = 2 biological replicates, each measured with 3 technical replicates. **(B)** Representative RNA FISH images of wild-type or mutant *NORAD* transcripts expressed from the *AAVS1* locus in HCT116 CRISPRi cells after knockdown of endogenous *NORAD*. Probe 10 was used for full-length *NORAD*, PREmut, and 5′ del constructs; probe 1 was used for 5′ frag; and probe 6 was used for ND4. **(C)** HCT116 CRISPRi cells stably expressing the indicated *AAVS1* knock-in construct were infected with

lentivirus expressing control or endogenous *NORAD*-targeting sgRNA. Aneuploidy was assayed 18-21 days later using DNA FISH for chromosome 7 and 20, and the frequency of interphase cells exhibiting a non-modal (2n) chromosome number was scored. Replicates represent two independently-derived *AAVS1* knock-in and sgRNA-infected cell lines. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001, chi-square test.

Figure 5. RBMX is not required for *NORAD* expression or localization. (A) qRT-PCR analysis of *RBMX* and *NORAD* transcript levels in HCT116 CRISPRi cells after introduction of the indicated lentivirally-expressed sgRNA with or without doxorubicin treatment (1 μM for 24 hours). Quantification relative to *GAPDH*. n = 3 technical replicates. **(B)** Subcellular fractionation and qRT-PCR of *NORAD*, *GAPDH* (cytoplasmic control), or *NEAT1* (nuclear control) following introduction of control or *RBMX*-targeting sgRNAs. n = 3 biological replicates each with 3 technical replicates.

SUPPLEMENTAL FIGURE LEGENDS

Figure 2 – figure supplement 1. *NORAD* remains predominantly cytoplasmic following doxorubicin-induced DNA damage. (A) Subcellular fractionation followed by qRT-PCR in HCT116 cells after treatment with doxorubicin for the indicated number of hours. n = 3 biological replicates each with 3 technical replicates. **(B)** As in (A) except cells were treated with doxorubicin plus actinomycin D. n = 3 technical replicates.

Figure 3 – figure supplement 1. Reanalysis of *NORAD* RAP-MS data. Analysis of previously published *NORAD* RAP-MS data (Munschauer et al., 2018) using a combined 3 search engine algorithm (MS Amanda, Sequest HT, Mascot) identifies isoforms of PUM1 (green), PUM2 (red), and RBMX (blue) as significantly enriched *NORAD* interactors. Volcano plot showing the

529 average fold-change compared to control *RMRP* pull-down and significance from two biological

530 replicates.

531

532 **Figure 4 – figure supplement 1. Representative western blots of PUM1, PUM2, and RBMX**

533 **in RIP experiments**.

Figure 1

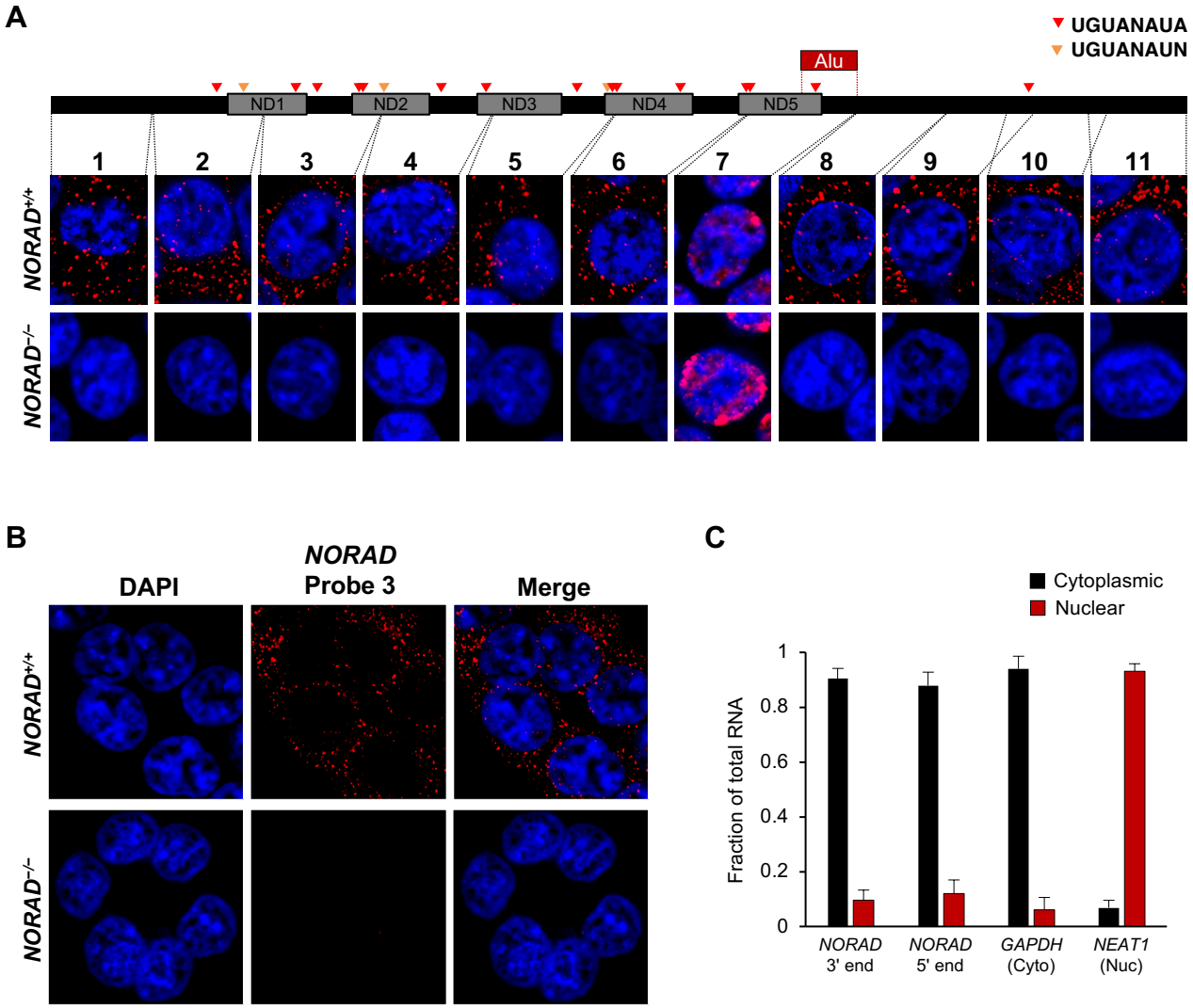


Figure 2

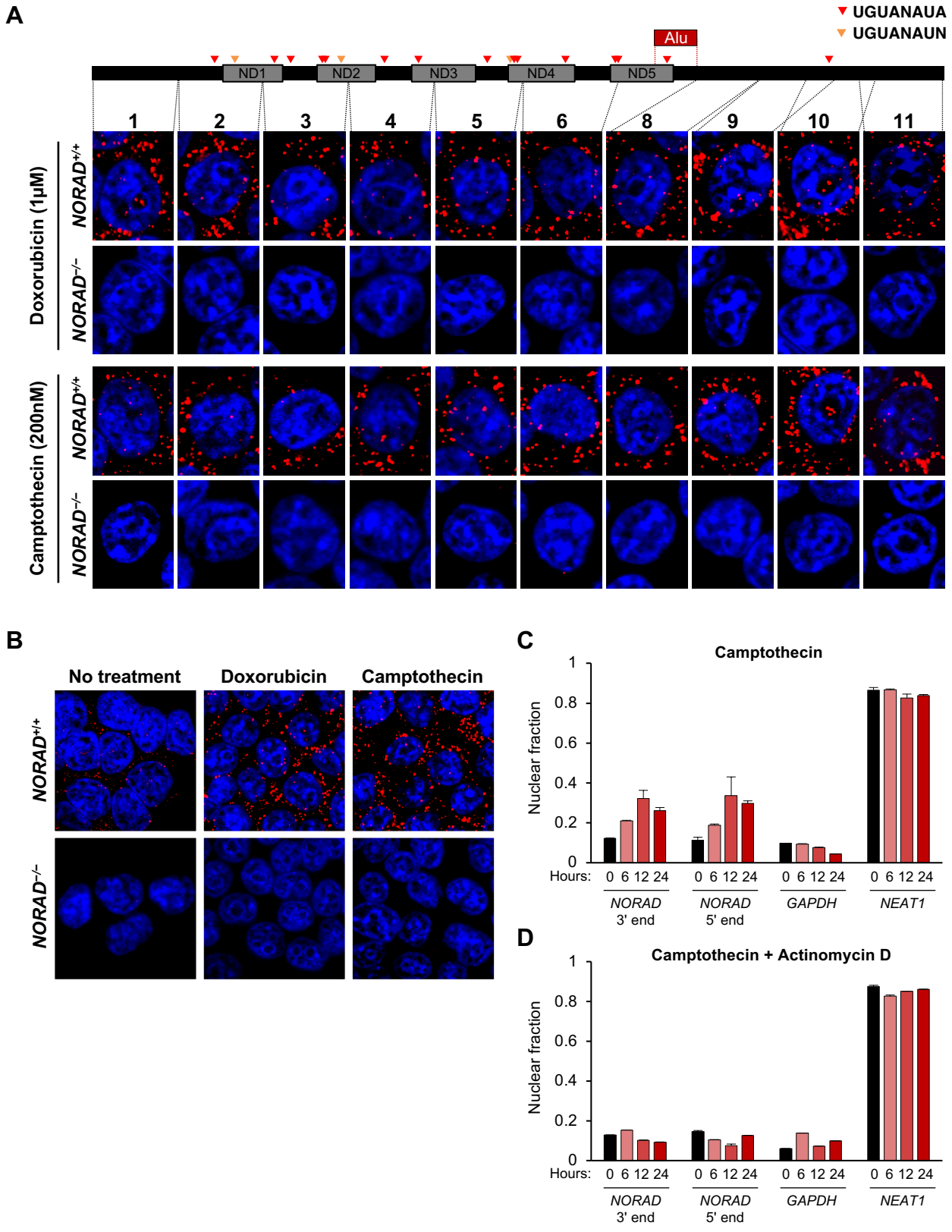


Figure 3

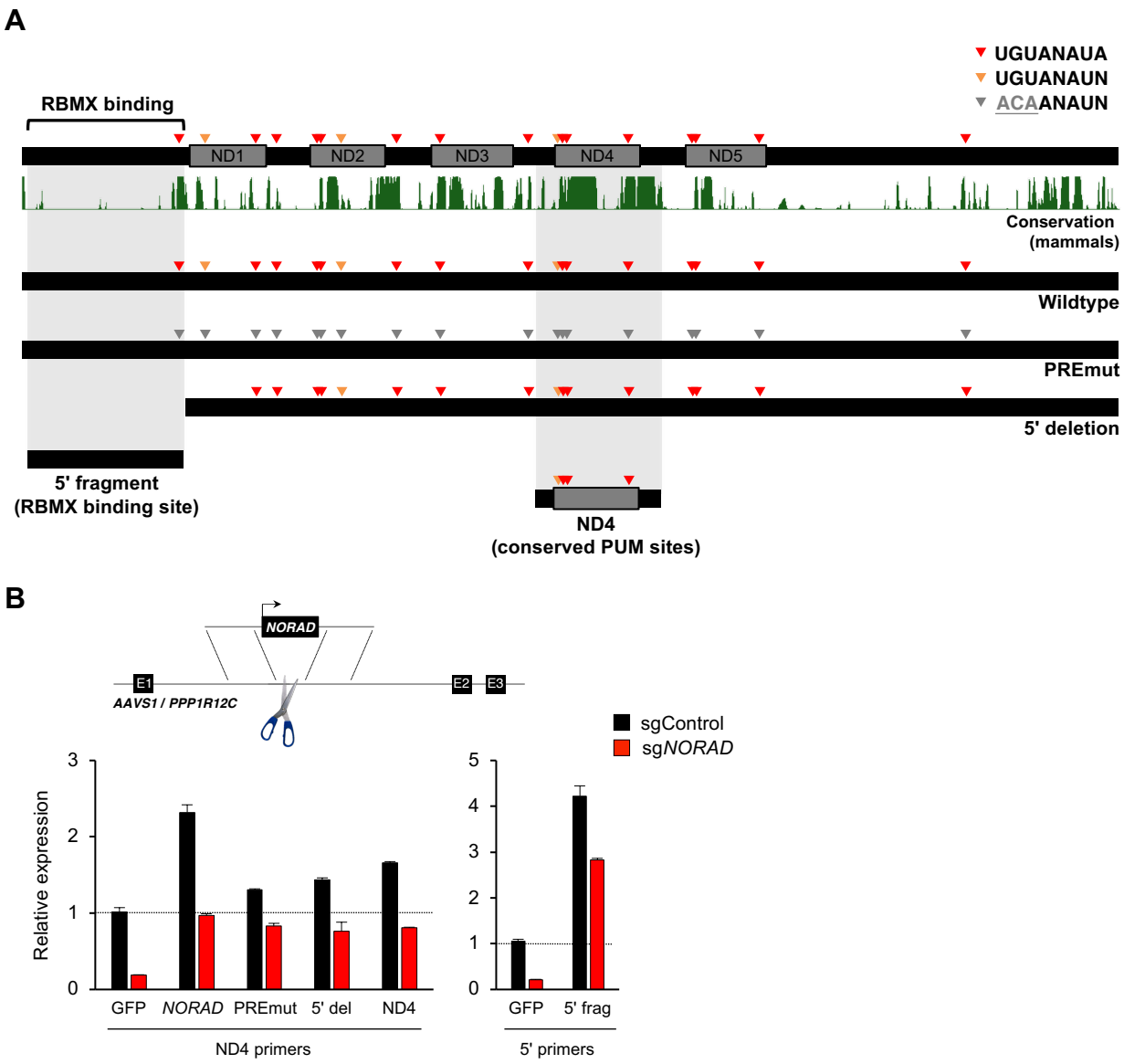


Figure 4

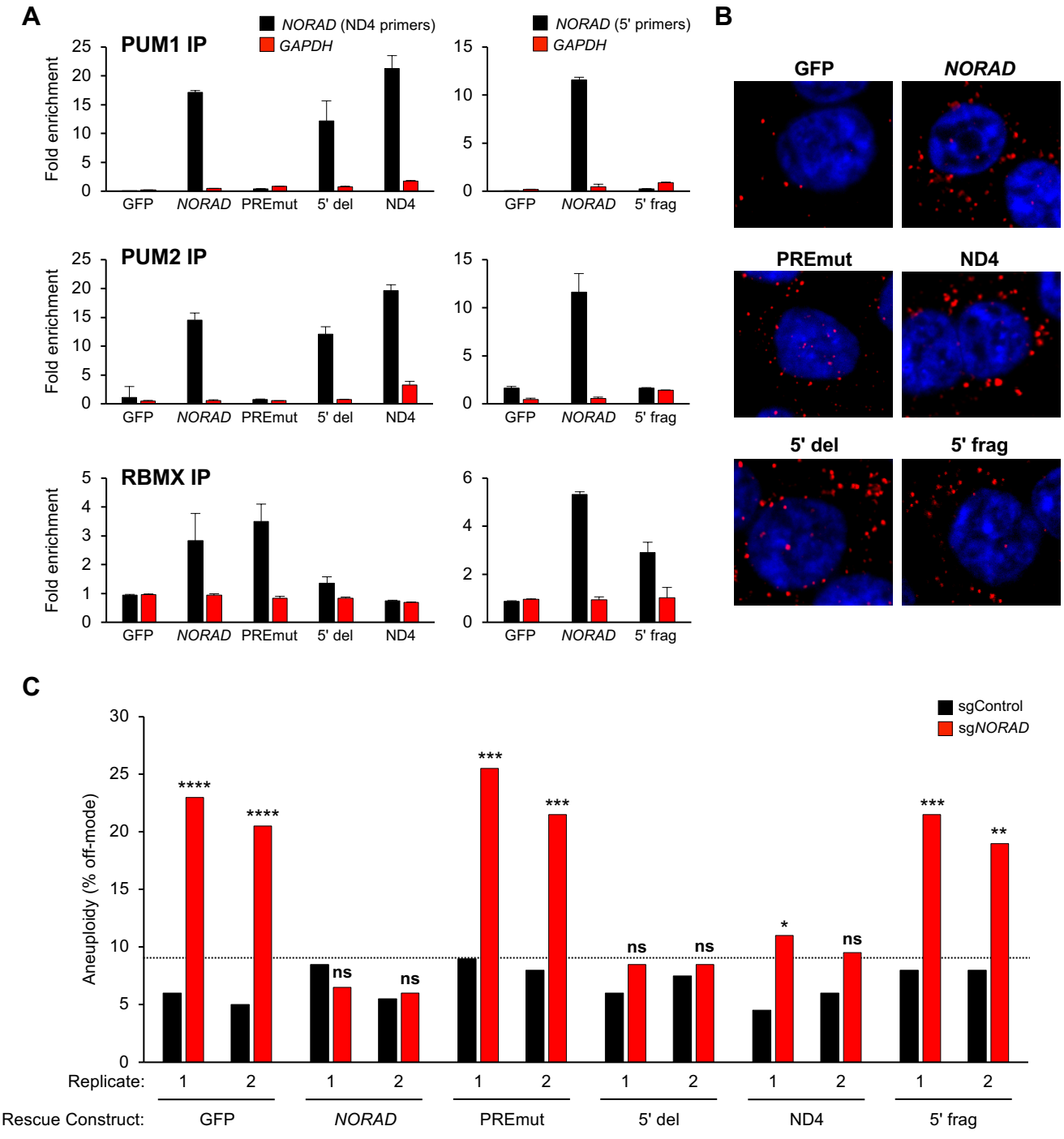


Figure 5

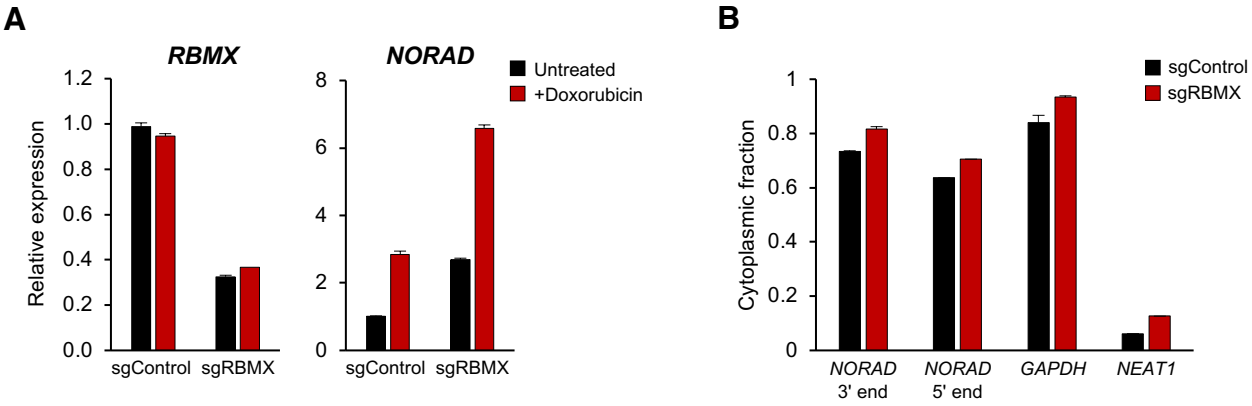


Figure 2 - figure supplement 1

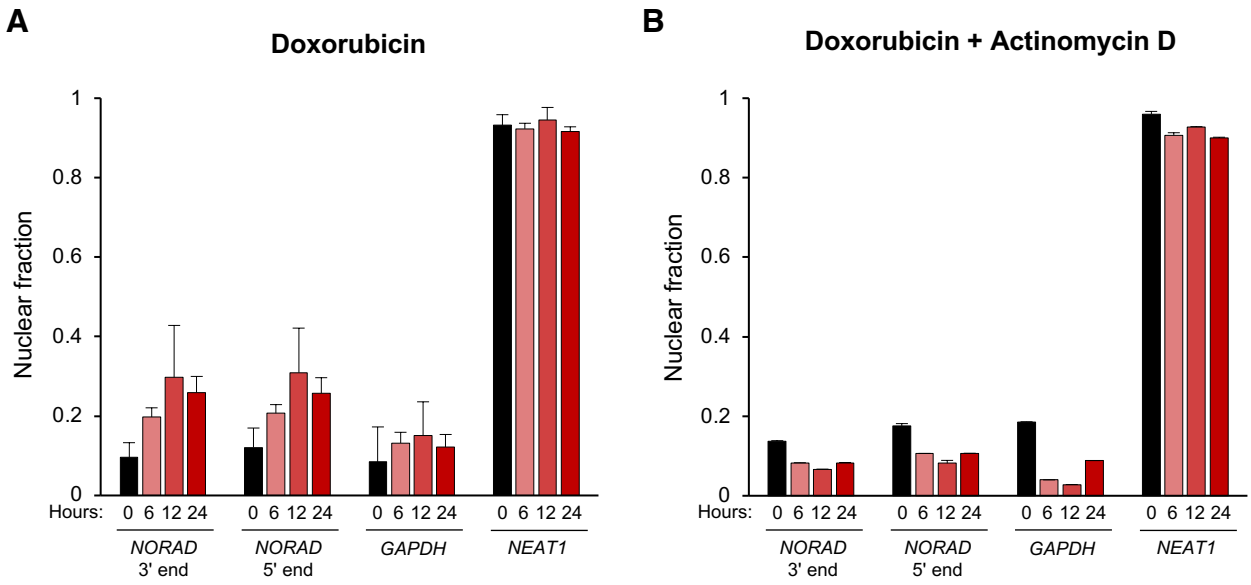


Figure 3 - figure supplement 1

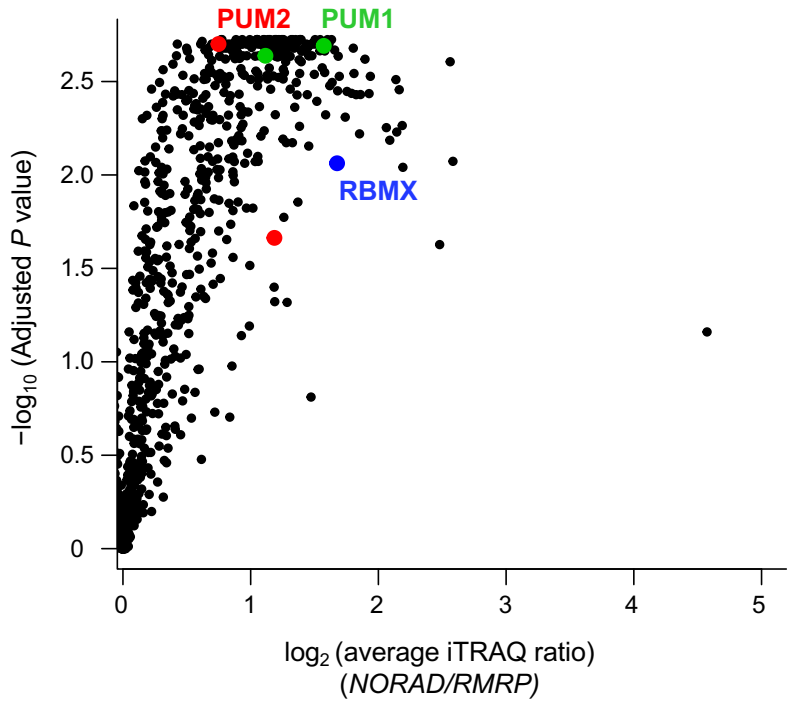
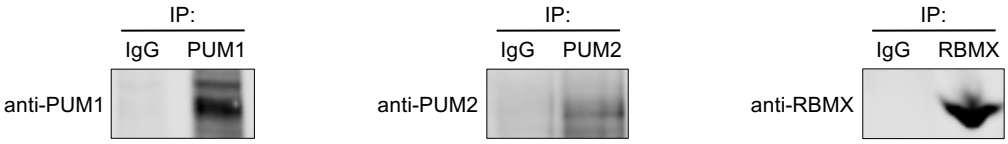


Figure 4 - figure supplement 1



Supplementary Table 1: Oligonucleotides

Oligonucleotide name/description	Sequence	Notes
NORAD FISH probe#1 fwd primer	AGTTCCGGTCCGGCAGAGAT	
NORAD FISH probe#1 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**TCGTCAGGACTAGGTAGGTC	bold represents SP6 promoter
NORAD FISH probe#2 fwd primer	CAACGGACAAAGGCCTTAAGGG	
NORAD FISH probe#2 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**GACCAGTCTAGCATAGAACCTTCTT	bold represents SP6 promoter
NORAD FISH probe#3 fwd primer	GGGTTAGATGACATGGAGCTGGAA	
NORAD FISH probe#3 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**TTCTGGCCTAGAACCCTTCCCAT	bold represents SP6 promoter
NORAD FISH probe#4 fwd primer	GGGTTCTAGGCCAGAATGTTCACA	
NORAD FISH probe#4 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**GCTTAGAGATGGTCAAACAAATTCC	bold represents SP6 promoter
NORAD FISH probe#5 fwd primer	CCATCTCTAAGCATTACACGTGCC	
NORAD FISH probe#5 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**GTCCTTTTCAGAAGACAGCCTTTCTA	bold represents SP6 promoter
NORAD FISH probe#6 fwd primer	GGCTGTCTTCTGAAAAGGACTTTTG	
NORAD FISH probe#6 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**GTAATGCTTAGGGGGGGTTTTAAC	bold represents SP6 promoter
NORAD FISH probe#7 fwd primer	CAGATGGCTTATAGCTGTCCACG	
NORAD FISH probe#7 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**TTTTGAAACGGAGCCTCGCTC	bold represents SP6 promoter
NORAD FISH probe#8 fwd primer	CCGTTTCAAAAAAAAAAGTGCACAAT	
NORAD FISH probe#8 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**GGTGGACTAATAAAGGTCACTCCC	bold represents SP6 promoter
NORAD FISH probe#9 fwd primer	CCACCCCTTGGAGCTAGACAT	
NORAD FISH probe#9 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**ACACACAGCAACAGAATACAGTATG	bold represents SP6 promoter
NORAD FISH probe#10 fwd	TGTTCATTAGGTTGGCAGCA	
NORAD FISH probe#10 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**TTCTTCTAGATCCTGTGTGTAGGC	bold represents SP6 promoter
NORAD FISH probe#11 fwd	GAGCATTAAGGGAATGCAGCAT	
NORAD FISH probe#11 rev primer (**SP6 promoter**)	gac**ATTTAGGTGACACTATAG**CAATGGAGGGAGGAGTCAAAGATG	bold represents SP6 promoter
NORAD 5' end qRT-PCR fwd primer	CTCTGCTGTGGCTGCCC	
NORAD 5' end qRT-PCR rev primer	GGGTGGGAAAGAGAGGTTCG	
NORAD 3' end qRT-PCR fwd primer	TGATAGGATACATCTTGGACATGGA	
NORAD 3' end qRT-PCR rev primer	TGGACACATCTGCATACATCTCT	
NORAD ND4 qRT-PCR fwd primer	GGAGAGGAGTTGGAAGGAATG	
NORAD ND4 qRT-PCR rev primer	TGTCCGCTATATACACAGTAGG	
GAPDH qRT-PCR fwd primer	AGCCACATCGCTCAGACAC	
GAPDH qRT-PCR rev primer	GCCCAATACGACCAAATCC	
NEAT1 qRT-PCR fwd primer	AGGCAGGGAGAGGTAGAAGG	
NEAT1 qRT-PCR rev primer	TGGCATGGACAAGTTGAAGA	
RBMX qRT-PCR fwd primer	CAGTTCGCAGTAGCAGTGGA	
RBMX qRT-PCR rev primer	TCGAGGTGGACCTCCATAAC	
NORAD full length construct fwd primer (cloning into AAVS1 plasmid)	ttcgaattctgcagtcgacggtaccAGTTCCGGTCCGGCAGAGAT	lower case sequence represents homology for HiFi assembly
NORAD full length construct rev primer (cloning into AAVS1 plasmid)	caataaacaagttaacaacaacaattgCAATGGAGGGAGGAGTCAAA	lower case sequence represents homology for HiFi assembly
NORAD 5' deletion construct fwd primer (cloning into AAVS1 plasmid)	ttcgaattctgcagtcgacggtaccATTCTCATTTGTTTAAAAGA	lower case sequence represents homology for HiFi assembly
NORAD 5' deletion construct rev primer (cloning into AAVS1 plasmid)	caataaacaagttaacaacaacaattgCAATGGAGGGAGGAGTCAAA	lower case sequence represents homology for HiFi assembly
NORAD 5' fragment construct fwd primer (cloning into AAVS1 plasmid)	ttcgaattctgcagtcgacggtaccCAGAACGCAGCCCGCTCCTC	lower case sequence represents homology for HiFi assembly
NORAD 5' fragment construct rev primer (cloning into AAVS1 plasmid)	caataaacaagttaacaacaacaattgTTACAAGATGTGTAACTTTC	lower case sequence represents homology for HiFi assembly
NORAD ND4 fragment construct fwd primer (cloning into AAVS1 plasmid)	ttcgaattctgcagtcgacggtaccAATGCTGTTTGGAAGTGGAAT	lower case sequence represents homology for HiFi assembly
NORAD ND4 fragment construct rev primer (cloning into AAVS1 plasmid)	caataaacaagttaacaacaacaattgGCACAAATATCAAAATGGGTA	lower case sequence represents homology for HiFi assembly
sgRNA *NORAD* CRISPRi knockdown	GTTCTCTGCGCTGGCAAGAG	
sgRNA *RBMX* CRISPRi knockdown	GCGCAACGAGGGCGAACAA	