CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation.
Authors	Yordanova, Martina M.;Loughran, Gary;Zhdanov, Alexander V.;Mariotti, Marco;Kiniry, Stephen J.;O'Connor, Patrick B. F.;Andreev, Dmitry E.;Tzani, Ioanna;Saffert, Paul;Michel, Audrey M.;Gladyshev, Vadim N.;Papkovsky, Dmitri B.;Atkins, John F.;Baranov, Pavel V.
Publication date	2018-01-03
Original Citation	Yordanova, M. M., Loughran, G., Zhdanov, A. V., Mariotti, M., Kiniry, S. J., O'Connor, P. B. F., Andreev, D. E., Tzani, I., Saffert, P., Michel, A. M., Gladyshev, V. N., Papkovsky, D. B., Atkins, J. F. and Baranov, P. V. (2018) 'AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation', Nature, 553, pp. 356. doi: 10.1038/nature25174
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1038/nature25174
Rights	© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Download date	2026-11-03 03:45:57
Item downloaded from	https://hdl.handle.net/10468/5463

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation.

Martina M Yordanova[1#], Gary Loughran[1#], Alexander V Zhdanov[1], Marco Mariotti[1,2], Stephen J Kiniry[1], Patrick BF O'Connor[1], Dmitry E Andreev[1,3], Ioanna Tzani[1], Paul Saffert[1], Audrey M Michel[1], Vadim N Gladyshev[2], Dmitry B Papkovsky[1], John F Atkins[1,4] & Pavel V Baranov[1]*

[1] School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland

[2] Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[3] Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia

[4]Department of Human Genetics, University of Utah, Salt Lake City, UT, USA

# These authors contributed equally to this work

*Correspondent author: p.baranov@ucc.ie

In addition to acting as template for protein synthesis, mRNA often contains sensory sequence elements that regulate this process[1,2]. Here we report a novel mechanism that limits the number of complete protein molecules that can be synthesised from a single mRNA molecule of the human *AMD1* gene encoding adenosylmethionine decarboxylase 1 (AdoMetDC). A small proportion of ribosomes translating *AMD1* mRNA stochastically read through the stop codon of the main coding region. These readthrough ribosomes then stall close to the next in-frame stop codon eventually forming a ribosome queue whose length is proportional to the number of AdoMetDC molecules that were synthesized from the same *AMD1* mRNA. Once the entire spacer region between the two stop codons is filled with queuing ribosomes, the queue impinges upon the main *AMD1* coding region halting its translation. Phylogenetic analysis suggests that this mechanism is highly conserved in vertebrates and existed in their common ancestor. We propose that this mechanism is used to count and limit the number of protein molecules that can be synthesized from a single mRNA template. It could serve to safeguard from dysregulated translation that may occur due to errors in transcription or mRNA damage.

AdoMetDC catalyses the decarboxylation of S-adenosylmethionine (AdoMet or SAM) which provides an aminopropyl group to polyamines such as spermidine and spermine[3]. SAM is an important metabolite as it serves as a major donor of methyl groups in numerous reactions involving methylation of DNA, RNA, proteins and metabolites[4-6]. Thus, in addition to its essential role in polyamine synthesis, AdoMetDC may also influence methylation reactions by affecting SAM availability. AdoMetDC is critical for embryonic stem cell self-renewal and differentiation to the neural lineage[7]. Its dysregulation is linked to tumorigenesis[8], and its

overexpression in rodent fibroblasts gives rise to aggressive transformants with extremely high invasive capacity in nude mice[9]. The synthesis of AdoMetDC is tightly controlled at the translational level allowing for quick adjustment in response to changes in polyamine concentrations. In vertebrates, control is dependent on translation of an upstream Open Reading Frame (uORF) encoding a micropeptide MAGDIS[10]. MAGDIS stalls ribosomes at the uORF stop codon and the duration of the ribosome arrest depends on the concentration of polyamines. At high concentrations, extended pausing of ribosomes at the MAGDIS uORF inhibits translation initiation at the downstream ORF encoding AdoMetDC. At lower concentrations, ribosomes terminate at the uORF stop codon and can then efficiently reinitiate translation at the AdoMetDC ORF. This mechanism provides a simple negative feedback control loop for AdoMetDC autoregulation[10].

Analysis of publicly available ribosome profiling data[11] not only confirms translation of the MAGDIS uORF but also reveals that it has the highest density of ribosome protected fragments within the *AMD1* mRNA (Fig. 1a). However, the ribosome density profile of the *AMD1* mRNA also revealed an unexpected feature: a strong isolated peak of ribosome footprint density in its 3' trailer (a.k.a. 3' UTR), 384 nucleotides downstream of the *AMD1* stop codon (Fig. 1a). In general, prominent isolated peaks of ribosome footprint density are indicative of an extended translational pause but could instead result from mRNA protection not related to genuine translation, e.g. within a nucleoprotein complex with similar sedimentation properties to that of ribosomes. The latter was recently proposed as an explanation for the peak in the *AMD1* 3' trailer[12]. The former could potentially occur if ribosomes read through the stop codon of the annotated *AMD1* coding region (CDS) and then stall downstream. Occurrences of

efficient stop codon readthrough have been observed in the decoding of many viruses, but also in cellular genes of many organisms, including humans[13-15]. There are no in-frame stop codons downstream of the *AMD1* stop until the peak of high density that occurs at the next in-frame stop codon. Thus, stop codon readthrough is a plausible explanation for the observed ribosome density peak.

Evidence supportive of functionally significant stop codon readthrough was obtained by phylogenetic analysis of the downstream ORF which we will refer to hereafter as the *AMD1 tail* and its product as the AdoMetDC extension. We initially examined a UCSC 100 species genomic alignment[16] with CodAlignView (Extended Data Fig. 1). The *AMD1 tail* ORF appeared to be conserved in the genomic sequence from ~80 tetrapods with a small variation in the naked mole rat. Given the low quality of some genomic sequences we expanded our analysis to all available vertebrate genomes and applied filtering to improve the quality (see Methods). This produced an alignment of 146-species (Extended Data Fig. 2). The origin of the *AMD1 tail* dates back to at least the root of vertebrates although it has been lost in a small group of amphibious fish species. The level of nucleotide conservation in the *AMD1 tail* is similar to that of the *AMD1* main ORF, peaking towards the middle. Analysis of the Ka/Ks ratio revealed weak purifying selection acting on amino acids encoded by the *AMD1 tail* which is strongest towards the end (last ~20 amino acids – see Fig. 1b).

Examination of all available vertebrate ribosome profiling data in GWIPS-viz browser[11] confirmed the existence of a ribosome density peak at the 3' end of the *AMD1 tail* in mouse, rat, frog and fish (Extended Data Fig. 3), strongly suggesting evolutionary conservation of, not only the *tail* ORF, but also of the potential ribosome stalling at its 3' end. The *AMD1 tail*

ribosome density peak is also present in ribosome profiling data obtained from cells treated with drugs that preferentially block ribosomes at sites of translation initiation[17,18]. This is characteristic of stalling sites and a similar ribosome peak can be observed for the well characterized stalling site at the end of the *XBP1* coding region[19] (Extended Data Fig. 4).

To determine whether the observed *AMD1 tail* peak indeed occurs as a result of ribosome arrest or because of protection by an RNA binding protein as suggested earlier[12], we tested for ribosome stalling by monitoring the formation of stable peptidyl-tRNA complexes[20] during translation of full length and truncated *AMD1 tails* (Fig. 1c). Translation of *AMD1 tails* resulted in the formation of RNase-sensitive complexes that depended upon the 63 nucleotides immediately 5' of the *AMD1 tail* stop, although the stop codon itself is not essential (Fig. 1c). Therefore, translation is stalled in a sequence-specific but stop codon-independent manner at the 3' end of the *AMD1 tail* ORF.

In an attempt to verify *AMD1* stop codon readthrough and concomitant translation of the *AMD1 tail*, we designed a construct where a HA-tag was fused to the AdoMetDC N-terminus (Fig. 2a). Stop codon readthrough would be expected to yield an additional product, longer by 128 residues. No readthrough product was detected by Western blotting using HA-tag antibodies (Fig. 2a). Surprisingly, however, we observed an almost complete loss of product in the positive control where the *AMD1* stop codon was replaced with a sense codon (UGG construct). These observations suggest that translation of the *AMD1 tail* leads to a dramatic decrease in the corresponding protein product levels. To investigate this further and to quantitatively characterise this observation, the *AMD1 tail* was cloned into a dual luciferase vector downstream of, and in-frame with, sequence encoding Renilla luciferase (R Luc).

Translation of downstream internal control Firefly luciferease (F Luc) is initiated at an introduced EMCV internal ribosome entry site (IRES) (Fig. 2b) allowing monitoring of both luciferase activities from the same mRNA. We observe a massive, about 65-fold, reduction in relative R Luc activity in those constructs where the UGA stop codon was replaced by a UGG codon (Fig. 2b). IRES-driven F Luc activity is similar from both constructs, arguing against the possibility that *tail* translation and ribosome stalling affect mRNA stability (Extended Data Fig. 5). However, to further rule out this possibility we performed RT-qPCR of the corresponding constructs and do not observe significant changes that could explain the large reduction in UGG construct product (Fig. 2b and Extended Data Fig. 5)

The reduction of protein levels when the *AMD1 tail* is translated could be explained by the effect of the extension on AdoMetDC stability, localization or by inhibition of its translation. Indeed, protein destabilization *via* readthrough extension has been reported earlier for the yeast *PDE2* mRNA[21] and was recently proposed to be a general phenomenon[22]. To explore the effect of the extension on protein stability we designed additional constructs in which the AdoMetDC extension is fused to the C-terminus of GFP (Fig. 2c). Live cell confocal imaging and Western blotting analyses demonstrated a dramatic decrease in GFP levels in cells transfected with the UGG construct (Fig. 2d-f) that cannot be explained by the 2-fold reduction in RNA levels of the UGG mRNA (Fig. 2g). The dramatic GFP reduction was also observed in constructs where UGA is replaced with sense codons other than UGG (Extended Data Fig. 6). The *AMD1 tail* inhibitory effect is consistent across several reporter constructs suggesting that it is independent of the main *AMD1* coding sequence.

Treatment of cells with the proteasome inhibitor MG132 did not restore the levels of GFP-UGG (Fig. 2d-f). Similarly, GFP-UGG levels remained low upon lysosome inhibition by dissipating their acidic pH with Concanamycin A (CMA) (Fig. 2h,i). Cell-free translation of the luciferase reporters in HEK293T lysates again resulted in significant reductions in R Luc levels from UGG mRNA compared to UGA mRNA, thus excluding the possibility that the inhibition is due to a non-canonical secretion signal in the AdoMetDC extension (Extended Data Fig. 7). Furthermore, we have excluded the possibility that the *AMD1 tail* effect may be a result of the *AMD1 tail* translation acting in *trans* (Extended Data Fig. 8)

If the AdoMetDC extension does not affect protein stability and does not serve as a secretion signal, what could be responsible for the reduction in reporter expression? To explain our observations, we propose a model illustrated in Fig. 3a. Infrequent readthrough ribosomes encounter a strong stalling site close to the *AMD1 tail* stop codon and form a stable complex with the mRNA. Trailing ribosomes that also read through the *AMD1* stop codon are stymied upstream of the stalling site and form a queue. The length of the queue is expected to be proportional to the number of AdoMetDC molecules produced from the same mRNA. Once the entire *AMD1 tail* is filled with queued ribosomes, all ribosomes translating *AMD1* would be unable to finish synthesis of AdoMetDC unless the roadblock of queueing ribosomes is cleared. Such a mechanism could be used as a safeguard against dysregulated *AMD1* mRNA molecules, e.g. those where, due to a synthesis error or damage, uORF mediated repression does not work properly. Translation of such dysregulated molecules would stop after a defined number of AdoMetDC molecules had been synthesized (Fig. 3a).

To test this hypothesis, we first took advantage of the StopGo peptide motif (a.k.a. Stop-CarryOn or 2A) which effectively results in the skipping of a peptide bond by causing release of a nascent peptide in the absence of a stop codon and then continued translation[23,24]. We fused StopGo sequences to the 3' end of R Luc sequences before the stop codon (or sense codon control) and the *AMD1 tail* (Fig. 3b). Therefore, the amino acid sequences of R Luc reporters produced from these constructs are identical irrespective of whether the *AMD1 tail* is translated. Despite this, we consistently observed a 3-fold reduction in relative R Luc activity in the construct where the wild type UGA codon was substituted with UGG (Fig. 3b). Western blotting analysis of R Luc and F Luc confirmed that the luciferase products from both constructs have identical sizes (Extended Data Fig. 9). This reduction in R Luc levels cannot be explained by an effect of the AdoMetDC extension on protein product properties (due to the presence of StopGo) or by mRNA stability, since the levels of IRES-driven F Luc observed with the UGA and UGG constructs are comparable (also confirmed by RT-qPCR, Extended Data Fig. 9c). The lower reduction in relative R Luc activities with (Fig. 3b) or without StopGo (Fig. 2b) when UGA is substituted with UGG is likely because StopGo is a slow process and it may decelerate queue formation. We also can not completely rule out the partial involvement of an uncharacterized protein degradation pathway.

According to our model, an increased readthrough efficiency should accelerate formation of the queue that results in reducing reporter expression. To test this, we measured the activity of R Luc in constructs containing the *AMD1 tail* with R Luc stop codons in contexts known to permit varying levels of readthrough[14,29]. To measure readthrough efficiencies of the different stop codon contexts, it was necessary to first eliminate the *AMD1 tail* effect by

removing the last 50 codons (Fig. 3c), which includes sequence essential for stalling (Fig. 1c). A readthrough efficiency of ~2.3% was observed for the wild type *AMD1* context (Fig. 3c) which is consistent with the low footprint density observed in ribosome profiles on the *AMD1 tail* (Fig. 1a). The readthrough levels observed at other contexts were ~5%, 9% and 13.5% (Fig. 3c). As predicted from the model in Fig. 3a, reductions in relative R Luc activity were much greater than what would be expected solely due to inactivation of readthrough products when the stalling site is present (Fig. 3d).

The scheme shown could be a simplification of the real situation. Provided that the stalled ribosomes are released with a certain rate *s*, ribosomes would accumulate in the *tail* only if $s < i/n$, where *i* is the rate of initiation and *1/n* is the probability of stop codon readthrough. In this case, the proposed mechanism would be predicted to block translation on only those mRNA molecules at which the synthesis of AdoMetDC exceeds a certain rate (i>s/n). Such regulation would be particularly effective in *AMD1* mRNA since the AdoMetDC half-life is less than 1 h[25] and its cellular concentration is largely determined by its synthesis. It is conceivable that similar mechanisms regulate expression of other genes where tight control is required. Indeed, we identify a number of genes with ribosome footprint peaks between the protein coding ORF stop and the next in-frame stop codon, see Extended Data File 3, which, intriguingly contains *EEF1A2* (see Extended Data Fig. 10 for further details). The exact function of stalling following stop codon readthrough needs to be investigated in a case by case manner, as it may vary. Queueing at the end of yeast antizyme (*oaz*) ORF, for example, has been suggested to reduce the efficiency of programmed ribosomal frameshifting in a polyamine dependent manner[26]. However, as the formation of the long queues required for both

proposed models has not been observed directly, the possibility of alternative mechanisms responsible for long range coordination between stalled ribosomes and translation far upstream on the same mRNA cannot be excluded.

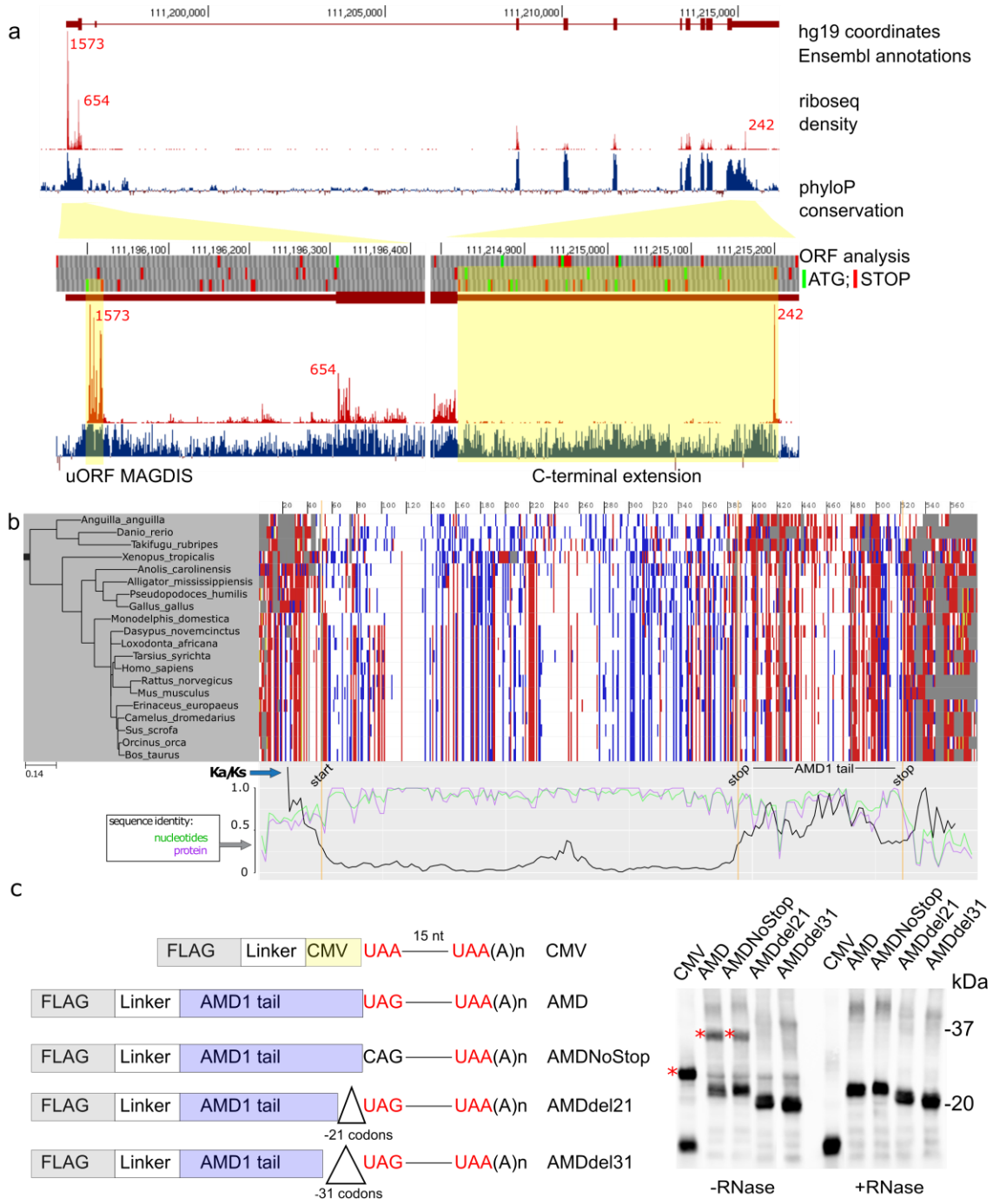


Figure 1 | Translation of phylogenetically conserved *AMD1 tail* results in ribosome stalling. **a.** A profile of ribosome footprint (riboseq) density from aggregated publicly available datasets in GWIPS-viz browser. Top plot shows all ribosome footprints aligned in the *AMD1* locus of the hg19 reference sequence assembly of the human genome. Below it are magnifications of the

areas in the vicinity of the regulatory MAGDIS uORF (left) and downstream of the *AMD1* stop codon (right). Exact number of footprints at the highest peak for each region is indicated in red. **b.** Alignment of *AMD1* sequences in a set of representative species. Synonymous and non-synonymous substitutions are shown in blue and red, respectively, and gaps are in grey. Ka/Ks ratios and sequence identities (see Methods) are shown at the bottom. **c**. Anti-Flag Western blotting of FLAG tag fusions with *AMD1 tail* and its mutants. Cytomegalovirus (CMV) gp48 uORF2 was used as a positive control. Covalent complexes of peptidyl tRNA with nascent peptide can be seen on the gel for the samples untreated with RNase.

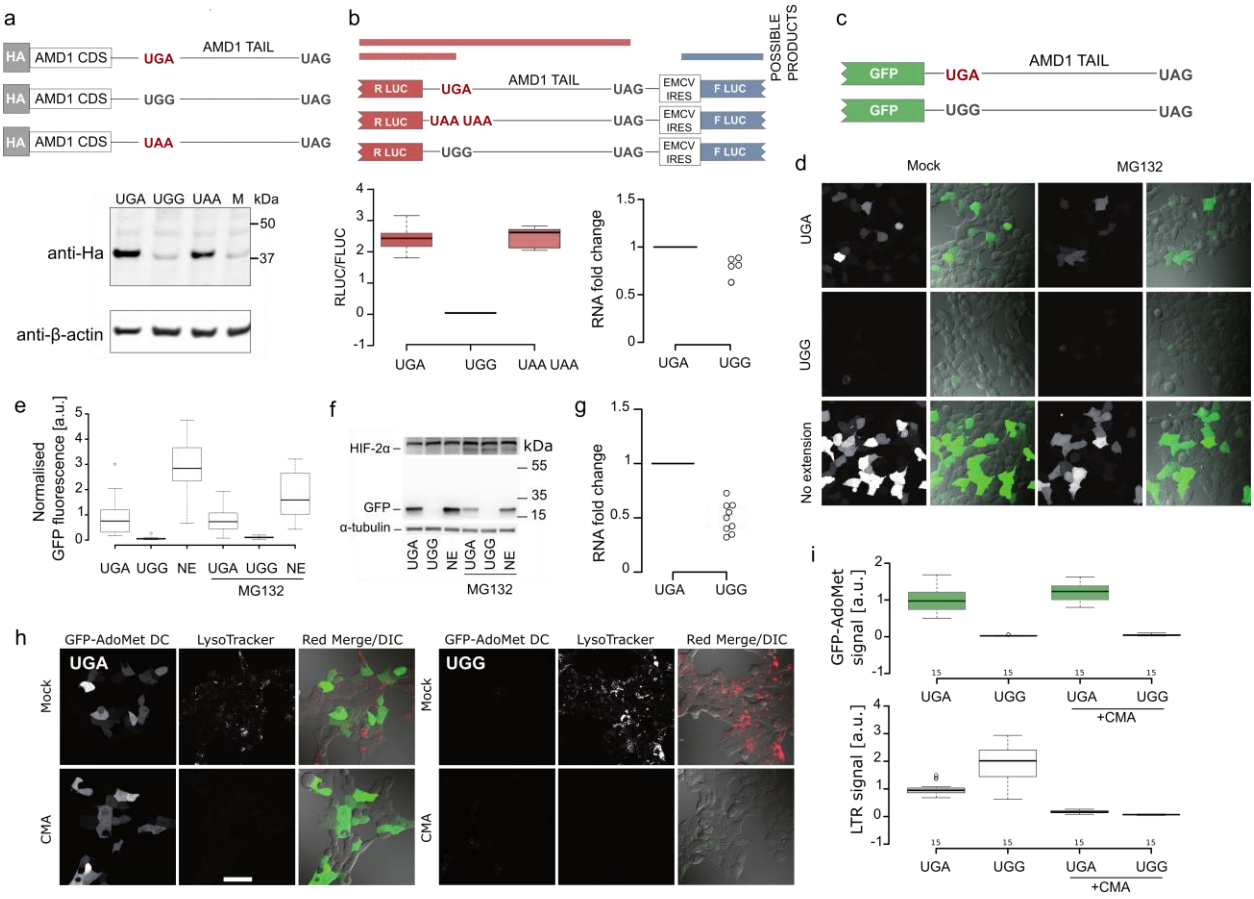


Figure 2 | *AMD1 tail* reduces expression independent of proteasomes or lysosomes pathways. a. Anti-HA immunoblots of protein lysates from HEK293T cells either mock-transfected (M) or transfected with constructs expressing HA-AMD1-UGA, HA-AMD1-UGG or HA-AMD1-UAA as indicated. **b.** Analysis of *AMD1 tail* translation effect on expression of dual reporter constructs (top); luciferase activity (bottom left), mRNA stability (bottom right). n=12 for the boxplots. **c**. GFP constructs used in (d-h). **d**. Images of live cells expressing GFP constructs treated or untreated with proteasome inhibitor MG132 (10 µM). **e.** Quantitative analysis of (**d**). **f**. Western blotting analysis of cells expressing GFP constructs, HIF-2α was used as a positive control for proteasome inhibition, NE stands for GFP expressed with no extension. **g.** RT-qPCR analysis of RNA levels for GFP constructs from (**c**). **h**. Images of live cells expressing

GFP constructs not treated or treated with the lysosome inhibitor Concanamycin A (CMA, 1µM). Acidic compartments were stained with LysoTracker Red (LTR). **i**. Quantitative analysis of (**h**) n=15 for each boxplot. Scale bar is 20 µm.

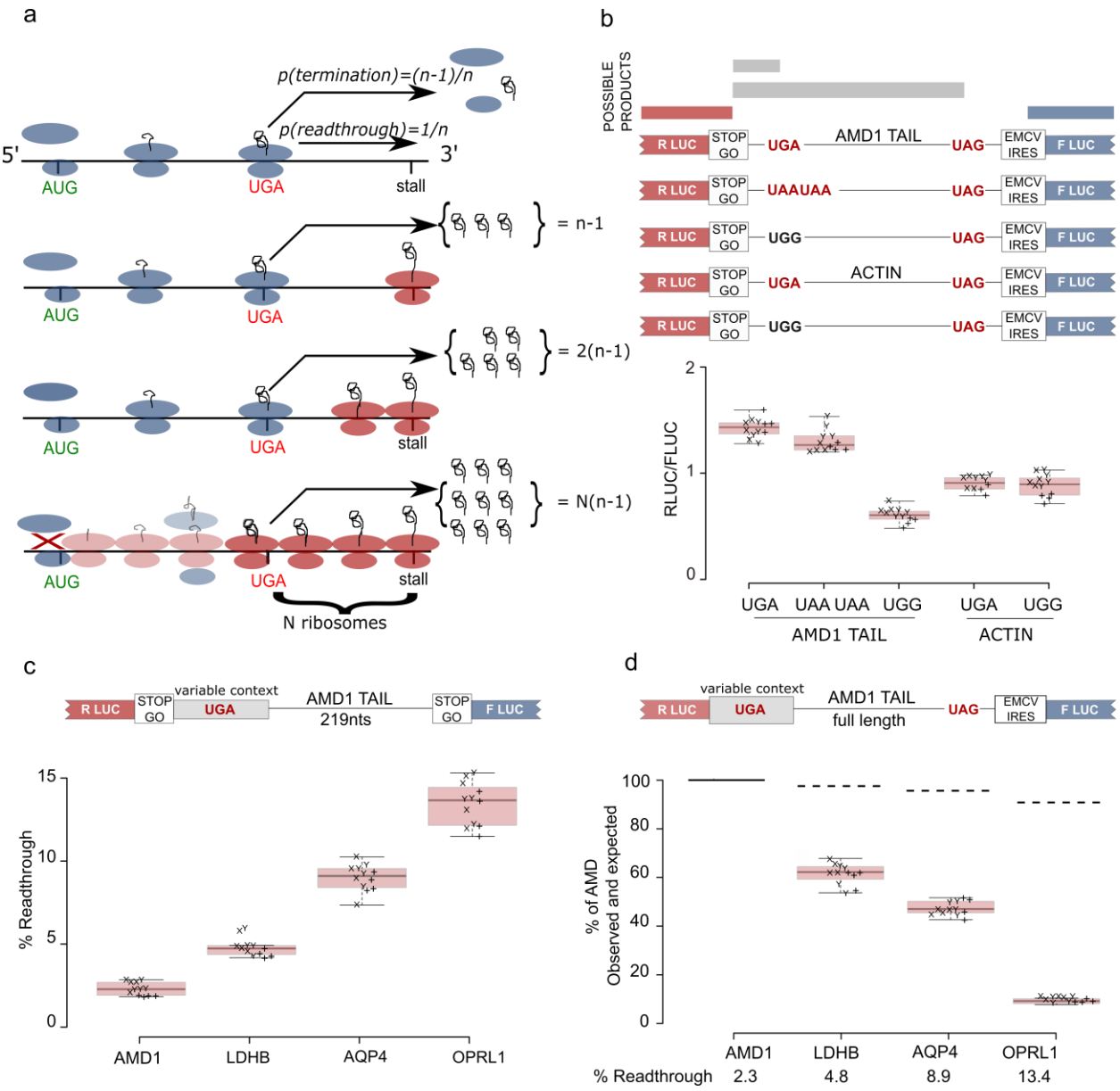


Figure 3 | *AMD1 tail* modulates translation of an upstream reporter ORF. a. Memory formation mechanism. Ribosomes read through the stop codon with frequency *1/n* and their progression is arrested at the stall site (stalled ribosomes are shown in red). The number of proteins synthesized from a single mRNA molecule is proportional to the number of ribosomes queued in the *tail*. Once the entire *tail* is filled with ribosomes (maximum *N*) translation of main ORF is impaired. This mechanism sets a limit on the number ((*n*-1)*N* on average) of AdoMetDC

molecules that can be synthesized from a single mRNA. Not drawn to scale. **b.** Analysis of *AMD1 tail* translation effect on upstream reporter translation in dual luciferase reporters containing the StopGo motif. **c**. Readthrough efficiencies determined by dual luciferase assay after transfection of HEK293T cells with indicated constructs. Normalized (Renilla/Firefly) luciferase activities were calculated for each as a percentage of the normalized UGG in-frame control. All constructs contain the first 219 nts of the *AMD1 tail*. **d**. The effect of stop codon readthrough context on translation of the upstream reporter. Broken lines indicate R Luc activities that would be expected if the extended proteins were immediately degraded. All data points are indicated as boxplots with the same symbols used for technical replicates that belong to the same biological replicate.

METHODS

Cloning

Oligonucleotides were synthesized by IDT, Belgium. Primer sequences are listed in Extended Data Table 1. The sequence of the *AMD1* coding region and the part of the 3' trailer encoding the tail was obtained as a gBlock from IDT and its sequence is provided in Extended Data File 1. The amplicons were generated by standard one or multiple PCR reactions. Plasmids used in this study include pEGFP-C1 (Clontech), pcDNA3-HA (Invitrogen), pDluc[27,28] and pSDLuc[29]. pDluc was modified such that the second luciferase reporter (firefly) is expressed under the control of the EMCV IRES. For StopGo constructs, the StopGo sequence[24] was inserted in place of the Renilla stop codon. For readthrough measurements pSGDluc[29] was used. Construct sequences are provided in Extended Data File 1. All constructs were transformed in *E. coli* strain DH5-α and were verified by sequencing.

Tissue culture and cell treatment

Human Embryonic Kidney 293T cells (ATCC) were maintained as monolayer cultures, grown in DMEM (Sigma-Aldrich) supplemented with 10% FBS, 1mM L-glutamine and antibiotics at 37°C in an atmosphere of 5% CO_2. For dual luciferase assay $4x10^6$ HEK293T cells were plated on 10 cm tissue culture dishes. After 24 hours the cells were detached with trypsin, suspended in fresh media and transfected in triplicate with Lipofectamine 2000 reagent (Invitrogen), using the 1-day protocol in which suspended cells are added directly to the DNA complexes in 96-well plates. For each transfection the following was added to each well: 50 ng plasmid DNA, 0.4 µl lipofectamine 2000 in 50 µl OptiMem (Gibco). $8x10^4$ cells in 150 µl DMEM, were added to the

transfecting DNA complexes in each well. Transfected cells were incubated at 37°C in 5% CO2 for 21 h and assayed using the dual luciferase assay.

Transfections for Western blotting of EGFP encoding constructs for Extended Data Fig. 6 and for HA-AMD1 and RT-qPCR analysis were performed in 6 well plates scaled-up from the method described for 96 well plate transfections above. The following was added to each well: 1ug plasmid DNA, 7µl lipofectamine 2000 in 1 ml OptiMem. $1x10^6$ cells in 3 ml DMEM, were added to the transfecting DNA complexes in each well. Transfected cells were incubated at 37° C in 5% CO_2 for 36 h for Western blotting and 21 h followed by RNA extraction for RT-qPCR.

For Western blotting analysis of plasmids encoding EGFP (Fig. 2f), cells were seeded at 1×10^6 cells per well on 12-well plates and grown for 16 h before transfection. For confocal fluorescence microscopy, cells were seeded at 2×10^4 on MatTek glass bottom dishes pre-coated with collagen IV / poly-D-lysine. Loading of the cells with fluorescent indicator LysoTracker Red DND-99 (Invitrogen) (100 nM) was performed in OptiMEM medium for 30 min. in CO_2 incubator. For proteasome inhibition, 10µM MG132 (Sigma-Aldrich) was added to transfected cells for 5 h prior to microscopy imaging or cell lysis and Western blotting analysis. Concanamycin A (Sigma-Aldrich) treatment (1µM) was performed for 5 h prior to microscopy imaging.

Confocal microscopy

Live cell imaging was conducted on an Olympus FV1000 confocal laser scanning microscope with controlled CO_2, humidity and temperature. EGFP was excited at 488 nm (2.5-10% of laser power) with emission collected at 500-540 nm. LysoTracker Red was excited at 543

nm (15% of laser power); emission was collected at 560-600 nm. Acquisition of each spectral signal was done in sequential laser mode. Fluorescence and differential interference contrast (DIC) images were collected with a 60X oil immersion objective in 12 planes using 0.5 µm steps. The resulting single images were analysed using FV1000 Viewer (Olympus), Adobe Photoshop and Illustrator software.

Protein Isolation and Western Blot Analysis

For Western blotting analysis of plasmids encoding EGFP for Fig. 2f, whole cell lysates were prepared in a standard RIPA buffer (Thermo Fisher) containing protease and phosphatase inhibitors. After lysate clarification, protein concentration was measured using BCA™ Protein Assay kit (Thermo Fisher) and equalised. Proteins were separated by 4-20% polyacrylamide gel electrophoresis on pre-made acrylamide gels (GenScript), transferred onto a 0.2 µm Immobilon™-P PVDF membrane (Sigma-Aldrich) using wet mini-transfer system Hoefer™ TE 22 (Hoefer) and probed with antibodies against HIF-2α (R&D systems), α-tubulin (Sigma-Aldrich) and GFP (Novex) in 5% fat-free milk in TBST (0.8% Tween-20) overnight at 4°C (primary) and for 2 h at room temperature (secondary). Immunoblots were analysed with HRP-conjugated secondary antibodies (Sigma-Aldrich) and Amersham™ ECL™ Prime reagents using the LAS-3000 Imager (Fujifilm) and Image Reader LAS-3000 2.2 software. Quantitative analysis was conducted using ImageJ program.

For EGFP encoding plasmids for Extended Data Fig 6, HA and luciferase encoding plasmids, cells were washed with 1x PBS and lysed in 1x PLB (Passive Lysis Buffer, Promega). Proteins were separated by 4-12% polyacrylamide gel electrophoresis on pre-made Bolt™ 4-

12% BisTris Plus gels (Thermo Fisher), transferred onto nitrocellulose membranes (Protran) and incubated with primary antibodies in 5% fat-free milk in PBST (1% Tween-20) overnight at 4°C. Primary antibodies were against HA (clone 16B12: Covance), Renilla (MBL), firefly (Promega) and GFP (Santa Cruz). Incubation with fluorescently labelled secondary antibodies was for 0.5 h at room temperature.

Dual luciferase assay

Firefly and Renilla luciferase assay buffers were prepared as described in Dyer *et al*[30]. Relative light units were measured on a Veritas Microplate Luminometer fitted with two injectors (Turner Biosystems). Cells transfected in 96 well plate were washed once with 1× PBS and then lysed in 18 µl of 1× passive lysis buffer (PLB; Promega), and light emission was measured following injection of 50µl of each luciferase substrate buffer.

RNA extraction and qPCR

For cells transfected with dual luciferase constructs, RNA was extracted using TRIzol Reagent® (Ambion) according to manufacturer's protocol followed by precipitation with isopropanol. 200 ng RNA were DNase treated (RQ1-DNase, Promega) and reverse transcribed with oligo dT-Primer and Superscript III (Thermo Fischer). Quantitative Real-time PCR (qRT-PCR) was performed in 10µl Reactions using the QuantiFast SYBR Green (Qiagen). The primers used are listed in the table S1. RNA levels of test constructs were normalised to beta-Actin mRNA levels.

For EGFP encoding constructs, cells were washed with 1 x PBS and lysed with 1 x PLB (Promega). RNA was extracted using the phenol-chlorophorm method and precipitated with

isopropanol. 1µg RNA was DNase treated (RQ1-DNase, Promega). 200ng RNA were reverse transcribed with random oligo (IDT) and Superscript III (Thermo Fischer) in 20µl reactions according to the manufacturer's recommendations. Quantitative Real-time PCR (qRT-PCR) was performed in 10µl reactions using the QuantiFast SYBR Green (Qiagen). The primers used are listed in the Extended Data Table 1. RNA levels of test constructs were normalised to the levels of the co-transfected control construct. The experiments were performed in two or three biological replicas each with two or three technical replicas at the stage of the RT-qPCR reaction. The fold difference was calculated by the Comparative CT Method (ΔΔCT Method)[31]. All data points were plotted.

***In vitro* transcription**

mRNA for in vitro translation was produced with T7 RiboMAX™ Express Large Scale RNA Production System (Promega) following the manufacturer's instructions with a PCR product serving as a template (primers used to generate the PCR templates are listed in extended data table 1). For translation in HEK293T cell free system mRNA was capped using the Vaccinia Capping System (NEB). For translation in RRL mRNA was not capped.

***In vitro* translation in HEK293T cell free system**

HEK293T cell-free translation system was prepared as described in Terenin *et al*[32]. Briefly, HEK293T cells at ~ 75% of confluence were quickly harvested on ice and resuspended in Lysolecitin lysis buffer (20 mM HEPES-KOH pH 7.4, 100 mM KOAc ,2.2 mM $Mg(OAc)_2$, 2 mM DTT, 0.1mg/ml Lysolecitin). The cells were then spun down and resuspended in Hypotonic extraction

buffer (20mM HEPES pH 7.5, 10mM KOAc, 1mM Mg(OAc)2, 4mM DTT, Complete Protease Inhibitor Cocktail (EDTA-free; Roche)). The cells were then disrupted in a pre-cooled 2ml Dounce homogenizer. The lysates were collected following centrifugation for 10 min. at 10000 g. 10µl in vitro translation reactions were assembled in the presence of 50% v/v HEK293T cell-free lysate, 1 x translation buffer (20mM Hepes–KOH pH 7.5, 1mM DTT, 0.5mM spermidine–HCl, 0.6mM Mg (OAc)2, 8mM creatine phosphate, 1mM ATP, 0.2mM GTP, 120mM KOAc and 25µM of each amino acid) and 200ng capped mRNA, as described in Andreev *et al*[33].

***In vitro* translation in RRL**

RNA was translated in 10µl *in vitro* reactions using Flexi® Rabbit Reticulocyte Lysate System (Promega) according to the manufacturer's recommendations.

Monitoring of peptidyl-tRNA complexes

Translation reactions in RRL were carried out at 30° C for 1 h and then placed on ice. Half of the reaction (5µl) was then subjected to RNaseA treatment for 20 min. on ice. 2.5µl of RNase treated and 2µl of untreated samples were combined with 2 x Sample buffer supplemented with RNase inhibitor and loaded onto NuPage Bis-Tris neutral Gels (Thermo Fisher). The neutral pH prevents hydrolysis of peptidyl-tRNA bond. The products were detected with an ANTI-FLAG antibody (F1804, Sigma). The sequence of the human cytomegalovirus (CMV) gp48 uORF2 was used as a positive control for co-translational ribosome stalling[34,35].

Gene finding and evolutionary analysis

All publicly available genome sequences of vertebrates were downloaded from NCBI. We then used Selenoprofiles[36] to identify *AMD1* orthologs using a manually curated protein profile alignment that spans the main ORF. Gene structures were completed by extending homologous coding regions to the upstream methionine and first in-frame downstream stop. Results were filtered to exclude retrotransposed pseudogenes, abundant in mammals and recognizable for their lack of introns. The predictions were further filtered through manual inspection to obtain a *bona fide* set of 146 complete gene sequences across vertebrates with clear orthology. These gene structures were then extended by 120 nt upstream and 510 nt downstream, in order to include a similar length of non-coding sequence at each side (considering the *AMD1 tail* downstream of *AMD1* main ORF stop). The phylogenetic tree of the investigated species was extracted from NCBI taxonomy[37] and standardized arbitrarily to a dichotomic tree with ETE 3[38]. Evolutionary analysis was then performed on the resulting alignment and tree using pycodeml, available at https://github.com/marco-mariotti/pycodeml. The rate of nonsynonymous vs synonymous substitutions (Ka/Ks) was computed with codeml[39] using a fixed rate. This metric was computed in sliding windows throughout the alignment, each 30 codons wide and with a 3 codon step. Sequence identity at nucleotide and protein level was also computed on the same alignment in sliding windows, each 3 codons wide and with a 3 codon step.

Codon alignments for Extended Data Figs. 1 and 10b were produced using CodAlignView ("CodAlignView: a tool for visualizing protein-coding constraint", I Jungreis, M Lin, M Kellis, in preparation.)

Identification of transcripts with ribosome footprint density profiles similar to *AMD1* mRNA

For each protein coding transcript in the Gencode v. 22[40], the genomic coordinates of the region between the annotated CDS stop and the next in frame stop were extracted. Any nucleotide position within this region which overlapped with another annotated coding region in the Gencode annotations was discarded. The number of footprints at each remaining position was extracted from the global aggregate track of ribosome profiling data (hg38 assembly and hg18 assembly with coordinates lifted over) in GWIPS-viz[11]. The positions with the highest number of footprints were recorded for each transcript. Any transcript with less than 500 footprints was discarded.

Code availability

All custom computer codes used in this study are available upon request.

Data availability

All data generated during this study are included in this published article (and its supplementary information files). In addition, publicly available data have been analysed in this study. Alignments of ribosome profiling data were obtained through GWIPS-viz (http://gwips.ucc.ie), genomic sequences have been obtained through GenBank and all relevant

sequences are provided in supplementary files.

References

1 Gebauer, F. & Hentze, M. W. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* **5**, 827-835, doi:10.1038/nrm1488 (2004).

2 Sonenberg, N. & Hinnebusch, A. G. New modes of translational control in development, behavior, and disease. *Mol Cell* **28**, 721-729, doi:10.1016/j.molcel.2007.11.018 (2007).

3 Pegg, A. E. S-Adenosylmethionine decarboxylase. *Essays Biochem* **46**, 25-45, doi:10.1042/bse0460003 (2009).

4 Chiang, P. K. *et al.* S-Adenosylmethionine and methylation. *FASEB J* **10**, 471-480 (1996).

5 Lu, S. C. & Mato, J. M. S-adenosylmethionine in liver health, injury, and cancer. *Physiol Rev* **92**, 1515-1542, doi:10.1152/physrev.00047.2011 (2012).

6 Roje, S. S-Adenosyl-L-methionine: beyond the universal methyl group donor. *Phytochemistry* **67**, 1686-1698, doi:10.1016/j.phytochem.2006.04.019 (2006).

7 Zhang, D. *et al.* AMD1 is essential for ESC self-renewal and is translationally down-regulated on differentiation to neural precursor cells. *Genes Dev* **26**, 461-473, doi:10.1101/gad.182998.111 (2012).

8 Scuoppo, C. *et al.* A tumour suppressor network relying on the polyamine-hypusine axis. *Nature* **487**, 244-248, doi:10.1038/nature11126 (2012).

9 Paasinen-Sohns, A. *et al.* Chaotic neovascularization induced by aggressive fibrosarcoma cells overexpressing S-adenosylmethionine decarboxylase. *Int J Biochem Cell Biol* **43**, 441-454, doi:10.1016/j.biocel.2010.11.018 (2011).

10 Law, G. L., Raney, A., Heusner, C. & Morris, D. R. Polyamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase. *J Biol Chem* **276**, 38036-38043, doi:10.1074/jbc.M105944200 (2001).

11 Michel, A. M. *et al.* GWIPS-viz: development of a ribo-seq genome browser. *Nucleic acids research* **42**, D859-864, doi:10.1093/nar/gkt1035 (2014).

12 Ji, Z., Song, R., Huang, H., Regev, A. & Struhl, K. Transcriptome-scale RNase-footprinting of RNA-protein complexes. *Nat Biotechnol* **34**, 410-413, doi:10.1038/nbt.3441 (2016).

13 Schueren, F. *et al.* Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *Elife* **3**, e03640, doi:10.7554/eLife.03640 (2014).

14 Loughran, G. *et al.* Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic acids research* **42**, 8928-8938, doi:10.1093/nar/gku608 (2014).

15 Stiebler, A. C. *et al.* Ribosomal readthrough at a short UGA stop codon context triggers dual localization of metabolic enzymes in Fungi and animals. *PLoS Genet* **10**, e1004685, doi:10.1371/journal.pgen.1004685 (2014).

16 Rosenbloom, K. R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic acids research* **43**, D670-681, doi:10.1093/nar/gku1177 (2015).

17 Gao, X. *et al.* Quantitative profiling of initiating ribosomes in vivo. *Nature methods* **12**, 147-153, doi:10.1038/nmeth.3208 (2015).

18 Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802, doi:10.1016/j.cell.2011.10.002 (2011).

19 Yanagitani, K. *et al.* Cotranslational targeting of XBP1 protein to the membrane promotes cytoplasmic splicing of its own mRNA. *Mol Cell* **34**, 191-200, doi:10.1016/j.molcel.2009.02.033 (2009).
20 Yanagitani, K., Kimata, Y., Kadokura, H. & Kohno, K. Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. *Science* **331**, 586-589, doi:10.1126/science.1197142 (2011).
21 Namy, O., Duchateau-Nguyen, G. & Rousset, J. P. Translational readthrough of the PDE2 stop codon modulates cAMP levels in Saccharomyces cerevisiae. *Mol Microbiol* **43**, 641-652 (2002).
22 Arribere, J. A. *et al.* Translation readthrough mitigation. *Nature* **534**, 719-723, doi:10.1038/nature18308 (2016).
23 Doronina, V. A. *et al.* Site-specific release of nascent chains from ribosomes at a sense codon. *Mol Cell Biol* **28**, 4227-4239, doi:10.1128/MCB.00421-08 (2008).
24 Ryan, M. D. & Drew, J. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J* **13**, 928-933 (1994).
25 Miller-Fleming, L., Olin-Sandoval, V., Campbell, K. & Ralser, M. Remaining Mysteries of Molecular Biology: The Role of Polyamines in the Cell. *J Mol Biol* **427**, 3389-3406, doi:10.1016/j.jmb.2015.06.020 (2015).
26 Kurian, L., Palanimurugan, R., Godderz, D. & Dohmen, R. J. Polyamine sensing by nascent ornithine decarboxylase antizyme stimulates decoding of its mRNA. *Nature* **477**, 490-494, doi:10.1038/nature10393 (2011).
27 Fixsen, S. M. & Howard, M. T. Processive selenocysteine incorporation during synthesis of eukaryotic selenoproteins. *J Mol Biol* **399**, 385-396, doi:10.1016/j.jmb.2010.04.033 (2010).
28 Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F. & Atkins, J. F. A dual-luciferase reporter system for studying recoding signals. *RNA* **4**, 479-486 (1998).
29 Loughran, G., Howard, M. T., Firth, A. E. & Atkins, J. F. Avoidance of reporter assay distortions from fused dual reporters. *RNA*, doi:10.1261/rna.061051.117 (2017).
30 Dyer, B. W., Ferrer, F. A., Klinedinst, D. K. & Rodriguez, R. A noncommercial dual luciferase enzyme assay system for reporter gene analysis. *Anal Biochem* **282**, 158-161, doi:10.1006/abio.2000.4605 (2000).
31 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).
32 Terenin, I. M., Andreev, D. E., Dmitriev, S. E. & Shatsky, I. N. A novel mechanism of eukaryotic translation initiation that is neither m7G-cap-, nor IRES-dependent. *Nucleic acids research* **41**, 1807-1816, doi:10.1093/nar/gks1282 (2013).
33 Andreev, D. E. *et al.* Differential contribution of the m7G-cap to the 5' end-dependent translation initiation of mammalian mRNAs. *Nucleic acids research* **37**, 6135-6147, doi:10.1093/nar/gkp665 (2009).
34 Degnin, C. R., Schleiss, M. R., Cao, J. & Geballe, A. P. Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J Virol* **67**, 5514-5521 (1993).
35 Bhushan, S. *et al.* Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell* **40**, 138-146, doi:10.1016/j.molcel.2010.09.009 (2010).
36 Mariotti, M. & Guigo, R. Selenoprofiles: profile-based scanning of eukaryotic genome sequences for selenoprotein genes. *Bioinformatics* **26**, 2656-2663, doi:10.1093/bioinformatics/btq516 (2010).
37 Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic acids research* **37**, D5-15, doi:10.1093/nar/gkn741 (2009).

38 Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol*, doi:10.1093/molbev/msw046 (2016).
39 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).
40 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774, doi:10.1101/gr.135350.111 (2012).

Acknowledgments.

We are grateful to Irwin Jungreis and Manolis Kellis (MIT) for permitting us to use CodAlignView. We acknowledge financial support from Science Foundation Ireland ([12/IA/1335] to PVB, [13/IA/1853] to JFA and [12/RC/2276] to DPB, Health Research Board [PhD/2007/04] to IT and Russian Science Foundation [16-14-10065] to DEA.

Author Contributions.

PBFO'C made initial observation of unusual ribosome footprint density. MM and PVB carried out phylogenetic analysis. MMY, GL, AVZ, IT, PS and DEA designed and carried out biochemical experiments. SJK and AMM carried out analysis of publicly available ribosome profiling data. PVB proposed the model. MMY and PVB drafted the manuscript. All authors contributed to interpretation of the experimental data and editing of the manuscript.

Author Information.

The authors declare no competing financial interests. Correspondence should be addressed to P.V.B (p.baranov@ucc.ie)

EXTENDED DATA

Note: Extended Data Figures 3-10 are available below, all other Extended Data items are submitted in separate files.

Extended Data File 1 | Sequences of vectors and plasmids used in this study in fasta format.

Extended Data File 2 | Genomic sequences of *AMD1* coding regions and surrounding areas in fasta format.

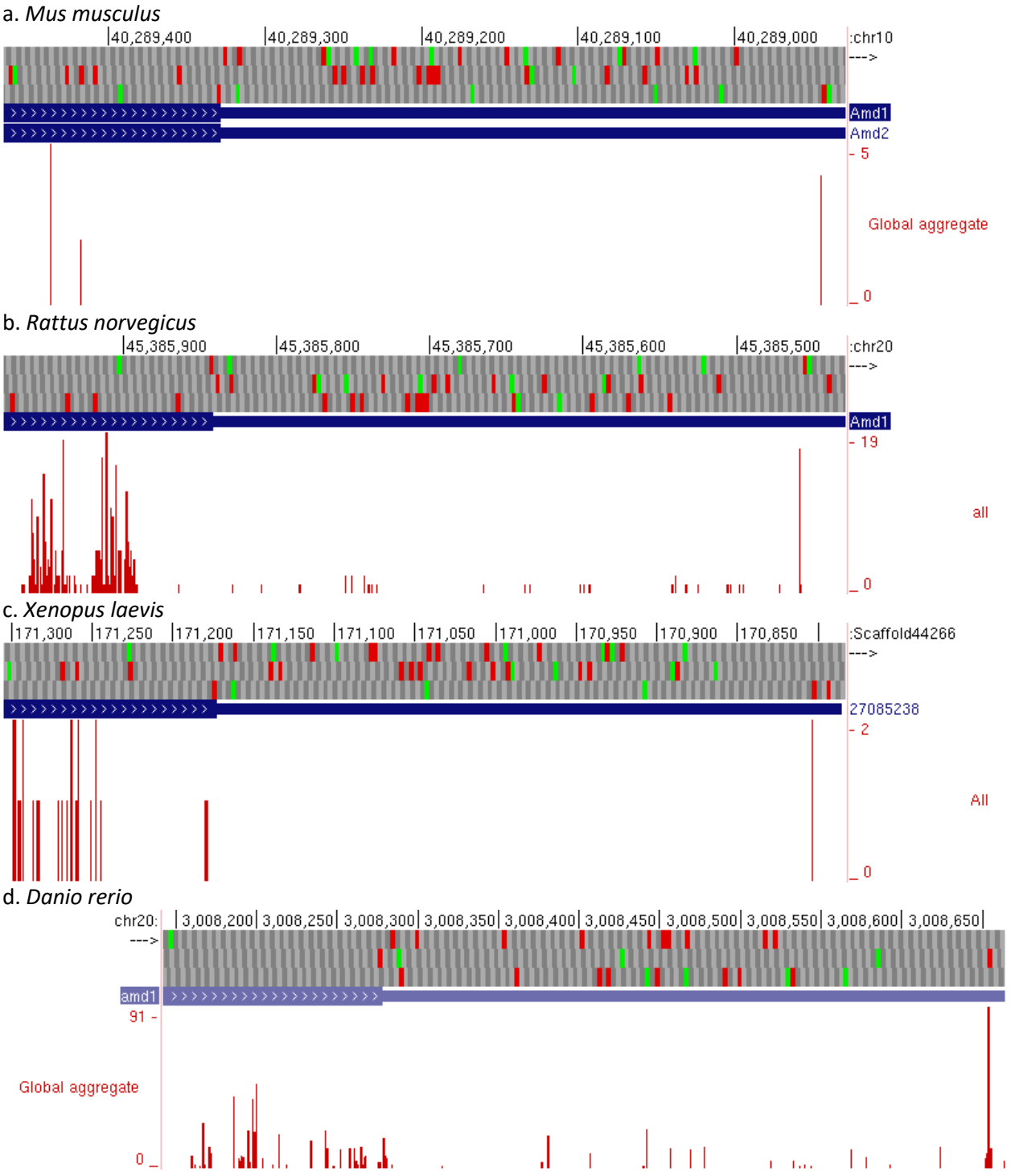
Extended Data File 3 | List of GENCODE transcripts containing peaks of ribosome density downstream and in-frame of protein coding regions.

Extended Table 1. Sequences of DNA primers used in this study.

Figure Legends

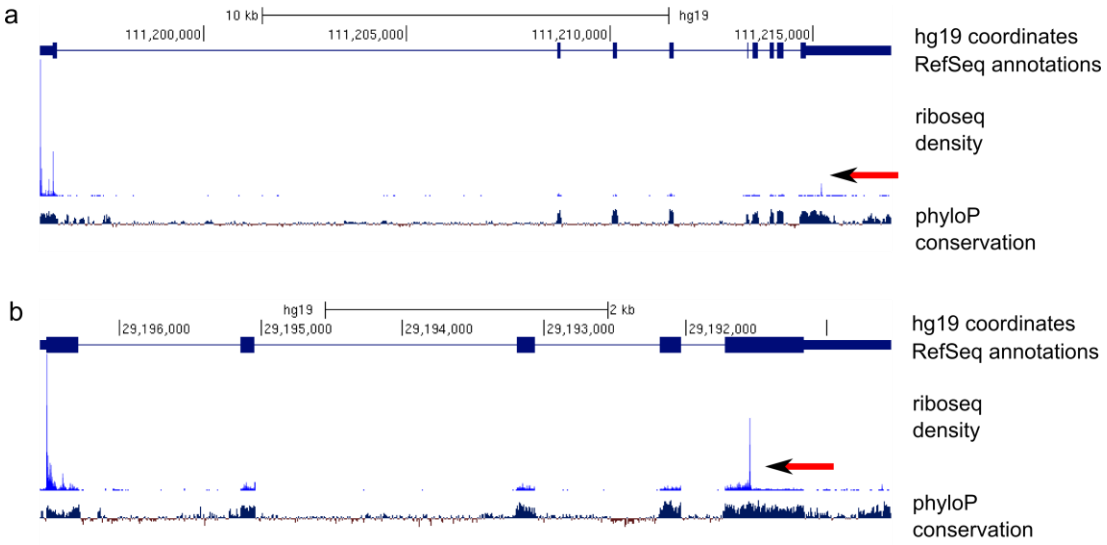
Extended Data Figure 1 | Genomic alignment of tetrapods from UCSC Genome browser 100 species alignment. Codon alignment was obtained with CodAlignView, positions of *AMD1* stop and *AMD1 tail* stop are annotated (second row).

Extended Data Figure 2| Alignment of *AMD1* coding region and surrounding areas from 146 vertebrate species. Synonymous and nonsynonymous substitutions are indicated by blue and red colours, respectively, and gaps are in grey. Ka/Ks ratio and sequence identity (see Methods) are shown at the bottom.

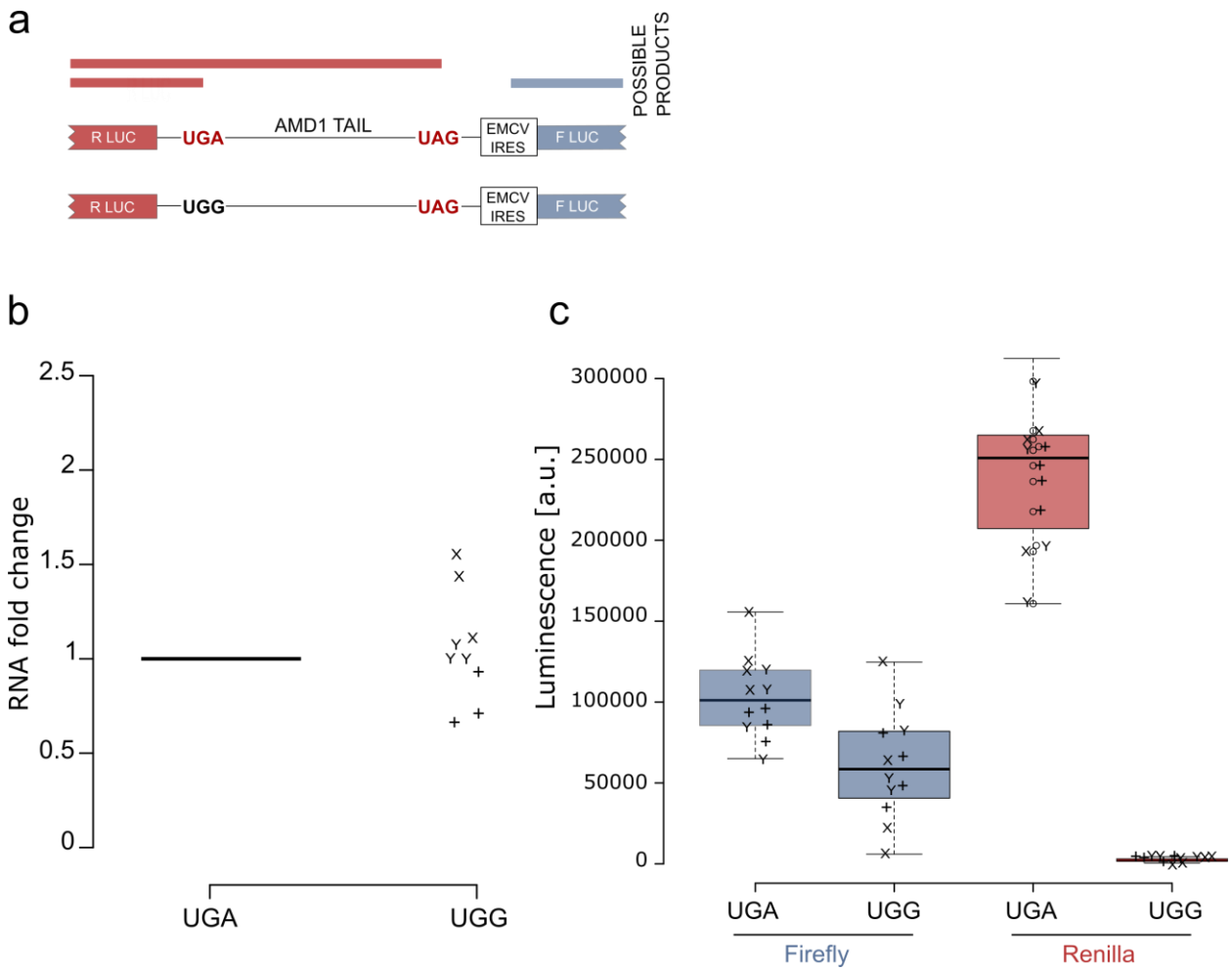


Extended Data Figure 3 | Cross-species examination of *AMD1 tail* using publicly available ribosome profiling data in GWIPS-viz browser. Available ribosome footprints aligned to the genomes of (**a**) mouse, (**b**) rat, (**c**) frog and (**d**) fish are shown along with gene annotation tracks

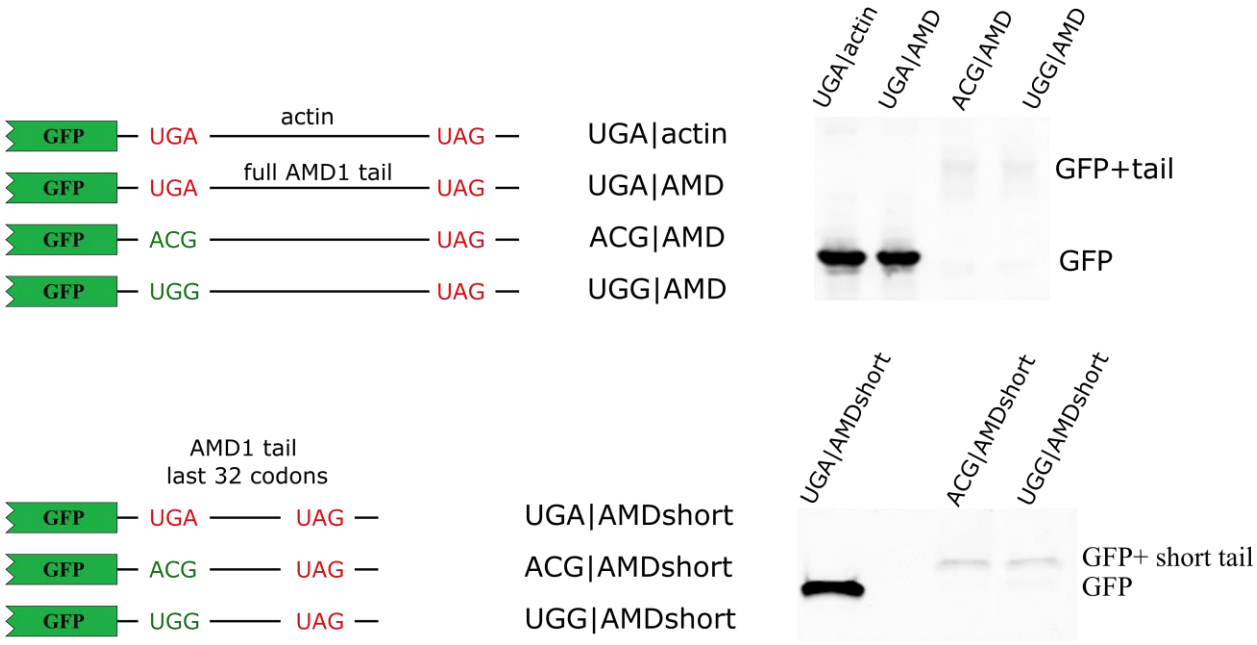
and ORF plots in which ATG codons are shown in green and stop codons are shown in red. Note that even under very low coverage, peaks of density are consistently present at the stop codon of *AMD1 tail*.



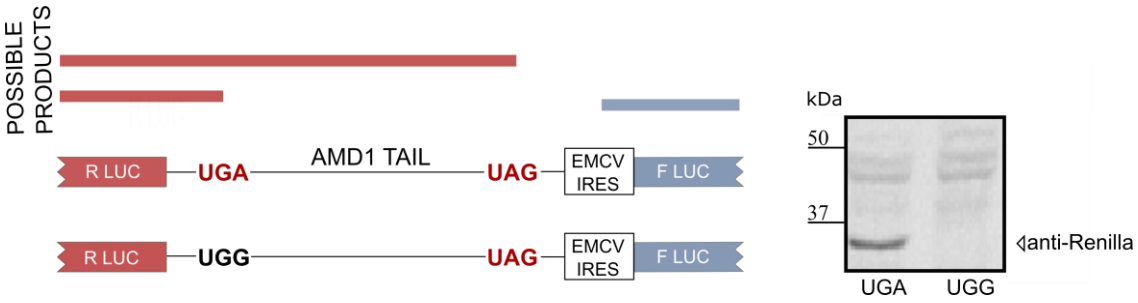
Extended Data Figure 4 | Human ribosome profiling data obtained with approaches that enrich ribosomes at the translation initiation sites. (a) *AMD1* locus, (b) *XBP1* locus. Three tracks are shown as indicated in the figure. Riboseq density corresponds to aggregated data obtained with drug treatments that preferentially arrest initiating ribosomes. Under these treatments actively elongating ribosomes run off. However, stalled ribosomes remain bound to mRNA and produce footprints along with initiating ribosomes blocked by these inhibitors. The peaks corresponding to ribosome stalling at the end of *AMD1 tail* and at the end of *XBP1* coding region (in unprocessed mRNA) are indicated with arrows.



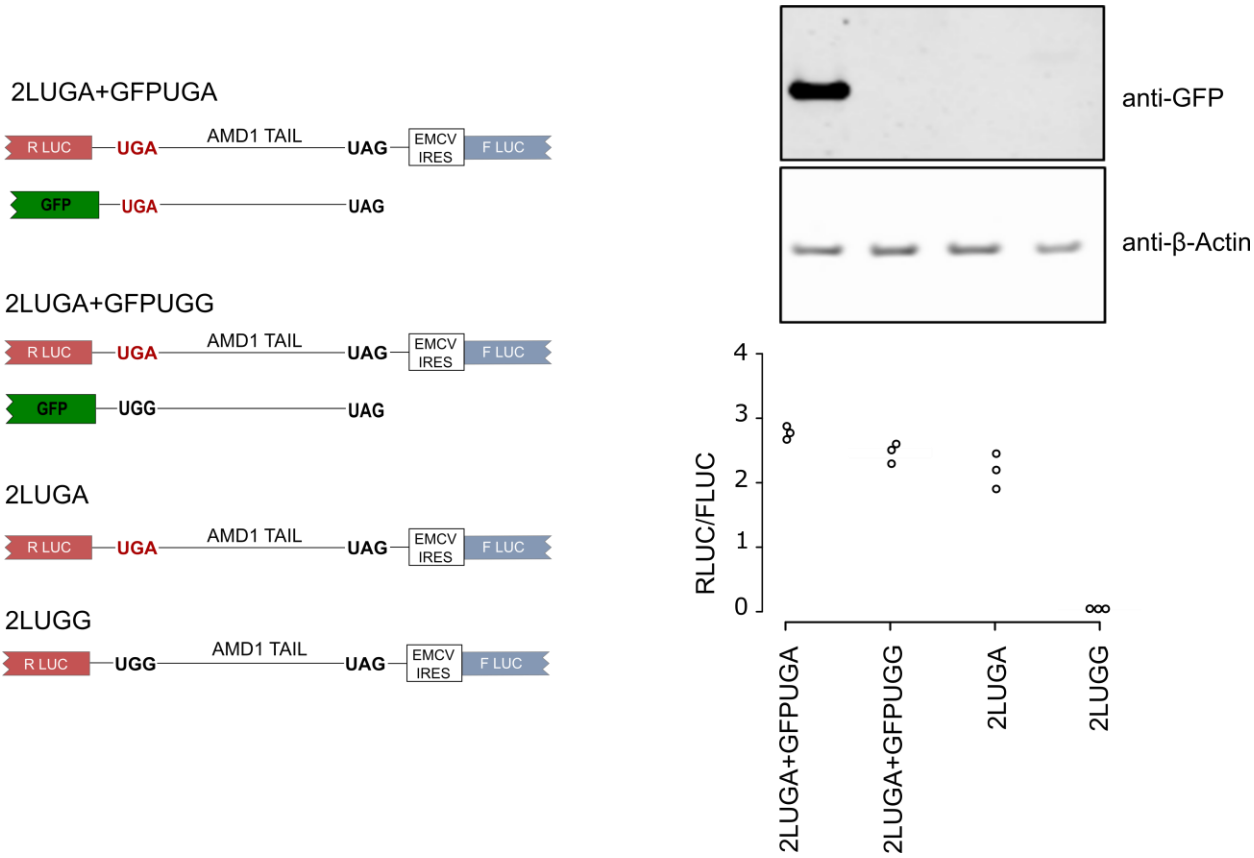
Extended Data Figure 5 | Assessment of dual luciferase mRNA stability. a. Scheme of constructs. **b**. RT-qPCR analysis with primers targeting R Luc sequence. **c.** Absolute values of R Luc and F Luc. Technical replicates that belong to the same biological replicate are indicated with the same symbols in panels **b** and **c**.



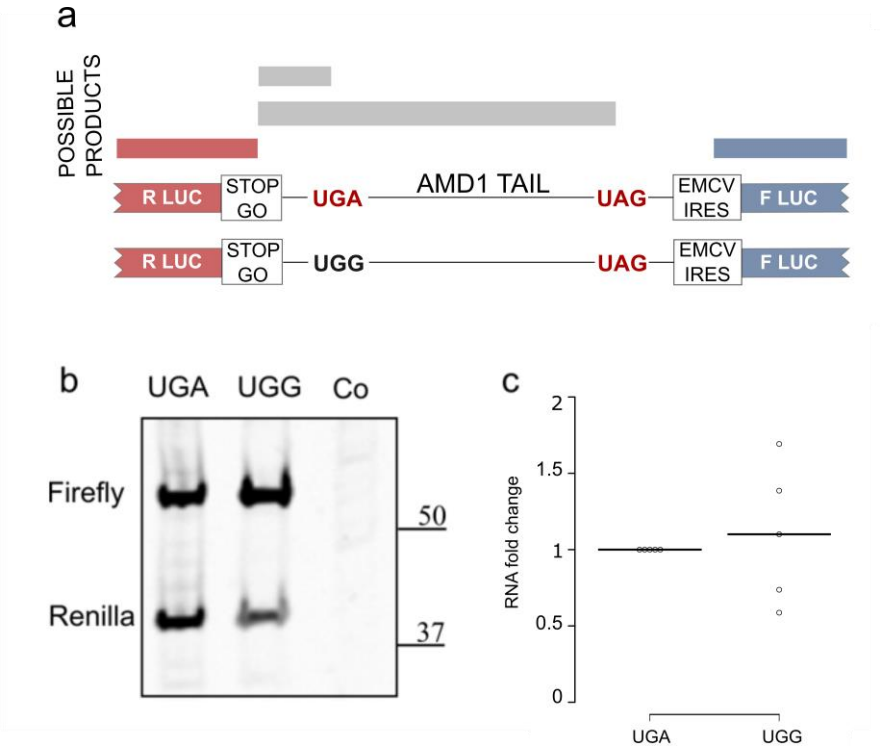
Extended Data Figure 6 | Expression of GFP constructs. Western blotting analysis of GFP fusions with a fragment of the actin 3' trailer of the same length as *AMD1 tail* (full length, top) and truncated from the 5' end tail, separated with stop and sense codons (bottom).



Extended Data Figure 7 | *In vitro* translation of *AMD1 tail* fusion reporters. Western blotting analysis of luciferase expressing mRNAs in HEK293T cell free translation system.



Extended Data Figure 8 | Potential trans effect of AMD1 tail translation. HEK293T cells were transfected in triplicate half-area 96 well plates with the indicated expression constructs (left) for 24 h. Cells were lysed in 15 µl PLB and incubated shaking for 15 min. at room temperature. 5 µl of each was removed for immunoblotting with both anti-GFP and anti-β-Actin (upper right) and the remaining lysate assayed for both R Luc and F Luc activities (lower right), n=3.



Extended Data Figure 9 | Expression of StopGo constructs. a. Scheme of constructs. **b**. Western blots with antibodies against Renilla and Firefly. **c**. RT-qPCR analysis.

a.

b.



Extended Data Figure 10 | EEF1A2 readthrough extension. a. GWIPS-viz screenshot of ribosome footprint density at the last 3' exon of *EEF1A2* (hg38). **b**. Multiple sequence alignment of 100 vertebrate genomes visualized with CodAlignView. It can be seen that the most conserved regions (no substitutions, white) coincides with the region upstream of the candidate stalling site.