One-shot analysis of translated mammalian lncRNAs with AHARIBO

Luca Minati†[1], Claudia Firrito†[1], Alessia Del Piano†[1], Alberto Peretti†[1], Simone Sidoli[2,] Daniele Peroni[3], Romina Belli[3], Francesco Gandolfi[4], Alessandro Romanel[4], Paola Bernabò[1], Jacopo Zasso[5], Alessandro Quattrone[5], Graziano Guella[6], Fabio Lauria[7], Gabriella Viero[7] and Massimiliano Clamer[1]*

[1]IMMAGINA BioTechnology S.r.l, Via Sommarive 18, Povo (Italy). [2]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [3]Mass Spectrometry Facility, Computational and Integrative Biology (CIBIO), University of Trento, Via Sommarive 9, Trento 38123, Italy. [4]Laboratory of Bioinformatics and Computational Genomics, Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Via Sommarive 9, Trento 38123, Italy. [5]Laboratory of Translational Genomics, Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Via Sommarive 9, Trento 38123, Italy. [6]Department of Physics, University of Trento, Via Sommarive 14, 38123 Povo Trento, Italy. [7]Institute of Biophysics, CNR Unit at Trento, Via Sommarive, 18 Povo, Italy.

*For correspondence: mclamer@immaginabiotech.com (MC)

† These authors contributed equally to this work

Abstract

A vast portion of the mammalian genome is transcribed as long non-coding RNAs (lncRNAs) acting in the cytoplasm with largely unknown functions. Surprisingly, lncRNAs have been shown to interact with ribosomes, encode peptides, or act as ribosome sponges. These functions still remain mostly undetected and understudied owing to the lack of efficient tools for genome-wide simultaneous identification of ribosome-associated and peptide-producing lncRNAs. Here we present AHARIBO, a method for the detection of lncRNAs either untranslated, but associated with ribosomes, or encoding small peptides. Using AHARIBO in mouse embryonic stem cells during neuronal differentiation, we isolated ribosome-protected RNA fragments, translated RNAs and corresponding *de novo* synthesized peptides. Besides identifying mRNAs under active translation and associated ribosomes, we found and distinguished lncRNAs acting as ribosome sponges or encoding micropeptides, laying the ground for a better functional understanding of hundreds lncRNAs.

Keywords

LncRNA, translation, proteogenomics, ribosome, RNA, proteomics, translatome, RIBO-seq

Introduction

An incredibly small fraction of the mammalian genome is protein-coding (< 3 %), while the number of potentially functional non-coding genes remains unclear (Djebali et al., 2012). LncRNAs are defined as non-coding RNA exceeding 200 nt. They have gained much attention because of their role in a variety of cellular processes, from chromatin architecture (Minajigi et al., 2015) to mRNA turnover (Kleaveland et al., 2018) and translation (Ingolia et al., 2011). Typically, lncRNAs are abundant transcripts (Iyer et al., 2015) that display short and not evolutionarily conserved ORFs with minimal homology to known protein domains (Guttman and Rinn, 2012). The majority of lncRNAs are localized in the cytoplasm (Carlevaro-Fita et al., 2016), where they are supposed to remain untranslated. Ribosome profiling (RIBO-seq), which provides positional information of ribosomes along transcripts (Clamer et al., 2018; Ingolia et al., 2012), identified several ribosome-associated lncRNAs (Bazzini et al., 2014; Ingolia et al., 2011; Lee et al., 2012; Zeng et al., 2018). A handful of lncRNAs have been shown to be involved in translation regulation (Carrieri et al., 2012; Yoon et al., 2012), while others are themselves potentially or partially translated (Anderson et al., 2015; Aspden et al., 2014; Bazin et al., 2017; Ingolia et al., 2011; Nelson et al., 2016; Ruiz-Orera et al., 2014; van Heesch et al., 2019). As coding RNAs, lncRNAs can be associated with actively translating or translationally silent ribosomes (Chandrasekaran et al., 2019; Chen et al., 2020; Jiao and Meyerowitz, 2010; Kapur et al., 2017). Hence, the potential involvement of lncRNAs in translation increases the complexity of the mammalian control of gene expression at the translatome and proteome level. Unfortunately, classical RIBO-seq approaches barely distinguish between lncRNAs producing peptides from those that sequester ribosomes (lncRNA bound to ribosomes without translation) and acting as ribosome sponges. Proteomics approaches, such as mass spectrometry, can help to define and quantitatively monitoring the production of peptides, but are less sensitive techniques than RNA sequencing (Slavoff et al., 2013; van Heesch et al., 2019). Therefore, proteomics and RIBO-seq alone cannot unravel the wide functional range of cytoplasmic lncRNAs associated with the translation machinery.

To fill this gap, we developed AHARIBO (AHA-mediated RIBOsome isolation), a combination of protocols that simultaneously isolate RNAs and nascent proteins associated with translationally active ribosomes. AHARIBO is based on the isolation of ribosomes trapped with their nascent peptides, by incorporating the non-canonical amino acid L-azidohomoalanine (AHA), followed by parallel RNA-seq, ribosome profiling and proteomics.

We applied AHARIBO to human and mouse cells and showed that it enables to: i) purify translating ribosomes via nascent peptide chains, ii) co-purify RNAs and proteins for transcriptome / *de novo* proteome-associated studies, and iii) detect the regulatory network of lncRNAs translated or associated with ribosomes.

Results

Nascent peptide labelling and separation of the ribosome complex with AHARIBO-rC

To simultaneously purify ribosomes under active translation, associated RNAs and corresponding growing peptide chains, we optimized a protocol in HeLa cells (Figure 1A). Briefly, the protocol consists of the following phases: i) incubation with a methionine-depleted medium, ii) addition of the methionine analogue AHA, iii) on-ribosome anchorage of nascent peptide chains with a small molecule, iv) cell lysis and AHA "copper-free click reaction" (Jewett and Bertozzi, 2010) for v) ribosome capture with magnetic beads. We reasoned that the protocol for isolating ribosomes through AHA can be used to obtain information about nascent peptides, constitutive components of ribosomes, mRNAs and lncRNAs associated with them. For this reason, we optimized several parameters from washing steps to nuclease treatments (Figure 1a), to isolate (1) the full translational complex (AHARIBO-rC, Ribosomal Complexes: ribosomes, ribosome-associated proteins, nascent peptides and RNAs) , (2) the *de novo* synthesized proteome (AHARIBO-nP, nascent Proteome) and (3) ribosome-protected fragments (RPFs) (AHARIBO RIBO-seq: RIBOsome profiling by sequencing).

To minimize the amount of AHA-tagged and fully synthesized proteins released from ribosomes and achieve optimal on-ribosome polypeptide stabilization, we tested multiple incubation-time of AHA exposure and compared the effect of two small molecules (namely cycloheximide (CHX) and sBlock, an anisomycin-based reagent). Anisomycin is known to inhibit the activity of eukaryotic ribosomes, while keeping polypeptides bound to translating ribosomes (Garreau de Loubresse et al., 2014; Grollman, 1967; Seedhom et al., 2016)

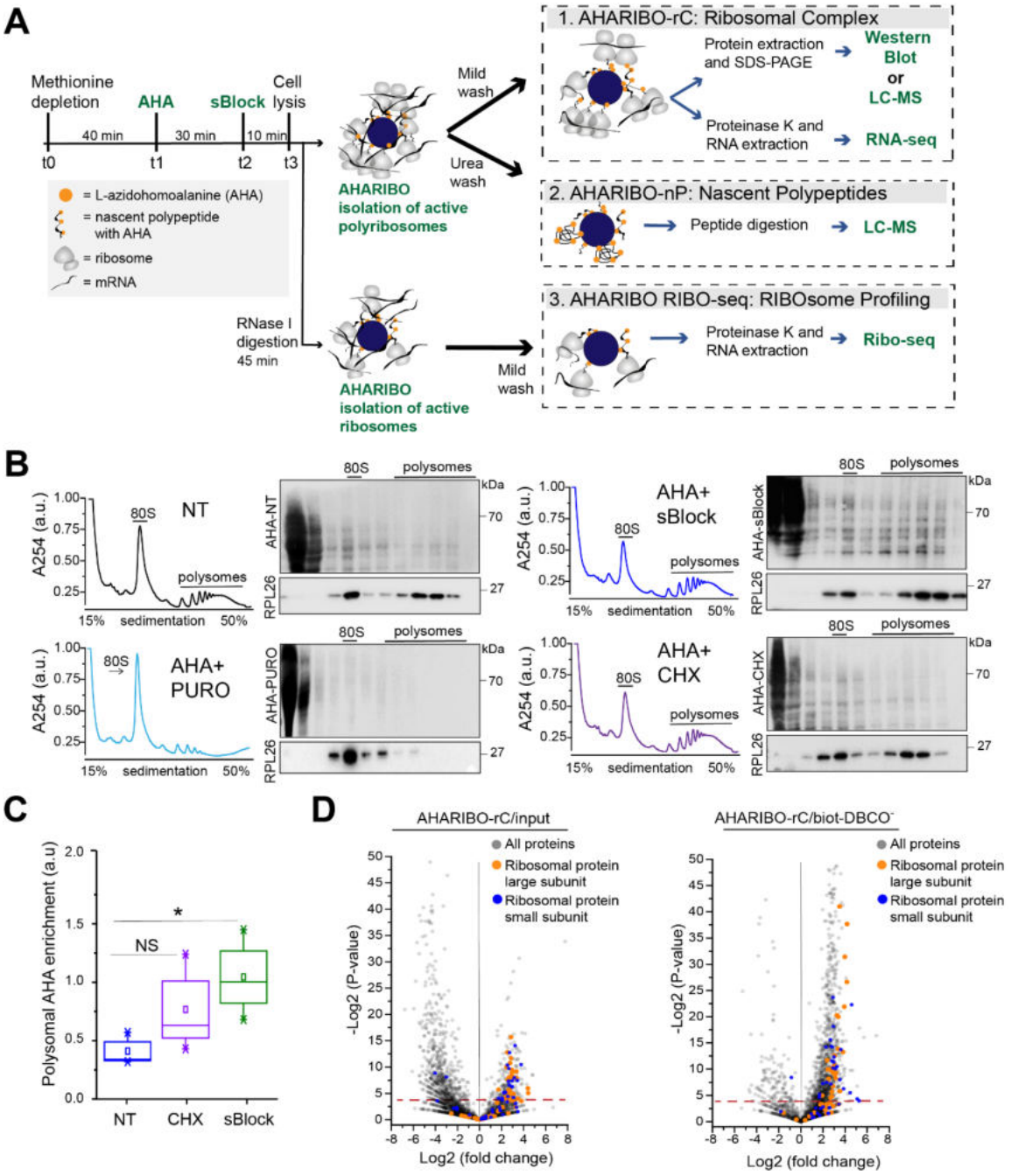

Figure 1. AHA labelling of nascent peptide chains and ribosome separation. (A) Schematic representation of AHARIBO workflow. After methionine depletion, AHA incubation and sBlock treatment, cell lysates can be processed for (1) AHARIBO-rC: isolation of translational complexes (ribosomes, ribosome-associated proteins, nascent peptides and RNAs); (2) AHARIBO-nP: isolation of *de novo* synthesized proteome; (3) AHARIBO RIBO-seq: for ribosome profiling. (B) Polysomal profiles in Hela cells. On the right of each profile, example of SDS-PAGE of protein extracts from each fraction of the profile. Staining of the membrane was performed by biotin cycloaddition followed by streptavidin-HRP. RPL26 protein was used as marker of the large ribosome subunit. (C) Box plot showing the AHA signal enrichment in the polysomal fractions of the profiles in cells untreated (NT) and treated with either CHX or sBlock. Results are shown as the median (± SE) of 3 independent experiments. NS, not significant. (*) P-value = 0.05 was obtained through an unpaired t-test. (D) Volcano plots of AHARIBO-rC-isolated proteins. Data are compared with input (AHA-containing lysate, left) or with streptavidin-coated beads without biotin-DBCO (right). DBCO, Dibenzocyclooctyne. Red line: t-test p-value < 0.05.

The online version of this article includes the following figure supplement(s) for Figure 1:

Figure 1-figure supplement 1. AHA incorporation, validation of AHA and RNA capture.

Figure 1-figure supplement 2. LC-MS analysis on AHARIBO-rC proteins and validation by western blot.

Figure 1-figure supplement 3. AHARIBO-rC efficiency test and validations.

Figure 1-source data 1. Relative abundance of AHARIBO-rC isolated proteins.

Figure 1-source data 2. Gene Ontology analysis.

We observed that 30 min is the optimal incubation time for sufficient AHA incorporation and maximum RNA recovery (Figure 1-figure supplement 1A-C). Next, we compared the efficiency of CHX and sBlock in stabilizing the nascent peptide by co-sedimentation analysis of AHA-tagged polypeptides with ribosomes along the sucrose gradient (Figure 1B). As a control, cells were treated in parallel with puromycin to cause ribosome disassembly and release of the growing peptide chains (Figure 1B) (Blobel and Sabatini, 1971; Enam et al., 2020). In agreement with literature, we found that both CHX and sBlock are able to stabilize AHA-peptides on ribosomes and polysomes (Biever et al., 2020; Mathias et al., 1964). The efficiency of anchoring polypeptides on ribosomes in CHX and sBlock treated cells was about 50% higher compared to untreated cells, confirming that the treatment effectively stabilize nascent polypeptides (Figure 1C). The high signal observed in lighter fractions is likely caused by AHA-labeled proteins released from ribosomes. To overcome this problem it is possible to perform a pre-cleaning of the cell lysate by sucrose cushioning. This step can increase the efficiency of total RNA isolation with AHARIBO compared with the control (no AHA) (Figure 1-figure supplement 1D). As expected, in puromycin-treated samples the AHA signal was mainly detected in the first two fractions of the gradient, proving that the signal observed in the heavier fractions of CHX- and sBlock-treated cells was not caused by diffusion of AHA-labeled peptides from lighter to heavier fractions. Since sBlock outperformed CHX in anchoring efficiency (Figure 1C), we used this compound in all further experiments.

Prompted by the evidence that nascent peptides can be stably anchored on ribosomes by a small molecule, we isolated RNAs and proteins associated with the translation complex. To this aim, we performed a label-free liquid chromatography-mass spectrometry (LC-MS) analysis of AHARIBO captured proteins relative to the input, to the background biotin-DBCO$^-$ (Figure 1d) or AHA$^-$ (Figure 1 – figure supplement 2A) (Figure 1D-source data 1) and to a sample treated with puromycin (AHA$^+$ puromycin) (Figure 1- figure supplement 2B), which

causes the release of nascent chains. We observed that ribosomal proteins belonging to both the large and the small ribosome subunits are indeed more abundant in AHARIBO-rC samples than in controls. LC-MS results were confirmed by Western blot analysis of proteins that are component of the large and small ribosomal subunits (RPS6, RPL26) (Figure 1- figure supplement 2B). Gene ontology (GO) analysis revealed that terms related to translation (biological process), nucleic acid binding (cellular function) and ribonucleoprotein complex (cellular component) are enriched in AHARIBO-rC compared to the control (no AHA), confirming efficient pulldown of translation-related proteins (Figure 1-source data 2).

Then, we used AHARIBO-rC to determine the translational status of cultured cells. To this aim, we down-regulated protein synthesis by treating HeLa cells with puromycin, heat shock (HS) (10 min at 42°C, during AHA incubation) or arsenite (Ar) treatment, which induces translational inhibition and stress granules formation (Wang et al., 2016). We observed a reduction of RNA captured in puromycin-, HS- and Ar-treated cells relative to the control (Figure 1-figure supplement 3A-C). In line with this finding, qRT-PCR analysis showed about 50 % reduction in 18S rRNA levels when translation was inhibited (Figure 1 - figure supplement 3D).

To further validate AHARIBO-rC, we took advantage of a micropeptide (176 aa) originating from an open reading frame of the TUG1 lncRNA, called TUG1-BOAT (Lewandowski et al., 2020). The wild-type (WT) ORF has a non-canonical start codon and a methionine 75 nt upstream the stop codon. We ectopically expressed the WT TUG1-BOAT transcript and two mutant constructs (Figure 1-figure supplement 3E): (a) the ΔTUG1-BOAT, without the methionine 75 nt upstream the stop codon; (b) the +1Met TUG1-BOAT with an ATG (methionine) as start codon. The +1Met TUG1-BOAT has two methionines, one at the N terminal and a second one at 25 aa (75 nt) before the C-terminal. Our RT-qPCR analysis performed 24h or 48h after transfection showed a good efficiency of AHARIBO in capturing the TUG1-BOAT RNA when methionines are present (about 50 times more in +1Met TUG1-BOAT than in ΔMet TUG1-Boat after 24h) (Figure 1- figure supplement 3E), confirming the efficiency of AHARIBO-rC in capturing translated RNA.

AHARIBO-nP: genome-wide portray of the *de novo* synthesized proteome

Motivated by the evidence that AHARIBO-rC can be used to isolate *bona-fide* active ribosomes, we further tested our method genome-wide in mouse embryonic stem cells

(mESCs) under basal condition and after differentiation into early neurons (EN) (Tebaldi et al., 2018) (Figure 2-figure supplement 1A). We analysed both AHARIBO-rC isolated RNA and newly synthesized polypeptides associated with actively translating ribosomes by RNA-seq and LC-MS, respectively. The protocol for the isolation of the *de novo* synthesised polypeptides (named AHARIBO-nP) is based on urea washing to remove all proteins that are not nascent peptides (Figure 2-figure supplement 1B). In parallel, we isolated and analysed the global translatome by extracting the RNA after 30% sucrose cushioning of cytoplasmatic lysates (Wang et al., 2013) and we analysed the global proteome by pulsed SILAC (pSILAC) (Schwanhäusser et al., 2009) (Figure 2A).

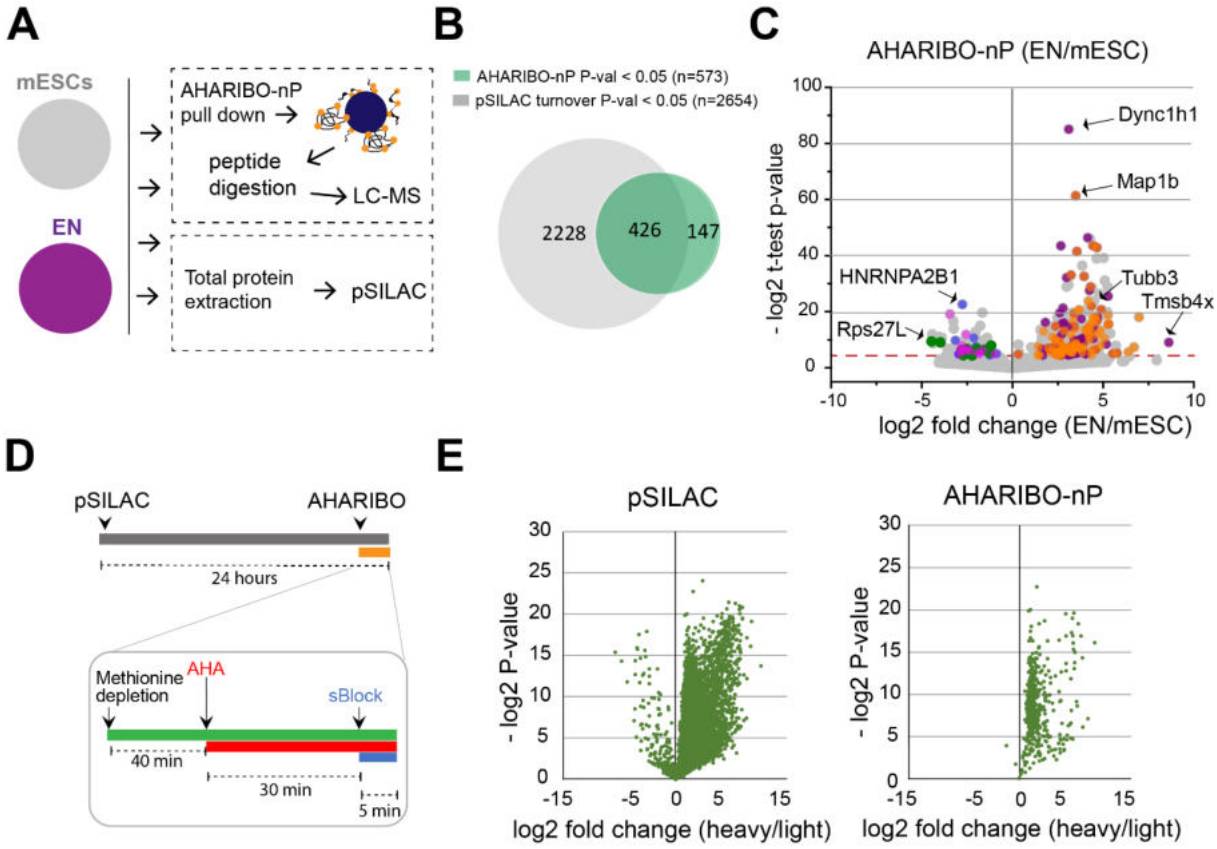

Figure 2. AHARIBO-nP and pSILAC. (A) Workflow for parallel AHARIBo-nP and pSILAC. mESCs, mouse embryonic stem cells. EN, mouse embryonic stem cells differentiated in early neurons. (B) Venn diagram representing the number of differentially expressed proteins (EN/mESCs) identified by AHARIBO-nP and pSILAC (p-value < 0.05). (C) Volcano plot for each differentially expressed protein (EN/mESC) the AHARIBO-nP proteome versus -log2(p-value). Red broken line indicates p-value < 0.05. Orange and purple dots represent up-regulated proteins involved in cytoskeleton organization (GO:0007010) and neurogenesis (GO:0022008) respectively. Blue, green and magenta dots represent down-regulated proteins related to RNA processing (GO:0006396), protein synthesis (GO:0006412) and mouse pluripotency (WP1763). Grey dots, represent all other proteins. (D) Schematic representation of combined cell treatments for pSILAC and AHARIBO-nP. (e) Volcano plots displaying for each protein the - log2 t-test p-value against the fold changes of protein turnover (heavy/light) in pSILAC proteome (left) and AHARIBO-nP (right) for double-treated mESCs.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure 2-figure supplement 1. Cell differentiation and additional proteomic analysis.

Figure 2-source data 1. pSILAC proteomic data.

Figure 2-source data 2. AHARIBO-nP differentially expressed proteins.

Quantitative proteomic analysis of ENs versus mESCs (EN/mESC) led to the identification of 2654 differentially expressed proteins (Figure 2B) (Figure 2-source data 1). As expected, differentiated cells (EN) showed a reduced turnover compared to mESCs (Figure 2-figure supplement 1C). In parallel, EN and mESC cells were analysed by AHARIBO-nP, which captured 1365 and 2215 proteins respectively. Of note, 74 % of proteins identified through AHARIBO-nP is in common with the pSILAC dataset. The smaller number of proteins identified with AHARIBO-nP compared to pSILAC is most probably related to the shorter time of incubation with AHA (30 min) compared to pSILAC (24 hours) and is consistent with previous observations from similar pulldown enrichment strategies (Bagert et al., 2014; Rothenberg et al., 2018). Differential expression analysis (EN/mESC) identified 573 proteins (p-value < 0.05) in AHARIBO-nP (Figure 2B) (Figure 2-source data 2). The GO analysis of differentially expressed proteins showed that proteins involved in cytoskeleton organization and neurogenesis were upregulated (Figure 2C), further confirming the reliability of AHARIBO-nP in monitoring *de novo* protein expression. We focused on proteins captured by AHARIBO-nP during differentiation (Figure 2C) (Figure 2-source data 2) and found that several are known to be expressed during early stages of development of the nervous system (e.g. Map1b, Tubb3 and, Dync1h1) (Fiorillo et al., 2014; Gonzalez-Billault et al., 2002; Latremoliere et al., 2018). In addition, we performed AHARIBO-nP pull-down in mESCs double-labelled for pSILAC (24 h) and AHA (30 min) (Figure 2D). Interestingly, we observed high fold changes of heavy amino acids in AHARIBO-nP (Figure 2E) and a significantly higher protein turnover in the AHARIBO-nP compared to the pSILAC proteins (Figure 2 - figure supplement 1D), suggesting that AHARIBO-nP is indeed able to capture the *de novo* synthesized polypeptides.

Collectively, these results show that AHARIBO-nP is captures *de novo* synthesized proteins and produce meaningful descriptions of phenotypic changes occurring upon cell differentiation. Moreover, these results demonstrate that our AHARIBO-nP protocol is suitable to monitor dynamic changes in protein expression by LC-MS analysis.

Combination of AHARIBO-rC and AHARIBO-nP: parallel genome-wide analysis of translated RNAs and *de novo* synthesized proteome

Prompted by previous results, we asked if mRNAs purified using AHARIBO-rC are a good proxy of protein levels. To this aim, we compared AHARIBO-rC RNA and the global translatome with AHARIBO-nP in mESCs during differentiation.

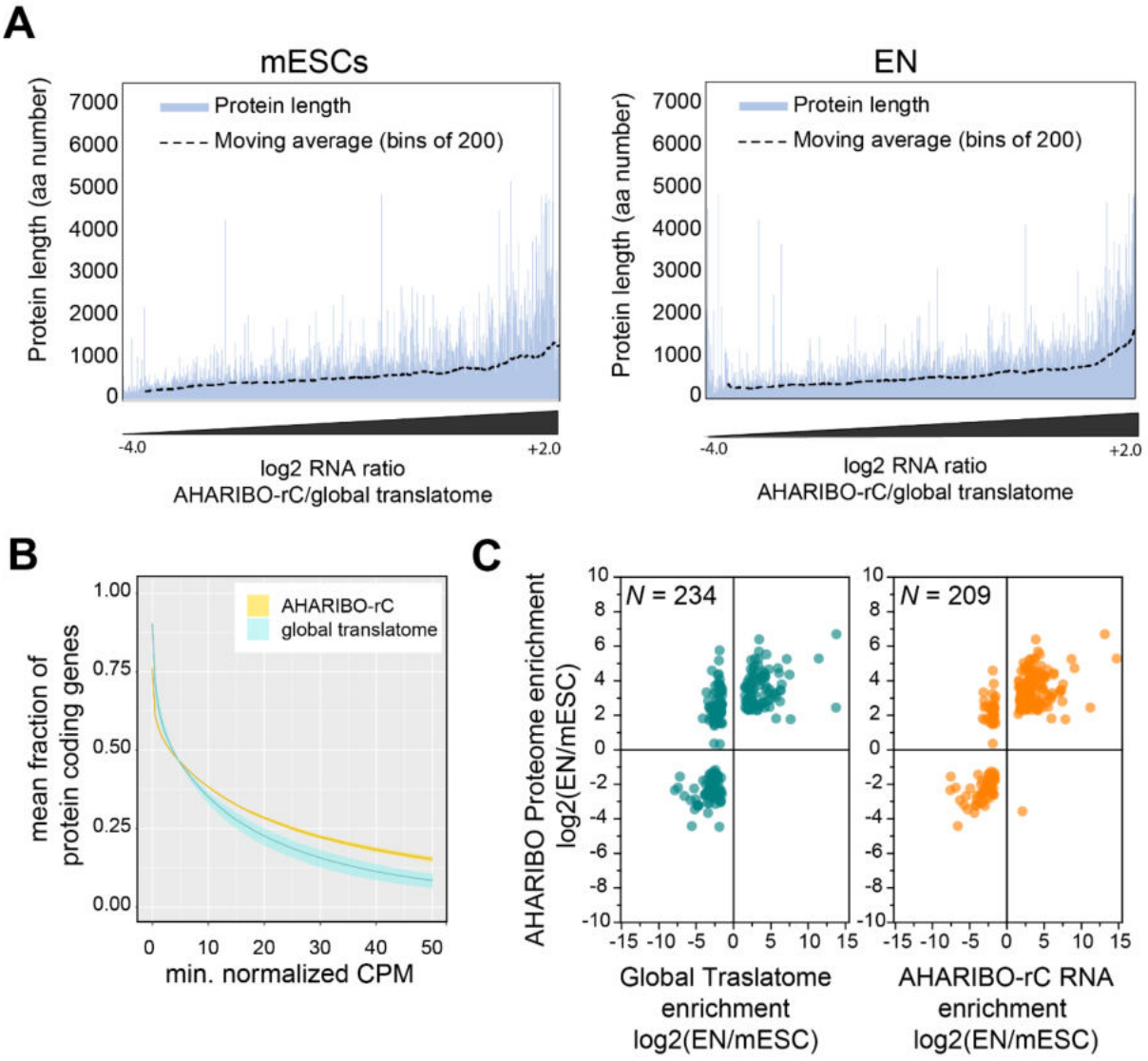

Figure 3. AHARIBO-rC RNA vs *de novo* proteome analysis. (A) Enrichment of a given transcript obtained with AHARIBO vs global translatome (x-axis) as a function of the theoretical protein length (y-axis) for mESCs (left) and ENs (right). Each bar represents the number of enriched transcripts with the defined theoretical protein length. (B) Fraction of coding genes expressed above a minimum threshold in EN. The AHARIBO-rC and the global translatome group are represented in yellow and cyan, respectively. For each group, the mean (solid line) and the SD (shades) of the fractions for a given count per million (CPM) threshold are calculated over all samples (n = 6) in that group. (C) Scatter plot of RNA fold change (global translatome on the left, AHARIBO-rC on the right) compared to protein fold change (AHARIBO-nP) obtained by comparing EN with mESC. N, number of DEGs with p-value < 0.05.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure 3-figure supplement 1. RNA-seq and protein coding RNA analysis.

Figure 3-source data 1. Differentially expressed genes (DEGs) from RNA-seq data.

Figure 3-source data 2. RNA and protein DEGs .

To exclude any bias related to protein length, we checked whether AHARIBO-nP preferentially captures long or short proteins. We plotted the peptide size against the enrichment resulting from AHARIBO-rC compared with the global transcriptome (Figure 3A). This value represents the extent to which AHARIBO-rC RNA differs from the standard method. Our results confirm that AHARIBO captures transcripts encoding for polypeptides in a wide range of length (Figure 3a). Since in all eukaryotes proteins are initiated with a methionine residue and the average protein size in eukaryotes is about 300 aa (Frith et al., 2006), virtually any protein can be captured as soon as the nascent peptide exits the ribosome (i.e., when it reaches a length of about 35-40 aa). In about 70 % of the proteome, the N-terminal methionine is co-translationally cleaved when the peptide is at least 50 aa long by the enzyme methionine aminopeptidase (Wild et al., 2020), while the remaining 30% retains the methionine (Martinez et al., 2008). Therefore, there is a reasonable probability for at least one AHA residue to be available for each peptide when the inhibitor of translation (sBlock) is added to the cell medium, enabling the capture of the polypeptide outside the ribosome exit tunnel.

To further prove the reliability of our method, we measured the efficiency of AHARIBO-rC to capture coding transcripts compared to a global translatome analysis. Using increasing abundance thresholds in EN we observed that AHARIBO-rC efficiency is comparable to the global translatome for low abundant transcripts in EN and for all transcripts in undifferentiated mESCs (Figure 3 - figure supplement 1A). Strikingly, AHARIBO captures abundant transcripts in EN with much higher efficiency than the global translatome (Figure 3B).

Finally, we tested whether the RNA isolated with AHARIBO-rC can predict the *de novo* synthesized proteome. After comparing differentially expressed genes (DEGs) during differentiation to the AHARIBO-nP proteome (Figure 3-source data 1), we observed that AHARIBO-rC RNA is a good proxy of the newly synthetized proteome (Pearson's correlation *r*

= 0.75, Figure 3C and Figure 3 - figure supplement 1B). In particular, we found that AHARIBO-rC RNA presents less uncoupled genes (up-RNA and down-protein or down-RNA and up-protein) than the global translatome (Figure 3 - figure supplement 1C), thus faithfully recapitulating proteome changes. The correlation of the global translatome with the global protein turnover measured with pSILAC shows a Pearson's *r* = 0.27 (Figure 3 - figure supplement 1D) (Figure 3-source data 2). This result demonstrate that AHARIBO-nP does reflect the labeling of peptides rather than completely synthesized proteins.

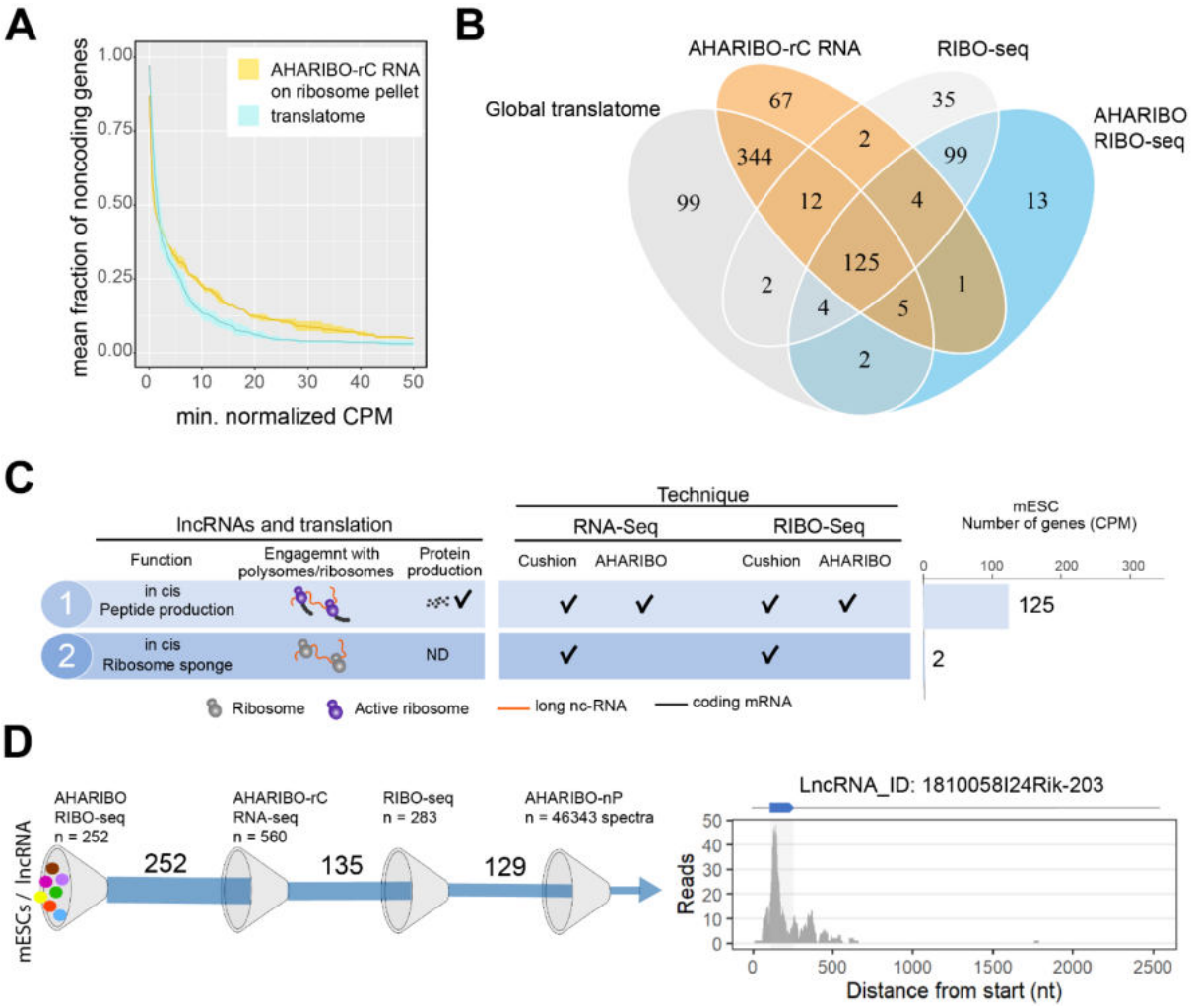

Figure 4. The AHARIBO platform can be used to detect ribosome-interacting lncRNAs. (A) Linear plot illustrating the fraction of noncoding genes expressed above a minimum threshold in EN. The AHARIBO-rC and the global translatome group are represented in yellow and cyan, respectively. For each group, the mean (solid line) and the SD (shades) of the fractions for a given CPM threshold are calculated over all samples (n = 3) in that group. Expression values are indicated as normalized CPM . AHARIBO-rC was performed on the ribosome pellet after sucrose cushioning. (B) Venn diagram of the number of lncRNAs genes with at least 1 CPM identified by RNA-seq, AHARIBO-rC, RIBO-seq, AHARIBO RIBO-seq. (C) Classification of lncRNAs interacting with ribosomes and relative detection through the multiple AHARIBO and standard approaches. ND, no detection of protein synthesis. (D) Left, schematic representation of the number of mESC lncRNAs in common between AHARIBO RIBO-seq, AHARIBO-rC RNA, standard RIBO-seq. These lnRNAs were validated by LC-MS. Right,

example a AHARIBO RIBO-seq ribosome occupancy profile of lncRNA 1810058I24Rik displaying the reads distribution along the entire transcript and the accumulation of reads at the known short open reading frame (shadow area and blue arrow on top).

The online version of this article includes the following figure supplement(s) for figure 4:

Figure 4-figure supplement 1. Isolation of lncRNAs with AHARIBO.

Figure 4-figure supplement 2. AHARIBO RIBO-seq data.

Figure 4-source data 1. lncRNAs identified by RNA-seq in mESCs.

Figure 4-source data 2. lncRNAs transcripts identified by RIBO-seq in mESCs.

Figure 4-source data 3. Matching peptides from lncRNAs .

Combined AHARIBO approaches define the functional role of lncRNAs in translation

Based on the evidence that a combination of AHARIBO approaches can simultaneously detect RNAs under active translation and peptides in the process of being produced, we applied our methods to detect ribosome-associated and translated native lncRNAs.

In AHARIBO-rC data, we identified a total of 687 lncRNA genes in mESCs and about 400 differentially expressed (DE) lncRNAs during neuronal differentiation (Figure 4 - figure supplement 1A) (Figure 4-source data 1). Among the top-5 DE lncRNAs (fold change > 10; p-value < 1×10^{-10}), we found *Pantr1* and *Lhx1os*, known to be involved in neuronal development (Biscarini et al., 2018; Carelli et al., 2019). To identify potentially translated lncRNAs, we applied the abundance threshold analysis to the subset of AHARIBO-rC non-coding RNAs in common with a published dataset (n = 270) of lncRNA identified by in ribosome profiling data in mESCs (Ingolia et al., 2011) (Figure 4-figure supplement 1B). The analysis of 100 lncRNAs in common between the two datasets showed a stronger enrichment of ribosome footprints in the AHARIBO-rC than in the global translatome (Figure 4A and Figure 4-figure supplement 1C). Altogether, these result suggests that a fraction of non-coding transcripts, which is efficiently isolated with AHARIBO-rC, is potentially translated.

To understand if and how lncRNAs interact with ribosomes, we performed ribosome profiling experiments after AHARIBO pulldown (named AHARIBO RIBO-seq), with parallel standard RNA-seq (on inputs) analysis in mESCs. For protein-coding genes, both standard and AHARIBO RIBO-seq show an enrichment of RPFs in the coding sequence (Figure 4 – figure supplement 2A). The two datasets show high correlation (Figure 4 – figure

supplement 2B) and the expected codon periodicity in the coding sequence in AHARIBO RIBO-seq (Figure 4 – figure supplement 2C). These results further confirms the capability of AHARIBO in capturing ribosomes. With AHARIBO RIBO-seq we identified a list of lncRNAs covered by ribosome footprints (Figure 4-source data 2). By intersecting our AHARIBO RIBO-seq data with those obtained from standard methods (RIBO-seq and RNA-seq after sucrose cushioning) or AHARIBO-rC, we identified 125 common putative translated lncRNAs (Figure 4B). Some of these lncRNA (n = 19) are known to be translated in mouse tissue (van Heesch et al., 2019). The vast majority of these lncRNAs do not have known function. Two of the identified lncRNAs (9330151L19Rik and Gm9776) were detected only by standard RIBO-seq and RNA-seq but not with AHARIBO (Figure 4C). This results may be due to the absence of translation events (i.e. transcripts loaded with idle ribosomes). Next we validated the coding potential of lncRNAs which are in common between AHARIBO and standard RIBO-seq (Figure 4d). We translated *in silico* the transcripts in all frames to find potential ORFs with a canonical start codon (AUG). Translated sequences were semi-trypsin digested *in silico*, and then manually annotated to find confident matching spectra from the AHRIBO-nP protein dataset. Out of the about 46,000 collected spectra (Figure 4-source data 3), our MS-based proteomics analysis detected peptides with highly corresponding ribosome footprints (e.g. Gm42743, Gm26518, B230354K17Rik, D030068K23Rik, 1810058I24Rik). From the list of 129 lncRNAs that are in common among all AHARIBO protocols and standard RIBO-seq (Figure 4D), we identified by MS analysis a micropeptide (Mm47) of 47 aa (Figure 4D) at a high degree of confidence. This micropeptide derives from a lncRNA expressed in murine macrophages, and recently characterized by an independent group (Bhatta et al., 2020) as a relevant peptide able to modulate the innate immunity in mice. Several other lncRNA show high confidence of translation events with in silico prediction, even if they were not perfectly matching our proteomic spectra (Figure 4-figure supplement 3), paving the way for a better characterization of translatable lncRNA that has not been reported before. These results, combined with (i) AHARIBO's efficiency in detecting an ectopically expressed micropeptide (TUG1-BOAT) and (ii) concordance with recently published data, prove that our approach could be useful to unravel translation events in lncRNAs that are mis-annotated as non-coding. Altogether, our data confirm that our three diverse and complementary AHARIBO approaches represent a unique method to identify ribosome-associated and translated RNAs.

Discussion

LncRNAs localize in the nucleus or in the cytoplasm. In the nucleus they modulate transcription, pre-mRNA splicing or act as scaffold for protein interaction during chromatin organization (Sun et al., 2018). In the cytoplasm, the majority of lncRNAs is associated with polysomes (Carlevaro-Fita et al., 2016) where they either can or cannot produce proteins (Chen et al., 2020; Ingolia et al., 2011). Numerous lncRNAs are misannotated as non-coding but contain short ORFs encoding for micropeptides with biological relevance in cancer (D'Lima et al., 2017; Huang et al., 2017), bone development (Galindo et al., 2007), immunity (van Solingen et al., 2018), metabolism (Magny et al., 2013; Nelson et al., 2016) and DNA repaire (Slavoff et al., 2014). Different methodological approaches have been developed to quantify the variations of RNA abundance by sequencing or imaging techniques (Amit Blumberg et. al, 2019; Jao and Salic, 2008; Morisaki et al., 2016; Wu et al., 2016), RNA engagement with the translational machinery by RIBOseq, or polysomal profiling (Arava et al., 2003; Clamer et al., 2018; Eden et al., 2011; Taniguchi et al., 2010), and protein synthesis by mass spectromenty or metabolic labelling (Aviner et al., 2013; Dieterich et al., 2006; Schwanhäusser et al., 2009; Yan et al., 2016). Despite these advantages, available technologies hardly capture in a single experiment the dynamics of translation across multiple biological conditions, the translation of unannotated coding transcripts and translation-related functions of lncRNAs. Now that it is widely accepted that a portion of the genome annotated as non-coding can result in a complex transcriptome partially engaged with ribosomes (Chen et al., 2020; Djebali et al., 2012; Iyer et al., 2015), RNA sequencing and ribosome profiling should include micropeptide detection.

Our data show that AHARIBO serves as a flexible tool to detect translated RNAs, to identify lncRNAs bound to elongating ribosomes and to detect *de novo* synthesized proteins. The intersection of standard RIBO-seq, RNA-seq and AHARIBO approaches allowed us to identify translated lncRNAs. We demonstrated that AHARIBO is efficient in capturing short translated open reading frames, both native or ectopically expressed. Although LC-MS technologies are not as sensitive as RNA sequencing, we successfully identified a mouse-specific micropeptide reported to originate form a native lncRNA ORF, confirming the effectiveness of AHARIBO. To overcome existing limitation in LC-MS detection, many other translation events on lncRNAs can be predicted combining AHARIBO approaches with in-silico translation of the

identified leads. This approach would likely allow to selectively validate a list of still uncharacterized lncRNAs. Although the unlabeled background cannot be avoided, a pre-cleaning of the cell lysate with a cushioning step can help to increase the resolution with difficult samples. Moreover, a puromycin treatment instead of sBlock could be added as control in proteomic experiments. A unique feature of AHARIBO is the possibility to simultaneously isolate ribosomes, RNA engaged with ribosomes and the corresponding proteins produced. Beside the versatility of the method, AHA labelling has the advantage of minimal interference with protein synthesis (Hodas et al., 2012; Tom Dieck et al., 2012).

The most prominently limitation of the method relay on the methionine starvation required for efficient AHA incorporation (Calve et al., 2016; Hodas et al., 2012; Saleh et al., 2019). This step can modify the physiological conditions of the cell and need to be taken into consideration when planning experiments requiring certain stimuli (e.g. drug treatment) during methionine depletion. The conditions used in the AHARIBO protocol give robust protein labeling, but AHA concentration can be conveniently tuned based on specific cell types or biological questions.

With AHARIBO we introduce a strategy for the selective isolation of active ribosomes using the nascent peptide chain as bait for more comprehensive interrogation of lncRNA biology and proteogenomic studies. Overall, we provide evidence that AHARIBO is a comprehensive and reliable toolkit suitable for downstream parallel RNA-seq, RIBO-seq and LC-MS analysis, empowering scientists to shed light onto the functional complexity of translation.

SUPPLEMENTARY INFORMATION

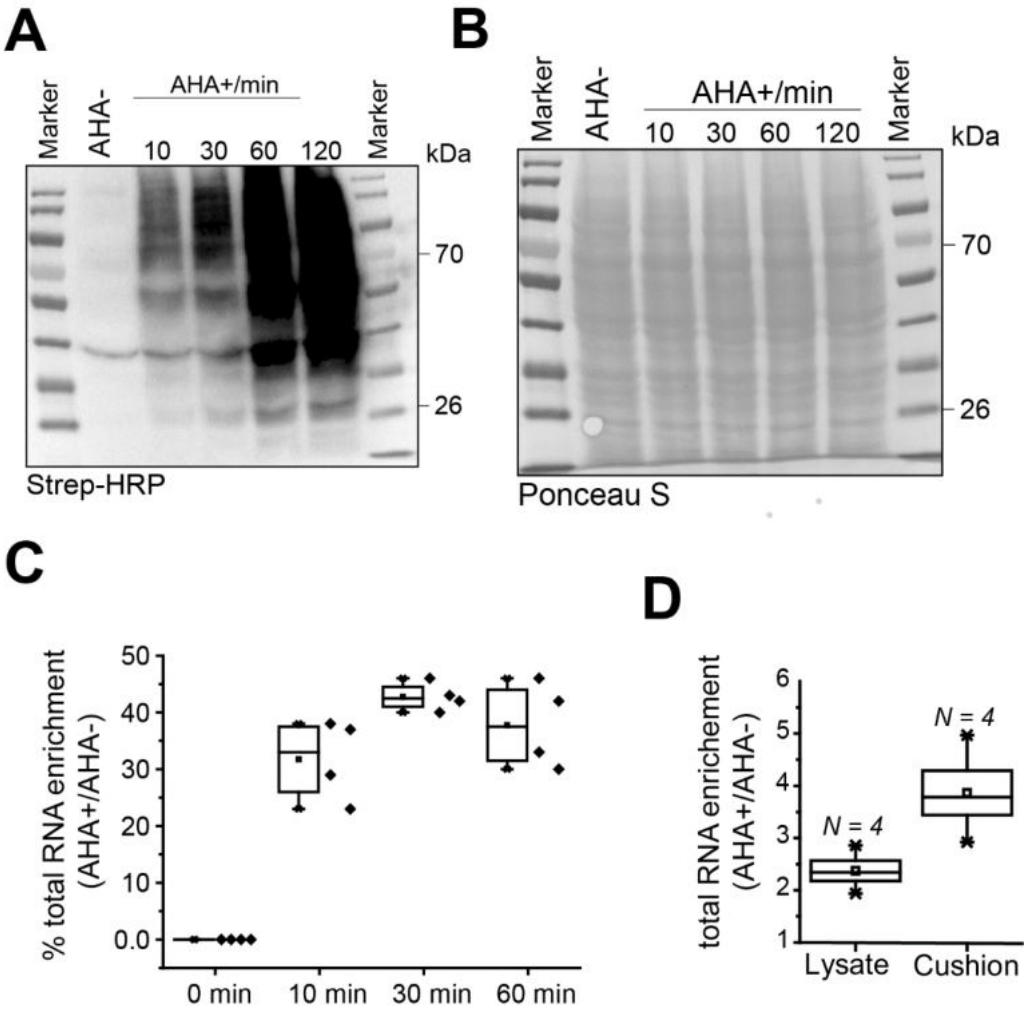

Figure 1-figure supplement 1. AHA incorporation, validation of AHA and RNA capture. (A) Labeling of nascent peptides in cells treated with AHA (250 µM) at different incubation times (10, 30, 60, 120 min). After SDS-PAGE of cell extracts, AHA residues were biotinylated by on-membrane cycloaddition based "click chemistry" and detected by streptavidin-HRP. (B) Ponceaus S staining of the membrane reported in A. (C) RNA enrichment in AHARIBO-rC pulldown at different AHA incubation times (10, 30, 60 min) compared to control (AHA^-) and reported as % of input (1/10 of total RNA). (D) RNA enrichment in AHARIBO-rC pulldown before or after sucrose cushioning compared to control (AHA-).

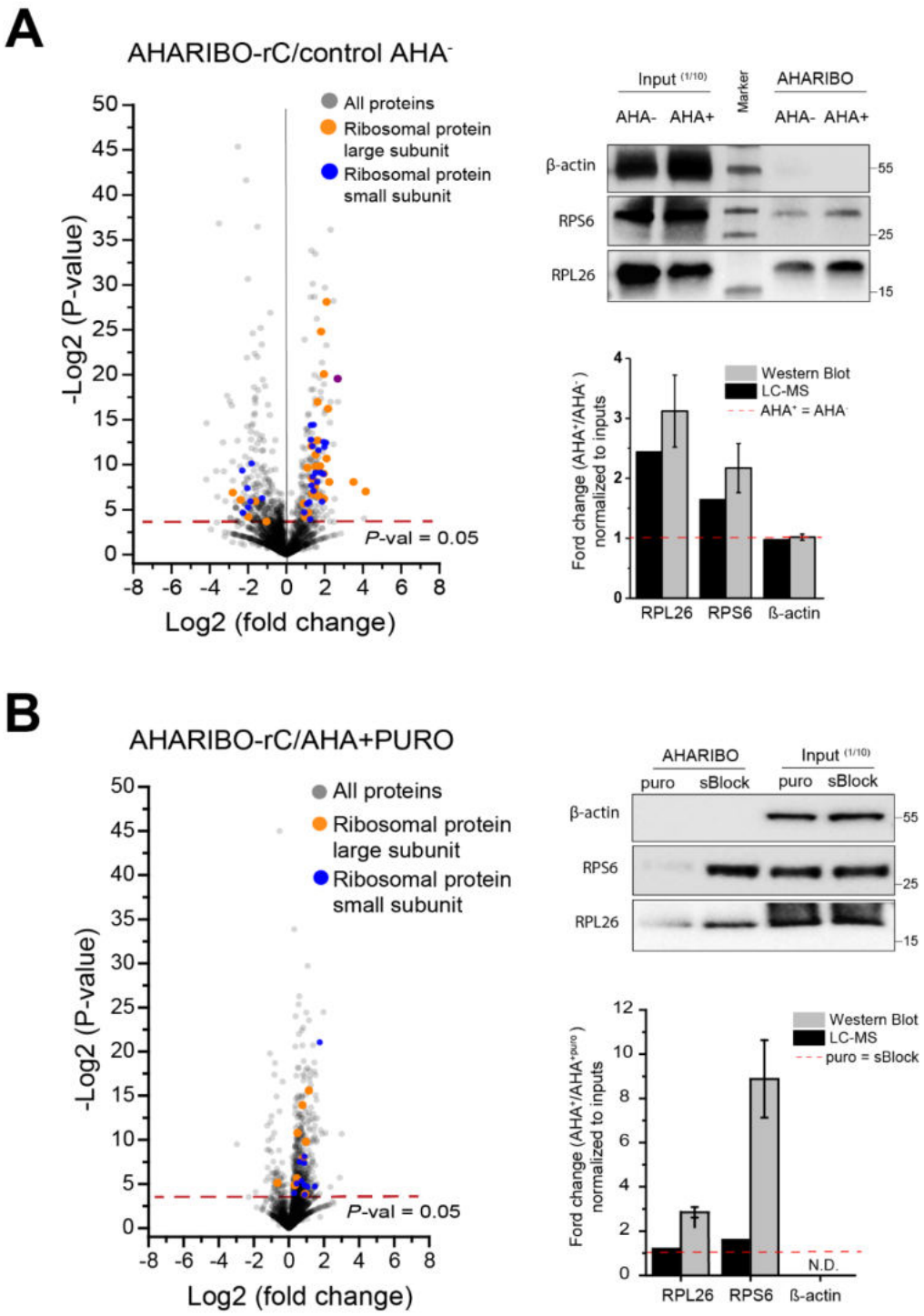

Figure 1-figure supplement 2. LC-MS analysis on AHARIBO-rC proteins and validation by western blot. Volcano plots showing the -Log (p-value) versus the relative abundance of AHARIBO-rC-isolated proteins. Data are compared with the non-specific signal derived from streptavidin-coated beads incubated with lysates from control (AHA-, without L-Azidohomoalanine) (A) and puromycin-treated cells (without sBlock) Red broken line, threshold p-value < 0.05 (B). Western blot on RPL26, RPS6 and Actin with related quantifications of band intensities are reported on the right of each dot blot.

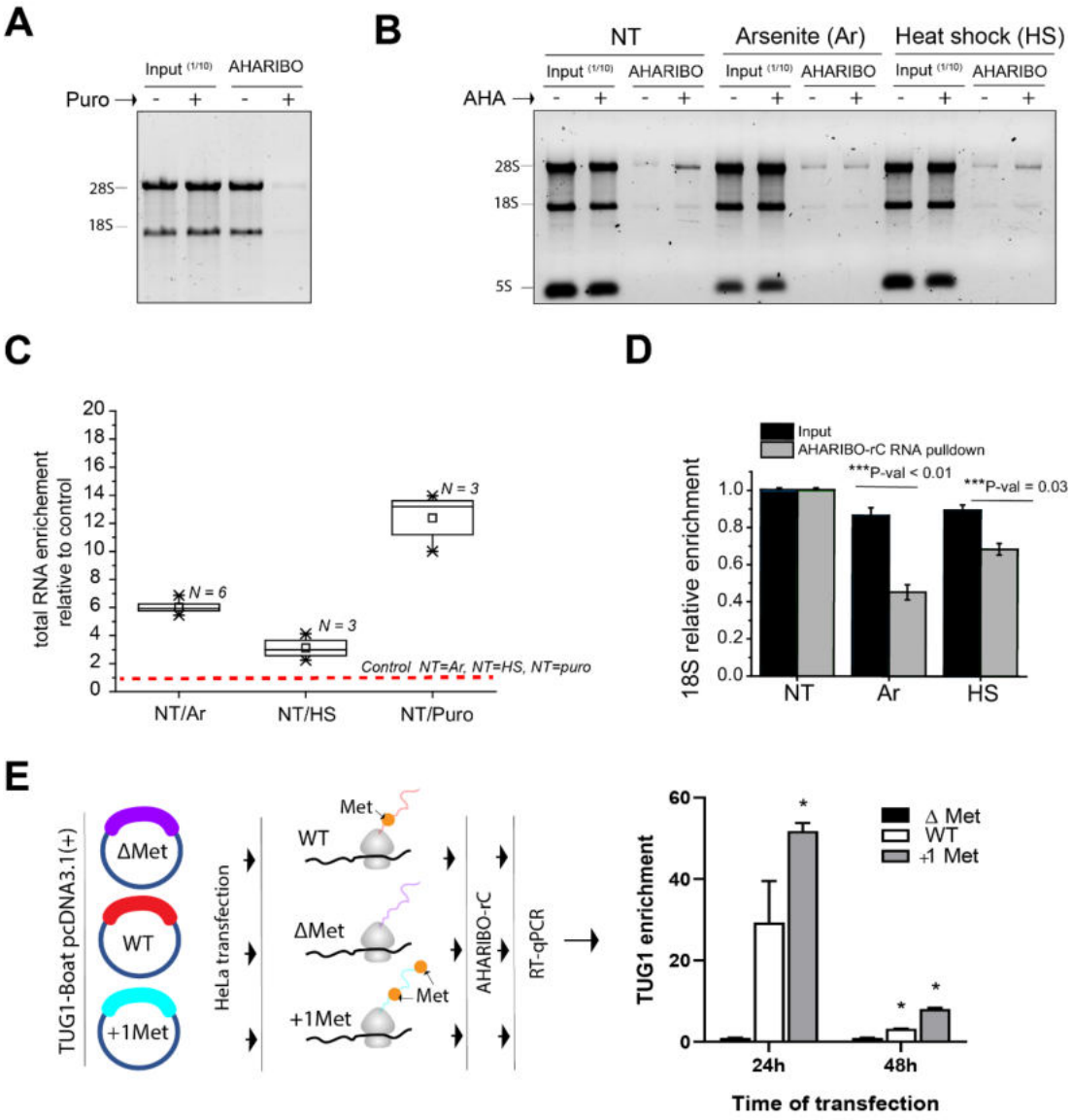

Figure 1-figure supplement 3. AHARIBO-rC efficiency test and validations. (A, B) Agarose gel electrophoresis of total RNA extracted from input lysates (1/10 of the total lysate volume) and lysates subjected to AHARIBO pulldown, obtained from cells either treated or not treated with AHA, with or without puromycin (50 µM) and with different stress. NT, non-treated cells; Ar, arsenite-treated cells, Puro, puromycin treatment (50 µM). Red broken line, no enrichment (C) Total RNA enrichment after AHARIBO-rC pulldown of lysates obtained from unstimulated cells over cells treated with arsenite and heat shock. For each condition, cells were either treated or not treated with AHA. Signal ratios (AHA+/AHA-) for each pulldown sample were normalized to the respective inputs. NT, non-treated. HS, heat shock-treated (42°C for 10 min). Puro, puromycin treatment (50 µM). Square box, mean; stars, 1-99% percentile. (D) 18S rRNA qRT-PCR analysis of RNA extracted from lysates subjected to AHARIBO-rC pulldown and input lysates, obtained from unstimulated cells or cells subjected to arsenite treatment. For each condition, cells were either treated or not treated with AHA. For each sample, 18S AHA+/AHA- signal ratios were normalized to the input and to the housekeeping gene HPRT1. NT, Non-treated; Ar, arsenite. (E) Detection of TUG1-Boat. Scheme of the experimental setup (left) and RT-pPCR enrichment for FUG1-Boat transcript (right) among the three different constructs normalized to the input and for two different time of transfection (24h and 48h) (*, p-value < 0.05 compared with ΔMet).

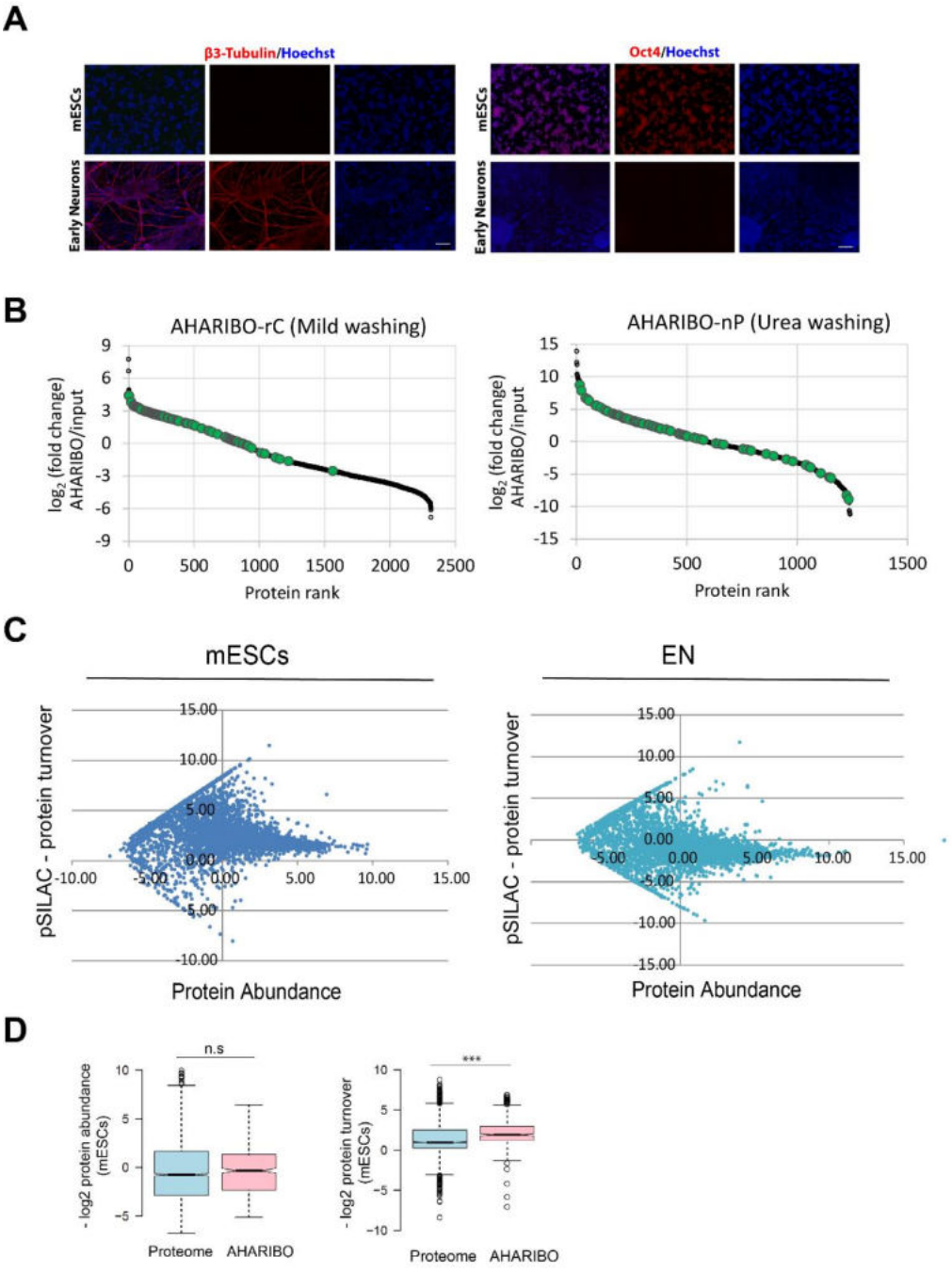

Figure 2-figure supplement 1. Cell differentiation and additional proteomic analysis. (A) Immunofluorescence for mESC (Oct4) and neuronal (β3-tubulin) marker expression on self-renewing mESCs and 15DIV mESC-derived neurons. Scale bar: 200 µm. (B) Rank plot of fold change of full proteome (black dots) and ribosomal proteins (green dots), comparing AHARIBO pulldown versus input samples, mild washing (left) and urea washing (right). Since AHARIBO-rC LC-MS analysis might cause an underestimation of the *de novo* synthesized proteome due to the enrichment of abundant ribosomal proteins, newly synthetized proteins bound to DBCO-conjugate magnetic beads were separated from ribosome subunits by harsh washing conditions (8M urea) before tryptic digestion and LC-MS analysis. The effectiveness of the washing procedure was confirmed, since no evident enrichment of ribosomal proteins in the pull-down was observed**.** (C) The scatter plots represent protein abundance versus protein turnover in mESCs (left) and ENs (right). (D) Normalized protein abundance (left) and turnover distribution (right) as determined by pSILAC and AHARIBO. ***p-value < 0.001.

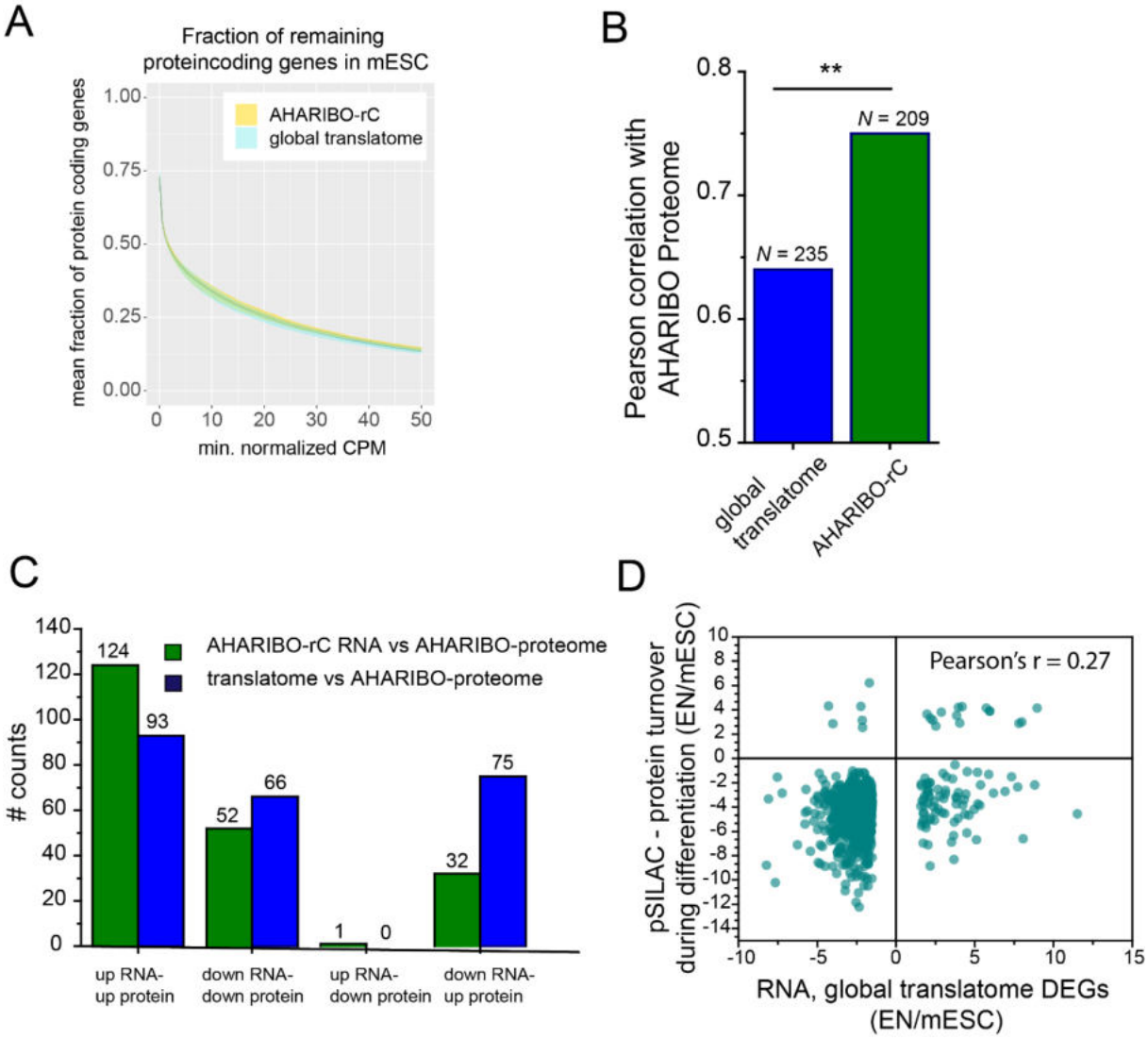

Figure 3-figure supplement 1. RNA-seq and protein coding RNA analysis. (A) Linear plot illustrating the fraction of coding genes (y-axis) expressed above a minimum threshold (x-axis) in mESC. The AHARIBO-rC and the global translatome group are respectively represented in yellow and cyan as indicated. For each group, the mean (solid line) and the SD (shades) of the fractions for a given CPM threshold are calculated over all samples (n = 6) in that group. (B) Histogram showing Pearson's correlation analysis of AHARIBO-nP protein fold change (EN/mESC) determined by mass spectrometry versus global translatome and AHARIBO-rC RNA fold change (EN/mESC) determined by RNA-seq. N, number of DEGs. p-value < 0.05. (C) Histogram of the number of DEGs (EN/mESCs) up- and down- regulated in AHARIBO-rC RNA or global translatome relative to the AHARIBO- nP proteome. (D) Scatter plot of RNA fold change (global translatome) compared to protein turnover (pSILAC).

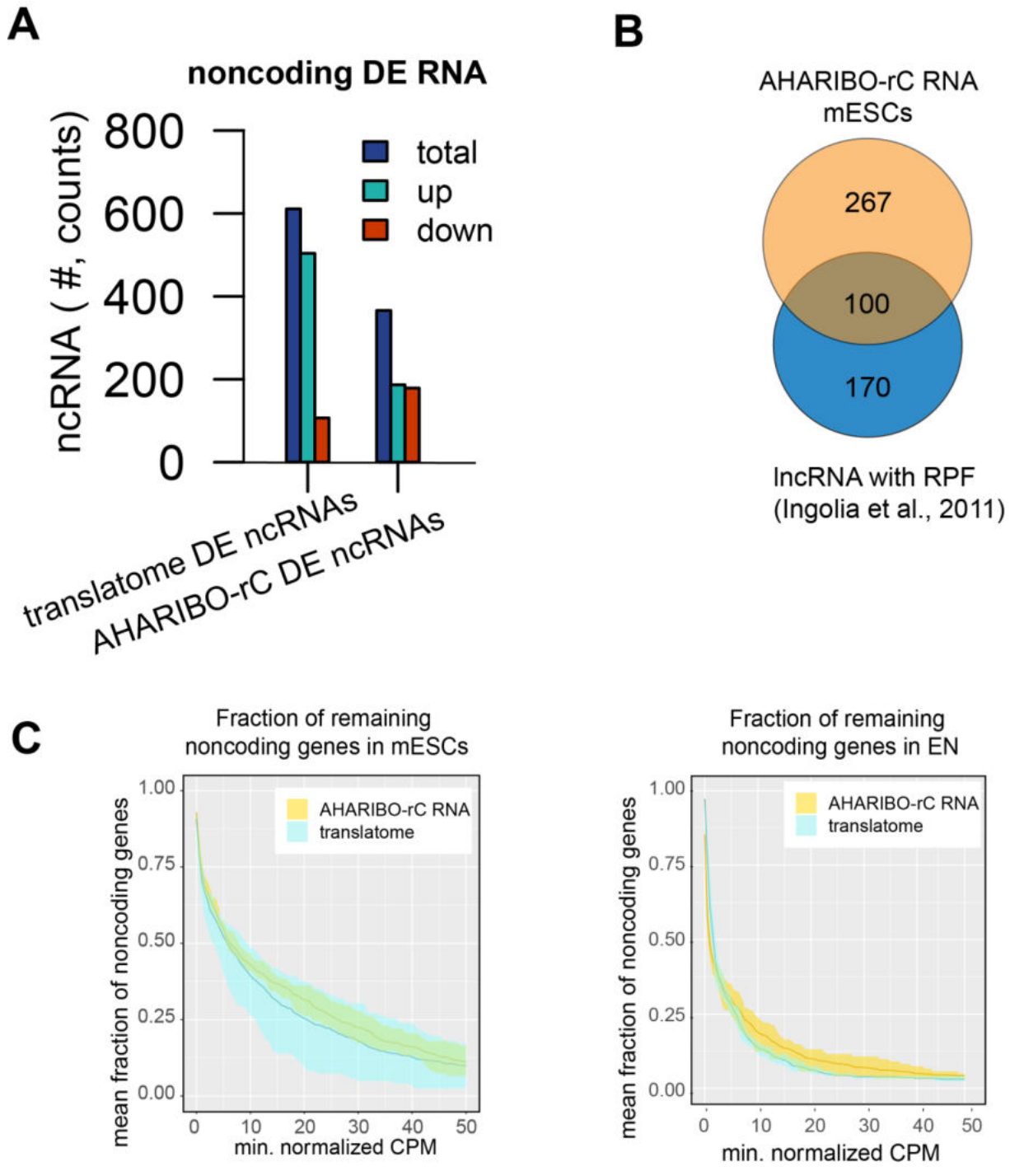

Figure 4-figure supplement 1. **Isolation of lncRNAs with AHARIBO**. (A) Number of up- and down- regulated differentially expressed non-coding RNAs in the global translatome and in AHARIBO-rC RNA DE, differentially expressed. ncRNA, non-coding RNA. (B) Venn diagram representing the number of differentially expressed lncRNAs identified by AHARIBO-rC (orange) and number of lncRNAs with at least 1 CPM in Ingolia et al., 2011 (blue). (C) Linear plot illustrating the fraction of non-coding genes (y-axis) expressed above a minimum threshold (x-axis) in mESC (left) and EN (right). The AHARIBO-rC and the global translatome group are respectively represented in yellow and cyan as indicated. For each group, the mean (solid line) and the SD (shades) of the fractions for a given CPM threshold are calculated over all samples (n = 6) in that group.

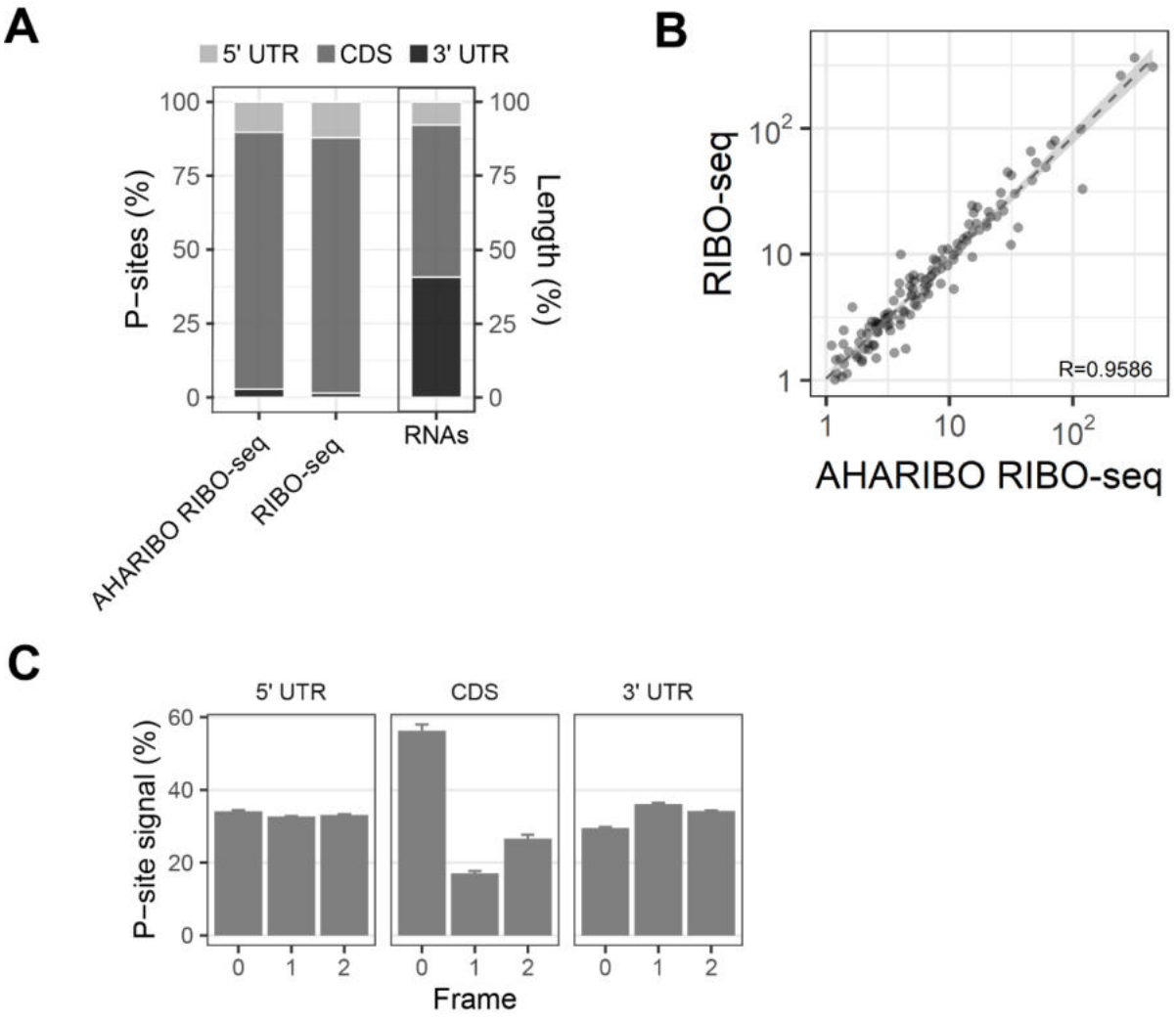

Figure 4-figure supplement 2. AHARIBO RIBO-seq data. (A) Percentage of ribosome P-sites mapping to the 5' UTR, coding sequence (CDS), and 3' UTR of mRNA from AHARIBO RIBO-seq and standard RIBO-seq data. The percentage length of each mRNA region is indicated on the right-hand y axis. (B) Data correlation of AHARIBO Ribo-seq and standard RIBO-seq (performed on the input) obtained in mESCs. Results are representative of two independent replicates for each method. (C) Percentage of P-sites according to the three reading frames for the 5' UTR, the 3' UTR and for CDS for AHARIBO RIBO-seq data, reflecting the codon periodicity along the CDS.

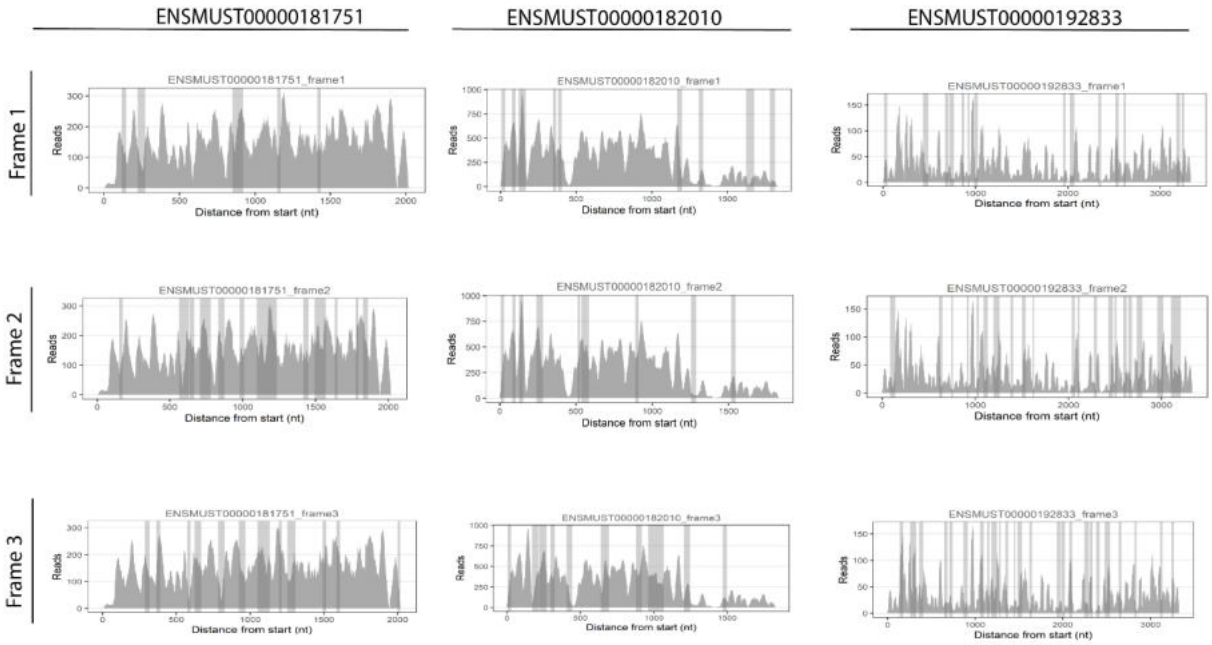

Figure 4-figure supplement 3. Translated lncRNAs. Representative data of three different lncRNAs (from left to right) displaying massive hallmarks of translation along the entire transcript. In silico translation in three different frames (from top to bottom) was performed to predict potential peptide. Shadow area: predicted in silico micriopeptides. The lncRNA reported are representative of a list of translated lncRNA identified by the combination of AHARIBO approaches (between brackets the unique peptide or the number of putative peptides predicted): ENSMUST00000051089 (NSFVNDIFER), ENSMUST00000181328 (KIDNQINLPK), ENSMUST00000181149 (KINQLQNMVKDNK), ENSMUST00000099446 (NLMNVINVVKLLHFS), ENSMUST00000180524 (MSPSQLLELKRNQ), ENSMUST00000182499 (VCVALIINICHIMI), ENSMUST00000134140 (NGGGLLMSYVIK), ENSMUST00000180432 (ELAEQPSSALKTSNREQ), ENSMUST00000181251 (QLTDNQRVNQKA), ENSMUST00000179344 (KELQLK), ENSMUST00000181443 (KGPNDISLAQSYLPI), ENSMUST00000071101 (KNNPPPQNAKPK), ENSMUST00000180407 (IELRENLQTY), ENSMUST00000180489 (EISASANLELNGAPSQQ), ENSMUST00000188038 (LALEELR), ENSMUST00000149246 (LLLPGVIK), ENSMUST00000180396 (23), ENSMUST00000181751 (61), ENSMUST00000182010 (43), ENSMUST00000192833 (94), ENSMUST00000200021 (27), ENSMUST00000223012 (86).

Materials and Methods

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
cell line (*Homo-sapiens*)	Papillomavirus-related endocervical adenocarcinoma	ATCC	RRID:CVCL_0030	
cell line (*M. musculus)*	46C embryonic stem cells	ATCC	RRID:CVCL_Y482	Quattrone A. Lab. (CIBIO)
antibody	anti-β3-tubulin (Mouse monoclonal)	Promega	Cat# #G712A RRID:AB_430874	(1:2000)
antibody	anti-Oct4 (Mouse monoclonal)	Santa Cruz Biotechnologies	Cat# #SC 5279 RRID:AB_628051	(1:2000)
antibody	anti-human RPL26 (Rabbit polyclonal)	Abcam	Cat#: #ab59567 RRID:AB_945306	(1:2000)
antibody	anti-human RPS6 (Rabbit polyclonal)	Abcam	Cat#: #ab40820 RRID:AB_945319	(1:2000)
antibody	anti-human beta actin (Rabbit polyclonal)	Abcam	Cat#: #ab8227 RRID:AB_2305186	(1:2000)
recombinant DNA reagent	WT TUG1-BOAT (plasmid)	PMID: 32894169		

recombinant DNA reagent	Δ TUG1-BOAT (plasmid)	This paper		See Materials and Methods section: Tug1-Boat ectopic expression and qPCR”
recombinant DNA reagent	+1Met TUG1-BOAT (plasmid)	This paper		See Materials and Methods section: Tug1-Boat ectopic expression and qPCR”
peptide, recombinant protein	Precision Protein StrepTactin-HRP Conjugate	BioRad	Cat#: #1610380	(1:5000)
chemical compound, drug	L-Arginine-13C6,15N4 hydrochloride	Sigma-Aldrich	Cat#: #608033	
chemical compound, drug	L-Lysine-13C6,15N2 hydrochloride	Sigma-Aldrich	Cat#: #608041	
chemical compound, drug	L-azidohomoalanine (Click-IT™ AHA)	Invitrogen	Cat. #: #C10102	
chemical compound, drug	Dibenzocyclooctyne-PEG4-biotin conjugate	Sigma Aldrich	Cat. #: #760749SML1656	
chemical compound, drug	sBlock	IMMAGINA BioTechnology	Cat. #: #SM8	

chemical compound, drug	puromycin	Sigma-Aldrich	Cat. #: #P8833	
chemical compound, drug	cycloheximide	Sigma-Aldrich	#C4859	
chemical compound, drug	Lipofectamine ™ 3000 Transfection Reagent	Thermo FisherScientific.	Cat. #: L3000001	
chemical compound, drug	Mag-DBCO beads	IMMAGINA BioTechnology	Cat. #: #MDBCO	
chemical compound, drug	eMagSi-cN beads	IMMAGINA BioTechnology	#018-eMS-001	
commercial assay or kit	SMART-Seq Stranded Kit	Takara	Cat. #: #634443	
commercial assay or kit	SuperScript ™ III Reverse Transcriptase	Thermo Fisher	Cat. #: #18080044	
commercial assay or kit	Kapa Probe Fast Universal qPCR Kit	Kapa Biosystems	#KK4702	
software, algorithm	Image Analysis	ImageJ	RRID:SCR_003070	
software, algorithm	statistical package	edgeR	RRID:SCR_012802	

Cell culturing and treatments

For protocol development, optimization and validation, HeLa cells were used. HeLa cells were maintained on adherent plates in Dulbecco's modified Eagle's medium (DMEM; EuroClone #ECM0728L) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 units/mL penicillin and 100 ug/mL streptomycin at 37°C, 5% CO_2. For passaging, cells were washed with 1x PBS, detached using 0.25% trypsin-EDTA and spun down at 260 x g for 5 minutes.

For treatments, 250,000-400,000 HeLa cells per well were seeded in 6-well plates and grown to 80% confluence. At the time of treatment, culture medium was removed and cells were washed once with warm 1x PBS. Subsequently, cells were incubated with Dulbecco's modified Eagle's limiting medium (DMEM-LM; Thermo Scientific #30030) supplemented with 10% fetal bovine serum and 800 µM L-leucine for 40 min to deplete methionine reserves. Methionine-free medium was then supplemented with L-azidohomoalanine (Click-IT™ AHA; Invitrogen #C10102) at a final concentration of 250 µM and incubation time (ranging from 10 min to 120 min; 30 min set as incubation time for the protocol). Cells were then treated with 1x sBlock® (IMMAGINA BioTechnology, catalog n° #RM8; sBlock is an anisomycin-containing proprietary reagent) for 10 min. Then, 6-well plates were placed on ice, medium was removed and cells were washed once with cold 1x PBS supplemented with 1x sBlock. After removing residual PBS with a pipette, hypotonic lysis buffer (0.01 M NaCl, 0.01 M $MgCl_2$, 0.01M Tris HCL, 1% Tx-100, 1x sBlock, 1% sodium deoxycholate, 5 units/mL DNAse I - Thermo Scientific #89836, 200 units/mL RiboLock RNase Inhibitor - Thermo Scientific #EO0381, 1x Protease Inhibitor Cocktail - Cell Signaling Technology #5871S) was added to each well, and cells were lysed with the aid of a scraper. After hypotonic lysis, nuclei and cellular debris were removed by centrifuging at 18000 x g, 4°C for 5 min. For quantification of the total absorbance value of cell lysates, the absorbance was measured (260 nm) using a Nanodrop ND1000 UV-VIS Spectrophotometer. Lysates were aliquoted and processed directly or stored at -80°C.

Arsenite pre-treatment was performed adding sodium arsenite (Sigma-Aldrich #S7400) at a final concentration of 500 µM for 1 hour.

For RNA-seq and proteomics experiments, two biological settings were assessed in triplicate experiments: i) undifferentiated mouse 46C embryonic stem cells (mESCs) (Ying et al., 2003) and ii) mESCs induced to differentiate into early neurons (ENs).

mESCs were maintained in mESC self-renewal medium composed of Glasgow's MEM (Thermo Scientific #11710-035) supplemented with 1000 units/ml ESGRO Recombinant Mouse LIF protein (Millipore #ESG1107), 10% fetal bovine serum, 55 µM 2-Mercaptoethanol, 1mM Sodium Pyruvate (Thermo Scientific #11360070), MEM Non-Essential Amino Acids (Thermo Scientific #11140050), GlutaMax (Thermo Scientific #35050061), and penicillin/streptomycin. For passaging, mESCs were washed twice with 1x PBS, detached using 0.02-0.05% trypsin-EDTA and spun down at 260 x g for 3 minutes. Pellet was resuspended in fresh medium and plated onto 0.1% gelatin-coated culture vessels.

For treatments, 5 x 10^5 mESCs/cm^2 were seeded in petri dishes and grown to 60% confluence. For pSILAC proteomics, 24 h before lysis mESCs were washed twice with 1x PBS and the medium was replaced with SILAC Advanced DMEM/F-12 Flex Medium (Thermo Scientific #A2494301), supplemented with 1000 units/ml ESGRO Recombinant Mouse LIF protein, 10% dialyzed fetal bovine serum, 4500 mg/L glucose, 17.25 mg/L proline, and penicillin/streptomycin. Either light or heavy L-arginine (Sigma-Aldrich #608033) and L-lysine (Sigma-Aldrich #608041) were added at 84 mg/L and 146 mg/L, respectively. For both AHA+ proteomics and RNA-seq experiments, treatments were performed as described above for HeLa cells, with the exception that methionine-free medium was supplemented with 1000 units/ml ESGRO Recombinant Mouse LIF protein and 10% dialyzed fetal bovine serum. After methionine depletion, cells were treated with 250 µM AHA for 30 min. The remaining treatment steps and hypotonic lysis were performed as detailed above.

Neuronal differentiation was performed according to a previously described protocol(Ying et al., 2003). Briefly, 2.000 mESCs/cm^2 were seeded on gelatin-coated culture vessels in N2B27 medium. Cells were gently washed with 1x PBS and medium was renewed every 1-2 days until 15DIV. N2B27 medium is composed of 1:1 mix of DMEM/F-12 (Thermo Scientific #21331020) and Neurobasal Medium (Thermo Scientific #21103049), supplemented with 0.5% N-2 (Thermo Scientific #17502048), 1% B-27 (Thermo Scientific #17504044), Glutamax, and penicillin/streptomycin.

Upon differentiation, early neurons were treated directly in culture vessels. For pSILAC proteomics, 24 h before lysis ENs were washed once with 1x PBS and the medium was replaced with SILAC Advanced DMEM/F-12 Flex Medium, supplemented with 0.5% N2, 1% B27, 4500 mg/L glucose, 17.25 mg/L proline, and penicillin/streptomycin, 4500 mg/L glucose, 17.25 mg/L proline, and penicillin/streptomycin. Either light or heavy L-arginine and

L-lysine were added at 84 mg/L and 146 mg/L, respectively. For both AHA+ proteomics and RNA-seq experiments, ENs were treated as described above for HeLa cells, with 250 µM AHA for 30 min. The remaining treatment steps and hypotonic lysis were performed as detailed above.

Cell lines were purchased directly from ATCC and passaged fewer than 15 times. M. musculus 46C ES were obtained from Quattrone A. Lab (CIBIO), RRID: CVCL_Y482. All cells tested negative for mycoplasma contamination.

Immunocytochemistry

For immunofluorescence assay, cells were fixed with 4% paraformaldehyde for 15 min at room temperature, permeabilized using 0.5% Triton X-100 in 1x PBS for 15 min at room temperature and blocked using 5% fetal bovine serum, 0.3% Triton X-100 in 1x PBS for 2h at room temperature. Cultures were then incubated overnight at 4°C with either anti-β3-tubulin (Promega #G712A) or anti-Oct4 (Santa Cruz Biotechnologies #SC-5279) primary antibodies diluted in 2% fetal bovine serum, 0.2% Triton X-100 in 1x PBS. Cells were then washed three times with 1x PBS and incubated with Alexa-555 anti-mouse secondary antibodies for 2h. Nuclei were counterstained with Hoechst 33258 before imaging with a Zeiss Axio Observer Z1 inverted microscope equipped with a 2.83 Megapixel AxioCam 503 mono D camera.

AHARIBO-rC/AHARIBO-nP: purification of active ribosomes for RNA/protein isolation

For RNAseq experiments, lysates were diluted in W-buffer (10mM NaCl, 10mM $MgCl_2$, 10 mM Hepes, 1x sBlock) to a final Nanodrop-measured absorbance (260 nm) of 1-2 a.u./mL, supplemented with 40 U of Superase-In RNase Inhibitor (Thermo Scientific #AM2696) and incubated with Dibenzocyclooctyne-PEG4-biotin conjugate (Sigma-Aldrich #760749; 50µM final concentration) in a reaction volume of 100 µl for 1 h on a rotator in slow motion (9 rpm) at 4°C. Lysates were then incubated with 50 µl of eMagSi-cN beads (IMMAGINA BioTechnology #018-eMS-001) for 30 min at 4°C on the rotator in slow motion (9 rpm). Subsequently, samples were taken off the rotator and placed on a magnetic rack on ice, and supernatants were discarded. Beads were washed two times with 500 µl of 1x PBS supplemented with 0.1% Triton-X100, 1x sBlock and 1:10,000 RiboLock RNase Inhibitor (Thermo Scientific #EO0381) on the rotator in slow motion at 4°C, removing supernatants

from the tubes sitting on the magnetic rack and gently adding new washing solution each time. After the final wash, beads were resuspended in 200 µl of W-buffer and transferred to a new vial. Then, 20 µl of 10% SDS and 5 ul of Proteinase K (Qiagen #19131) were added to each sample, and samples were incubated at 37°C for 75 min in a water bath. Subsequently, suspensions were transferred to a new vial and acid phenol:chloroform:isoamyl alcohol RNA extraction was performed. Briefly, an equal volume of acid phenol:chloroform:isoamyl alcohol (pH 4.5) was added, and samples were vortexed and centrifuged at 14,000 x g for 5 min. Aqueous phases were then transferred to new vials, 500 µl of isopropanol and 2 µl of GlycoBlue (Thermo Scientific #AM9516) were added, samples were mixed and incubated at room temperature for 3 min and then stored overnight at - 80 °C. The following day, samples were centrifuged at 14,000 x g for 30 min, supernatants were removed, 500 µl of 70 % ethanol were added to each sample and samples were then centrifuged at 14,000 x g for 10 min. Finally, pellets were air-dried and resuspended in 10 µl of nuclease-free water. When quality check and quantification was needed, RNA samples were run on a 2100 Bioanalyzer (Agilent) using the Agilent RNA 6000 Nano Reagents kit (Agilent #5067-1511) and assayed on the Qubit fluorometer using the Qubit RNA HS Assay Kit (Thermo Scientific # Q32852). For visualization of total RNA patterns, samples were run on a 1% agarose gel. ImageJ software (v 1.45s) was used for quantitation of signal intensities of ribosomal RNA bands.

For proteomics experiments, lysates were diluted in W-buffer to a final Nanodrop-measured absorbance (260 nm) of 1-2 a.u./mL in a final volume of 100 ul. Ribosome pulldown was performed using Mag-DBCO beads (IMMAGINA BioTechnology #MDBCO). Lysates were incubated with 50 µl of beads for 1h on a rotator in slow motion (9 rpm) at 4°C. Supernatants were discarded after placing samples on the magnetic rack. Beads were washed three times with 500 µl of 200 mM Tris, 4 % CHAPS, 1 M NaCl, 8M Urea, pH 8.0 at room temperature on a shaker at 1000 rpm, using the magnetic rack to replace the washing solution. After the final wash, beads were resuspended in 30 µl of water and transferred to a new vial.

qRT-PCR analysis

Total RNA was extracted from samples processed through the AHARIBO-rC protocol as described above. Depending on the available input material, RNA was retrotranscribed using either RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific #K1621) or

SuperScript™ III Reverse Transcriptase (Thermo Fisher #18080044), per manufacturer protocols. qPCR was run on CFX Connect Real-Time PCR Detection System (BioRad) using Kapa Probe Fast Universal qPCR Kit (Kapa Biosystems #KK4702). Reactions were performed in technical duplicates of biological triplicates. The following TaqMan probes were used: Hs99999901_s1 (18S), Hs02800695_m1 (HPRT1).

For normalization of qRT-PCR results, HPRT1 was used as housekeeping gene. The fold change in normalized 18S RNA levels between untreated (control) and treated (arsenite) samples was calculated. A second normalization to threshold cycles from non-AHA-treated samples was done to account for background signal.

Tug1-Boat ectopic expression and qPCR

We ectopically express the putative protein produced by the open reading frame of TUG1, called TUG1-BOAT (Tug1-Bifunctional ORF and Transcript) in HeLa cells. Briefly, construct generation and transfection was performed as in (Lewandowski et al., 2020) with some minor changes to adapt the experimental setup to the AHARIBO method. We synthesized three different constructs for human Tug1 ORF1 (Thermo Scientific):

1. The first (called *WT TUG1-BOAT*) is the one reported in Lewandowski J.P. et al., 2020. It has a non-canonical start codon and a methionine at 75 nt (25 aa) upstream the stop codon.
2. The second (called Δ TUG1-BOAT), is deleted by the only methionine of the sequence present at 75 nt from the stop codon. No methionines are present.
3. The third (called +1Met TUG1-BOAT), has an ATG start codon (methionine) instead of the non-canonical CTG start codon e. Therefore, the third construct has two methionines, one at the N terminal and a second one at 25 aa (about 75 nt) upstream the C-terminal.

We cloned the constructs in the pcDNA3.1(+) plasmid with Hind III and Eco RV restriction enzymes. For transfection of TUG1-BOAT constructs we seeded HeLa cells in a 6-well plate and we transfected the cells with 2.5 μg of plasmids (pcDNA3.1(+) containing each of the inserts) using 742 Lipofectamine™ 3000 Transfection Reagent (Thermo Fisher Scientific). After 24 or 48 hours post transfection, cells were processed with AHARIBO-rC protocol followed by RNA extraction.

We performed qPCR analysis on AHARIBO pull-downs and input for each vector to validate the efficiency in capturing short translated ORF deriving from RNA annotated as lncRNA (TUG1). Briefly, 200 ng of DNase I-treated RNA was used as input to generate cDNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems), according to the manufacture protocol. qPCR was run on CFX Connect Real-Time PCR Detection System (BioRad) using Powerup Sybr Master Mix (Applied Biosystems) and a couple of primers design to amplify 150 nt of the CDS of TUG1-BOAT transcript (see below). Reactions were performed in technical duplicates of biological duplicates. For normalization, 18S was used as housekeeping gene. Ct values were analysed using the ΔΔCt method (Livak and Schmittgen, 2001).

Fw PRIMER: GGCTCTTTCTCCTGCTCTGG

Rev PRIMER: CTCCTCGTCGAATCGCAAAC

Insert size: 150 nt

TUG1-BOAT sequences are listed below

Italic: 5' UTR leader sequence. Bold, canonical and not canonical start codons. Red, methionine.

> WT TUG1-BOAT

GGCCGAGCGACGCAGCCGGGACGGTAGCTGCGGTGCGGACCGGAGGAGCCATCTTGTCTCGTCGCCGGGGAGTCA
*GCCCCTAAATCGAAGAAGCC***CTG**GCGCGCCCTCCCCCCCTCCCGGGTCTGGTAGGGCGAAGGAACGGGCGTGCGG
TCGATCGAGCGATCGGTTGGCGGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGGCCGCCGC
GGCGCGCGCCCGGTGGCCGTTGGCGCTCGCGCCGCGTCTTTCTTCTCGTACGCAGAACTCGGGCGGCGGCCTATG
CGTTTGCGATTCGACGAGGAGTCGTCCGGGTGGTCGGCGGCGGCGGGCAGCTGCTCCGCCCCGCTCCGGGGGAGG
CGGCGGCGGCAGCGGCCGCGGGATTTGGAGCGGCCGGGGAGGCGGGGGTGGCCGGGGCCGGCTTGGAGGCCTGGC
GCCACCCTTCGGGGCCTGCAAGGACCCAGTTGGGGGGGCAGGAGGGGGCCGGAGGATGGTTGGTTGTGGGATTTC
TACTTTGCCTTTTCCTCCTT**ATG**CCGCCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACA*TCGACT*
*ACAAGGATGACGATG*ACAAGTAG

> ΔTUG1-BOAT

GGCCGAGCGACGCAGCCGGGACGGTAGCTGCGGTGCGGACCGGAGGAGCCATCTTGTCTCGTCGCCGGGGAGTCA
*GCCCCTAAATCGAAGAAGCC***CTG**GCGCGCCCTCCCCCCCTCCCGGGTCTGGTAGGGCGAAGGAACGGGCGTGCGG
TCGATCGAGCGATCGGTTGGCGGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGGCCGCCGC
GGCGCGCGCCCGGTGGCCGTTGGCGCTCGCGCCGCGTCTTTCTTCTCGTACGCAGAACTCGGGCGGCGGCCTATG
CGTTTGCGATTCGACGAGGAGTCGTCCGGGTGGTCGGCGGCGGCGGGCAGCTGCTCCGCCCCGCTCCGGGGGAGG
CGGCGGCGGCAGCGGCCGCGGGATTTGGAGCGGCCGGGGAGGCGGGGGTGGCCGGGGCCGGCTTGGAGGCCTGGC
GCCACCCTTCGGGGCCTGCAAGGACCCAGTTGGGGGGGCAGGAGGGGGCCGGAGGATGGTTGGTTGTGGGATTTC

TACTTTGCCTTTTCCTCCTTCCGCCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACA*TCGACTACA*
*AGGATGACGATGA*CAAGTAG

> +1Met TUG1-BOAT

GGCCGAGCGACGCAGCCGGGACGGTAGCTGCGGTGCGGACCGGAGGAGCCATCTTGTCTCGTCGCCGGGGAGTCA
*GGCCCCTAAATCGAAGAAGCC***ATG**GACTACAAGGATGACGATGACAAGGCGCGCCCTCCCCCCCTCCCGGGTCTG
GTAGGGCGAAGGAACGGGCGTGCGGTCGATCGAGCGATCGGTTGGCGGCTCTTTCTCCTGCTCTGGCATCCAGCT
CTTGGGGCGCAGGCCCGGCCGCCGCGGCGCGCGCCCGGTGGCCGTTGGCGCTCGCGCCGCGTCTTTCTTCTCGTA
CGCAGAACTCGGGCGGCGGCCTATGCGTTTGCGATTCGACGAGGAGTCGTCCGGGTGGTCGGCGGCGGCGGGCAG
CTGCTCCGCCCCGCTCCGGGGGAGGCGGCGGCGGCAGCGGCCGCGGGATTTGGAGCGGCCGGGGAGGCGGGGGTG
GCCGGGGCCGGCTTGGAGGCCTGGCGCCACCCTTCGGGGCCTGCAAGGACCCAGTTGGGGGGGCAGGAGGGGGCC
GGAGGATGGTTGGTTGTGGGATTTCTACTTTGCCTTTTCCTCCTT**ATG**CCGCCTGACTACAAAGACCATGACGGT
GATTATAAAGATCATGACATCTAG

RNA-seq

RNA samples were subjected to library preparation for the Illumina platform using the SMART-Seq Stranded Kit (Takara #634443) as manufacturer instructions, using 5 ng of RNA as starting material. For quality check and quantification, the final libraries were run on a 2100 Bioanalyzer (Agilent) using the Agilent DNA 1000 Kit (Agilent #5067-1504) and assayed on the Qubit fluorometer using the Qubit dsDNA HS Assay Kit (Thermo Scientific # Q32851). Libraries were sequenced on an Illumina HiSeq2500 by the NGS Core Facility (University of Trento).

Polysome profiling

HeLa cells were treated and lysed as described above, adding one the following blocking drugs: i) sBlock (IMMAGINA BioTechnology #SM8, final concentration 1x, 10 min treatment); ii) cycloheximide (Sigma-Aldrich #C4859; final concentration 30 µM, 5 min treatment); iii) puromycin (Sigma-Aldrich #P8833; final concentration 50 µM, 5 min treatment); iv) no blocking drug. Cleared supernatants obtained from cytoplasmic lysates were loaded on a linear 15%–50% sucrose gradient and ultracentrifuged in a SW41Ti rotor (Beckman) for 1 h and 40 min at 180,000 x g at 4°C in a Beckman Optima LE-80K Ultracentrifuge. After ultracentrifugation, gradients were fractionated in 1 mL volume fractions with continuous absorbance monitoring at 254 nm using an ISCO UA-6 UV detector. Each fraction was flash-frozen in liquid nitrogen and stored at - 80°C for subsequent protein extraction.

Polysome profiles were analysed as follows. The relative intensity of each individual fraction was determined for both on-membrane AHA and RPL26 signals, then the AHA/RPL26 relative intensity ratio was calculated for each fraction. For each profile, the relative intensity ratios of polysome-containing fractions (fractions 8/9-10/11) were averaged and normalized to the relative intensity ratio of the 60S fraction, which was chosen as internal baseline for background signal based on the fact that it should be devoid of translationally active ribosomes. To assess the effect of the different blocking drugs, averaged normalized relative intensity ratios for the profiles obtained from different blocking drugs and from the untreated control sample were compared. ImageJ software (v 1.45s) was used for quantitation of signal intensities of protein bands.

Sucrose cushioning for ribosome enrichment (global translatome)

HeLa cells were treated in petri dishes and lysed as described above, adding 1x sBlock as blocking drug. Sucrose cushioning was performed according to a modified version of a previously described protocol (Nicholas T Ingolia et al., 2012). For each sample, a volume of cell lysate corresponding to 1.7 a.u. (based on Nanodrop measurement of absorbance at 260 nm) was layered on top of 900 µl of 30% sucrose cushion (30 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 1M sucrose in nuclease-free water) supplemented with 1x sBlock. Samples were ultracentrifuged at 95,000 rpm at 4°C for 1 h and 40 min using a $TLA_{100.2}$ rotor (Beckman). Pellets were resuspended in 100 µl of nuclease-free water supplemented with 30 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$.

Protein extraction from sucrose gradient fractions

Polysomal fractions (1 mL) or pellet/supernatant fractions from 30% sucrose cushioning ($1/5^{th}$ of total amount, adjusted to 260 µl volume) were processed for methanol/chloroform protein extraction. Briefly, 600 µl of methanol and 150 µl of chloroform were added to each sample and samples were vortexed. Then, 450 µl of deionized water were added to each sample and samples were vortexed again. Samples were centrifuged at 14,000 x g for 1 min at room temperature and the resulting aqueous phase was removed without disrupting the underlying white ring (protein interface). Subsequently, 450 µl of methanol were added to each sample, samples were vortexed and then centrifuged at 14,000 x g for 2 min at room

temperature. After centrifugation, supernatants were removed and pellets air-dried. Finally, pellets were resuspended in deionized water supplemented with Pierce Lane Marker Reducing Sample Buffer (Thermo Scientific #39000) to a final volume of 15 µl and either stored at -80°C or heated at 95°C and directly used for SDS-PAGE.

On-membrane click chemistry

Cell lysate or protein extracts obtained from sucrose gradient fractions were supplemented with Pierce Lane Marker Reducing Sample Buffer (Thermo Scientific #39000), heated at 95°C for 10 min and separated by SDS-PAGE. Separated proteins were transferred to nitrocellulose membranes, then membranes were blocked overnight at 4°C in 5% milk prepared in 1x TBS - 0.1% Tween20 supplemented with Dibenzocyclooctyne-PEG4-biotin conjugate (Sigma-Aldrich #760749; 50 µM final concentration). Membranes were washed three times in 1x TBS - 0.1% Tween20 for 10 min each, then incubated with Precision Protein StrepTactin-HRP Conjugate (BioRad #1610380; 1:1000 in 5 % milk prepared in 1x TBS - 0.1 % Tween20) for 1 h at room temperature, then washed again. Membranes were subsequently developed using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare #RPN2236). Images were acquired through the ChemiDoc MP Imaging System. ImageJ software (v 1.45s) was used for quantitation of AHA signal intensities.

Immunoblotting

Aliquots of 10-20 µl of cell lysate or protein extracts obtained from sucrose gradient fractions were supplemented with Pierce Lane Marker Reducing Sample Buffer (Thermo Scientific #39000), heated at 95°C for 10 min and separated by SDS-PAGE. Separated proteins were transferred to nitrocellulose membranes, then membranes were blocked for 1 h at room temperature in 5% milk prepared in 1x TBS - 0.1% Tween20. Membranes were subsequently incubated for 1 h at room temperature with the following primary antibodies, diluted in 5% milk prepared in 1X TBS - 0.1% Tween20: anti-RPL26 (Abcam #ab59567; 1:2000), anti-RPS6 (Abcam #ab40820; 1:1000), anti-beta-actin (Abcam #ab8227; 1:2000). Membranes were washed three times in 1x TBS - 0.1% Tween20 for 10 min each, then incubated with the appropriate HRP-conjugated secondary antibodies for 1h at room temperature and washed again as before. Membranes were then developed using either

Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare #RPN2236) or SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific #34095), depending on signal intensities. Images were acquired through the ChemiDoc MP Imaging System. ImageJ software (v 1.45s) was used for quantitation of signal intensities of protein bands.

RNA-seq data analysis

FASTQ files from Illumina HiSeq2500 were firstly checked for adapters and quality-base distribution using FASTQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) followed by trimming with Trimmomatic-0.35(Bolger et al., 2014) with the following settings: ILLUMINACLIP:ADAPTOR_FILE:2:30:8:1 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15. Prior to sequencing data processing, technical replicates (different sequencing lanes) from the same library were merged together generating a unique FASTQ per sample. Reads were aligned onto mm10 Mouse genome using STAR-2.6.0a (Dobin et al., 2013) with a maximum mismatch of two and default setting for all other parameters. Once uniquely mapped reads were selected, the GRCm38.92 mouse gene annotation from Ensembl (www.ensembl.org) was incorporated in the HTSeq-count v0.5.4 (Anders et al., 2015) tool to obtain gene-level counts. Genes with CPM (Counts Per Million) < 1 in all samples were considered as not expressed and hence removed from the analysis. TMM (Trimmed Mean of M values) (Robinson and Oshlack, 2010) normalization and CPM conversion were then performed to obtain gene expression levels for downstream analyses. For each comparison, differential expression testing was performed using the edgeR-3.20.9 (Robinson et al., 2010) statistical package from Bioconductor. According to the edgeR approach, both common (all genes in all samples) and separate (gene-wise) dispersions were estimated and integrated into a Negative Binomial generalized linear model to moderate gene variability across samples. Finally, genes having a Log fold-change higher/smaller than 1.5/-1.5 and a FDR-corrected p-value of 0.01 (or smaller) were considered as significant.

Proteomics experiments

Proteomic analysis was performed on samples processed through the pSILAC and AHARIBO workflows, as described above. For pSILAC experiments, cells were prepared as described above (see Cell culturing and treatments). 50 µg of lysates were then subjected to acetone precipitation and protein pellets were dissolved in 50 mM ammonium bicarbonate and 6M urea. For AHARIBO enrichment, the beads used for ribosome pulldown were reconstituted in 100 µl 6M urea with 50 mM ammonium bicarbonate.

Samples were reduced using 10 mM DTT for 1 h at room temperature and alkylated with 20 mM iodoacetamide in the dark for 30 min at room temperature. Subsequently, proteins were digested at room temperature with 0.5 µg Lys-C (Promega, # VA1170) for 4 h, after which the solution was diluted 4 times in 50 mM ammonium bicarbonate. 1 µg of trypsin (Promega, # V5111) was then added to the samples and proteolysis was carried out overnight. Digestion was interrupted by adding 1% trifluoroacetic acid. Samples were then desalted by C18 stage-tip, lyophilized, and resuspended in 20 µl of buffer A (0.1% formic acid) for LC-MS/MS analysis.

Samples were analysed using an Easy-nLC 1200 system coupled online with an Orbitrap Fusion Tribrid mass spectrometer (both Thermo Fisher Scientific). Peptides were loaded onto a 25 cm long Acclaim PepMap RSLC C18 column (Thermo Fisher Scientific, 2 µm particle size, 100 Å pore size, id 75 µm) heated at 40 °C. For pSILAC samples, the gradient for peptide elution was set as follow: 5- to 25 % buffer-B (80 % acetonitrile, 0.1 % formic acid) over 90 min, 25 to 40 % over 15 min, 40 % to 100 % over 18 min and 100 % for 17 min at a flow rate of 400 nl/min. For AHARIBO pulldown samples, the same steps for peptide elution were set over a total gradient of 80 min. The instrument was set in a data-dependent acquisition mode. The full MS scan was 350-1100 m/z in the orbitrap with a resolution of 120,000 (200 m/z) and an AGC target of $1x10^6$. MS/MS was performed in the ion trap using the top speed mode (3 secs), an AGC target of $5x10^3$ and an HCD collision energy of 30.

MS raw files were analysed by using Proteome Discoverer (v2.2, Thermo Scientific). MS/MS spectra were searched by the SEQUEST HT search engine against the human or the mouse UniProt FASTA databases (UniProtKB 11/2018). Trypsin was specified as the digestive enzyme. Cysteine carbamidomethylation (+57.021 Da) was set as fixed modification, methionine oxidation (+15.995 Da) and N-term acetylation (+42.011 Da) as variable

modifications. SILAC labeling (Lys +8.014 Da, Arg +10.008 Da) was used as quantification method for pSILAC samples. All other values were kept as default.

Proteomics data analysis

Heteroscedastic T-test was used to assess the significant differences in peptide/protein abundance (p-value lower than 0.05) unless stated otherwise. Data distribution was assumed to be normal but this was not formally tested. Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis were performed using DAVID version 6.8, PANTHER 14.1 and Enrichr (http://amp.pharm.mssm.edu/Enrichr/).

Identification of lncRNA peptides from result spectra

Sequenced non-coding RNAs were in-silico translated into amino acid sequences using the EMBOSS Transeq tool from EMBL. Only the three forward frames were translated. Spectra obtained from the AHARIBO enrichment of newly synthesized proteins were searched against a database of typical contaminants like keratins, trypsin, bovine serum albumin provided by MaxQuant (PMID: 19029910). The software utilized for database searching was Proteome Discoverer (v2.4, Thermo Scientific); the Non-fragment filter and the Top N Peaks Filter (with N=4 per 100 Da) were also used in the workflow to eliminate noise signals from the MS/MS spectra. The spectra not matching with high confidence this database were searched against the human SwissProt database. Those not matching with both databases were used to match the in-silico translated database generated by EMBOSS Transeq using semi-specific tryptic cleavage to consider also unexpected translation start sites. We considered only those spectra that passed the 1% FDR threshold, and we created two distinct groups for those peptides with an AUG “in-frame” vs not in-frame.

Ribosome Profiling

mESCs at 80 % confluence were pre-treated with the elongation inhibitor cycloheximide before rapid harvest on ice and cell lysis (lysis buffer, IMMAGINA BioTechnology #RL001-1). Clarified cell lysates (1.7 total a.u., measured by Nanodrop) were treated with 1.3 U of RNase I (Thermo, #AM2295) in W-buffer (IMMAGINA BioTechnology #RL001-4) containing 1x sBlock to digest RNA not protected by ribosomes. Digestion was performed for 45 min at RT and then stopped with Superase-In RNase Inhibitor (Thermo Scientific #AM2696) for 10 min on

ice. Samples were then processed differentially according to the specific approaches described below.

Standard RIBO-seq. 80S ribosomes were isolated by centrifuging lysates through a 30% sucrose cushion at 95000 rpm, for 2 h at 4°C. The cushion was resuspended in W-buffer and treated with SDS 10% (final 1%) and 5 µL of proteinase K (20mg/mL), and incubated at 37 °C in a water bath for 75 min before acid phenol:chlorophorm:isoamyl alcohol (pH 4.5) RNA extraction.

AHARIBO RIBO-seq. The lysates was incubated with Dibenzocyclooctyne-PEG_4-biotin conjugate (Sigma-Aldrich #760749; 50 µM final concentration) in a reaction volume of 100 µl for 1 h on a rotator in slow motion (9 rpm) at 4°C. Lysates were then incubated with 50 µl of eMagSi-cN beads (IMMAGINA BioTechnology #018-eMS-001) for 30 min at 4°C on the rotator in slow motion (9 rpm). Subsequently, samples were taken off the rotator and placed on a magnetic rack on ice, and supernatants were discarded. Beads were washed two times with 500 µl of 1x PBS supplemented with 0.1 % Triton-X100, 1x sBlock and 1:10,000 RiboLock RNase Inhibitor (Thermo Scientific #EO0381) on the rotator in slow motion at 4 °C, removing supernatants from the tubes sitting on the magnetic rack and gently adding new washing solution each time. After the final wash, beads were resuspended in 200 µl of W-buffer and transferred to a new vial. Then, 20 µl of 10% SDS and 5 ul of Proteinase K (Qiagen #19131) were added to each sample, and samples were incubated at 37 °C for 75 min in a water bath. Subsequently, suspensions were transferred to a new vial and acid phenol:chloroform:isoamyl alcohol (pH 4.5) RNA extraction was performed.

For both approaches, protocol steps starting from RNA extraction were performed as follows. Briefly, an equal volume of phenol:chloroform:isoamyl alcohol was added, and samples were vortexed and centrifuged at 14,000 x g for 5 min. Aqueous phases were then transferred to new vials, 500 µl of isopropanol and 2 µl of GlycoBlue (Thermo Scientific #AM9516) were added, samples were mixed and incubated at room temperature for 3 min and then stored overnight at - 80 °C. The following day, samples were centrifuged at 14,000 x g for 30 min, supernatants were removed, 500 µl of 70 % ethanol were added to each sample and samples were then centrifuged at 14,000 x g for 10 min. Finally, pellets were air-dried and resuspended in 10 µl of nuclease-free water. Extracted RNA was then resolved by electrophoresis through denaturing TBE-urea gels, and fragments between 25 nt and 35 nt were excised. Libraries were prepared using the RiboLace kit_module 2 (IMMAGINA

BioTechnolgy #RL001_mod2) and sequenced on an Illumina HiSeq 2500 sequencer with a single-end 50 base pair run.

RIBO-seq data analysis

Reads were processed by removing 5’ adapters, discarding reads shorter than 20 nucleotides and trimming the first nucleotide of the remaining ones (using Trimmomatic v0.36). Reads mapping on the collection of M. musculus rRNAs (from the SILVA rRNA database, release 119) and tRNAs (from the Genomic tRNA database: gtrnadb.ucsc.edu/) were removed. Remaining reads were mapped on the mouse transcriptome (using the Gencode M17 transcript annotations). Antisense and duplicate reads were removed. All alignments were performed with STAR (v020201) employing default settings.

The identification of the P-site position within the reads was performed using riboWaltz (v1.1.0) computing the P-site offsets from the 3’ end of the reads. The percentage of P-sites falling in the three annotated transcript regions (5' UTR, CDS and 3' UTR) was computed using the function *region_psite* included in riboWaltz (v1.1.0). Transcript counts were normalized using the trimmed mean of M-values normalization method (TMM) implemented in the edgeR Bioconductor package. Transcripts displaying 1 CPM in at least one replicate were kept for further analyses.

Additional files

Supplementary files

- Figure 1-source data 1. A table with the relative abundance of AHARIBO-rC-isolated proteins.
- Figure 1-source data 2. Gene Ontology analysis data
- Figure 2-source data 1. A table with the pSILAC data
- Figure 2-source data 2. A table with AHARIBO differentially expressed proteins. Proteins are considered differentially expressed when adjusted p-values are smaller than 0.05
- Figure 3-source data 1. A table with differentially expressed genes (DEGs) from RNA-seq data comprising: logFC, LogCPM, LogFWER, LogPval. Genes are considered differentially expressed when both log fold-changes are higher/smaller than 1.5/-1.5 and FDR adjusted p-values are smaller than 0.01.
- Figure 3-source data 2. A table with RNA and protein DEGs from AHARIBO-nP, pSILAC, AHARIBO-rC and global translatome. Genes are considered differentially expressed when both log fold-changes are higher/smaller than 1.5/-1.5 and FDR adjusted p-values are smaller than 0.01. Proteins are considered differentially expressed when adjusted p-values are smaller than 0.05.
- Figure 4-source data 1. A table with the list of lncRNA identified by RNA-seq.
- Figure 4-source data 2. A table with the list of lncRNA identified by RIBO-seq
- Figure 4-source data 3 A table with the list of Matching peptides from AHARIBO's identified lncRNAs

Data availability

NGS raw fastq files and processed data are deposited on SRA with the dataset identifier PRJNA381797. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD022679 and 10.6019/PXD022679

Funding

IMMAGINA internal R&D funding and LP6/99 financial support from Autonomous Province of Trento and Banca Intesa.

References

Amit Blumberg, Yixin Zhao, Yi-Fei Huang, Noah Dukler, Edward J. Rice, Katie Krumholz, Charles G. Danko VOPS. 2019. Characterizing RNA stability genome-wide through combined analysis of PRO-seq and RNA-seq data. *BioRXiv*. doi:10.1101/690644

Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**:166–9. doi:10.1093/bioinformatics/btu638

Anderson DM, Anderson KM, Chang C-L, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN. 2015. A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell* **160**:595–606. doi:10.1016/j.cell.2015.01.009

Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D. 2003. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **100**:3889–3894. doi:10.1073/pnas.0635171100

Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MAS, Brocard M, Couso J-P. 2014. Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *Elife* **3**. doi:10.7554/eLife.03528

Aviner R, Geiger T, Elroy-Stein O. 2013. PUNCH-P for global translatome profiling. *Translation* **1**:e27516. doi:10.4161/trla.27516

Bagert JD, Xie YJ, Sweredoski MJ, Qi Y, Hess S, Schuman EM, Tirrell DA. 2014. Quantitative, Time-Resolved Proteomic Analysis by Combining Bioorthogonal Noncanonical Amino Acid Tagging and Pulsed Stable Isotope Labeling by Amino Acids in Cell Culture. *Mol Cell Proteomics* **13**:1352–1358. doi:10.1074/mcp.M113.031914

Bazin J, Baerenfaller K, Gosai SJ, Gregory BD, Crespi M, Bailey-Serres J. 2017. Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. *Proc Natl Acad Sci* **114**:E10018–E10027. doi:10.1073/pnas.1708433114

Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, Giraldez AJ. 2014. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J* **33**:981–993. doi:10.1002/embj.201488411

Bhatta A, Atianand M, Jiang Z, Crabtree J, Blin J, Fitzgerald KA. 2020. A Mitochondrial Micropeptide Is Required for Activation of the Nlrp3 Inflammasome. *J Immunol*

204:428–437. doi:10.4049/jimmunol.1900791

Biever A, Glock C, Tushev G, Ciirdaeva E, Dalmay T, Langer JD, Schuman EM. 2020. Monosomes actively translate synaptic mRNAs in neuronal processes. *Science (80-)* **367**:eaay4991. doi:10.1126/science.aay4991

Biscarini S, Capauto D, Peruzzi G, Lu L, Colantoni A, Santini T, Shneider NA, Caffarelli E, Laneve P, Bozzoni I. 2018. Characterization of the lncRNA transcriptome in mESC-derived motor neurons: Implications for FUS-ALS. *Stem Cell Res* **27**:172–179. doi:10.1016/j.scr.2018.01.037

Blobel G, Sabatini D. 1971. Dissociation of Mammalian Polyribosomes into Subunits by Puromycin. *Proc Natl Acad Sci* **68**:390–394. doi:10.1073/pnas.68.2.390

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–20. doi:10.1093/bioinformatics/btu170

Calve S, Witten AJ, Ocken AR, Kinzer-Ursem TL. 2016. Incorporation of non-canonical amino acids into the developing murine proteome. *Sci Rep* **6**:32377. doi:10.1038/srep32377

Carelli S, Giallongo T, Rey F, Latorre E, Bordoni M, Mazzucchelli S, Gorio MC, Pansarasa O, Provenzani A, Cereda C, Di Giulio AM. 2019. HuR interacts with lincBRN1a and lincBRN1b during neuronal stem cells differentiation. *RNA Biol* **16**:1471–1485. doi:10.1080/15476286.2019.1637698

Carlevaro-Fita J, Rahim A, Guigó R, Vardy LA, Johnson R. 2016. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* **22**:867–82. doi:10.1261/rna.053561.115

Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, Forrest ARR, Carninci P, Biffo S, Stupka E, Gustincich S. 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **491**:454–457. doi:10.1038/nature11508

Chandrasekaran V, Juszkiewicz S, Choi J, Puglisi JD, Brown A, Shao S, Ramakrishnan V, Hegde RS. 2019. Mechanism of ribosome stalling during translation of a poly(A) tail. *Nat Struct Mol Biol* **26**:1132–1140. doi:10.1038/s41594-019-0331-x

Chen J, Brunner A-D, Cogan JZ, Nuñez JK, Fields AP, Adamson B, Itzhak DN, Li JY, Mann M, Leonetti MD, Weissman JS. 2020. Pervasive functional translation of noncanonical human open reading frames. *Science (80-)* **367**:1140–1146. doi:10.1126/science.aay0262

Clamer M, Tebaldi T, Lauria F, Bernabò P, Gómez-Biagi RF, Marchioretto M, Kandala DT,

Minati L, Perenthaler E, Gubert D, Pasquardini L, Guella G, Groen EJN, Gillingwater TH, Quattrone A, Viero G. 2018. Active Ribosome Profiling with RiboLace. *Cell Rep* **25**:1097-1108.e5. doi:10.1016/j.celrep.2018.09.084

D'Lima NG, Ma J, Winkler L, Chu Q, Loh KH, Corpuz EO, Budnik BA, Lykke-Andersen J, Saghatelian A, Slavoff SA. 2017. A human microprotein that interacts with the mRNA decapping complex. *Nat Chem Biol* **13**:174–180. doi:10.1038/nchembio.2249

Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM. 2006. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci* **103**:9482–9487. doi:10.1073/pnas.0601637103

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See L-H, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR. 2012. Landscape of transcription in human cells. *Nature* **489**:101–108. doi:10.1038/nature11233

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. doi:10.1093/bioinformatics/bts635

Eden E, Geva-Zatorsky N, Issaeva I, Cohen A, Dekel E, Danon T, Cohen L, Mayo A, Alon U. 2011. Proteome Half-Life Dynamics in Living Human Cells. *Science (80-)* **331**:764–768. doi:10.1126/science.1199784

Enam SU, Zinshteyn B, Goldman DH, Cassani M, Livingston NM, Seydoux G, Green R. 2020. Puromycin reactivity does not accurately localize translation at the subcellular level. *Elife* **9**. doi:10.7554/eLife.60303

Fiorillo C, Moro F, Yi J, Weil S, Brisca G, Astrea G, Severino M, Romano A, Battini R, Rossi A, Minetti C, Bruno C, Santorelli FM, Vallee R. 2014. Novel Dynein DYNC1H1

Neck and Motor Domain Mutations Link Distal Spinal Muscular Atrophy and Abnormal Cortical Development. *Hum Mutat* **35**:298–302. doi:10.1002/humu.22491

Frith MC, Forrest AR, Nourbakhsh E, Pang KC, Kai C, Kawai J, Carninci P, Hayashizaki Y, Bailey TL, Grimmond SM. 2006. The Abundance of Short Proteins in the Mammalian Proteome. *PLoS Genet* **2**:e52. doi:10.1371/journal.pgen.0020052

Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. 2007. Peptides Encoded by Short ORFs Control Development and Define a New Eukaryotic Gene Family. *PLoS Biol* **5**:e106. doi:10.1371/journal.pbio.0050106

Garreau de Loubresse N, Prokhorova I, Holtkamp W, Rodnina M V., Yusupova G, Yusupov M. 2014. Structural basis for the inhibition of the eukaryotic ribosome. *Nature* **513**:517–522. doi:10.1038/nature13737

Gonzalez-Billault C, Owen R, Gordon-Weeks PR, Avila J. 2002. Microtubule-associated protein 1B is involved in the initial stages of axonogenesis in peripheral nervous system cultured neurons. *Brain Res* **943**:56–67. doi:10.1016/S0006-8993(02)02534-9

Grollman AP. 1967. Inhibitors of protein biosynthesis. II. Mode of action of anisomycin. *J Biol Chem* **242**:3226–33. doi:10;242(13):3226-33

Guttman M, Rinn JL. 2012. Modular regulatory principles of large non-coding RNAs. *Nature* **482**:339–346. doi:10.1038/nature10887

Hodas JJL, Nehring A, Höche N, Sweredoski MJ, Pielot R, Hess S, Tirrell DA, Dieterich DC, Schuman EM. 2012. Dopaminergic modulation of the hippocampal neuropil proteome identified by bioorthogonal noncanonical amino acid tagging (BONCAT). *Proteomics* **12**:2464–76. doi:10.1002/pmic.201200112

Huang J-Z, Chen M, Chen D, Gao X-C, Zhu S, Huang H, Hu M, Zhu H, Yan G-R. 2017. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol Cell* **68**:171-184.e6. doi:10.1016/j.molcel.2017.09.015

Ingolia Nicholas T., Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7**:1534–1550. doi:10.1038/nprot.2012.086

Ingolia NT, Lareau LF, Weissman JS. 2011. 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

Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu Y-M, Robinson DR, Beer DG, Feng FY, Iyer HK, Chinnaiyan AM. 2015. The landscape of long noncoding RNAs

in the human transcriptome. *Nat Genet* **47**:199–208. doi:10.1038/ng.3192

Jao CY, Salic A. 2008. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc Natl Acad Sci U S A* **105**:15779–84. doi:10.1073/pnas.0808480105

Jewett JC, Bertozzi CR. 2010. Cu-free click cycloaddition reactions in chemical biology. *Chem Soc Rev* **39**:1272–9. doi:10.1039/b901970g

Jiao Y, Meyerowitz EM. 2010. Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol Syst Biol* **6**:419. doi:10.1038/msb.2010.76

Kapur M, Monaghan CE, Ackerman SL. 2017. Regulation of mRNA Translation in Neurons—A Matter of Life and Death. *Neuron* **96**:616–637. doi:10.1016/j.neuron.2017.09.057

Kleaveland B, Shi CY, Stefano J, Bartel DP. 2018. A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. *Cell* **174**:350-362.e17. doi:10.1016/j.cell.2018.05.022

Latremoliere A, Cheng L, DeLisle M, Wu C, Chew S, Hutchinson EB, Sheridan A, Alexandre C, Latremoliere F, Sheu S-H, Golidy S, Omura T, Huebner EA, Fan Y, Whitman MC, Nguyen E, Hermawan C, Pierpaoli C, Tischfield MA, Woolf CJ, Engle EC. 2018. Neuronal-Specific TUBB3 Is Not Required for Normal Neuronal Function but Is Essential for Timely Axon Regeneration. *Cell Rep* **24**:1865-1879.e9. doi:10.1016/j.celrep.2018.07.029

Lee S, Liu B, Lee S, Huang S-X, Shen B, Qian S-B. 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci* **109**:E2424–E2432. doi:10.1073/pnas.1207846109

Lewandowski JP, Dumbović G, Watson AR, Hwang T, Jacobs-Palmer E, Chang N, Much C, Turner KM, Kirby C, Rubinstein ND, Groff AF, Liapis SC, Gerhardinger C, Bester A, Pandolfi PP, Clohessy JG, Hoekstra HE, Sauvageau M, Rinn JL. 2020. The Tug1 lncRNA locus is essential for male fertility. *Genome Biol* **21**:237. doi:10.1186/s13059-020-02081-5

Livak KJ, Schmittgen TD. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**:402–408. doi:10.1006/meth.2001.1262

Magny EG, Pueyo JI, Pearl FMG, Cespedes MA, Niven JE, Bishop SA, Couso JP. 2013. Conserved Regulation of Cardiac Calcium Uptake by Peptides Encoded in Small Open Reading Frames. *Science (80-)* **341**:1116–1120. doi:10.1126/science.1238802

Martinez A, Traverso JA, Valot B, Ferro M, Espagne C, Ephritikhine G, Zivy M, Giglione C,

1194 Meinnel T. 2008. Extent of N-terminal modifications in cytosolic proteins from
1195 eukaryotes. *Proteomics* **8**:2809–2831. doi:10.1002/pmic.200701191
1196 Mathias AP, Williamson R, Huxley HE, Page S. 1964. Occurrence and function of polysomes
1197 in rabbit reticulocytes. *J Mol Biol* **9**:154-IN11. doi:10.1016/S0022-2836(64)80097-8
1198 Minajigi A, Froberg JE, Wei C, Sunwoo H, Kesner B, Colognori D, Lessing D, Payer B,
1199 Boukhali M, Haas W, Lee JT. 2015. A comprehensive Xist interactome reveals cohesin
1200 repulsion and an RNA-directed chromosome conformation. *Science (80-)* **349**:aab2276–
1201 aab2276. doi:10.1126/science.aab2276
1202 Morisaki T, Lyon K, DeLuca KF, DeLuca JG, English BP, Zhang Z, Lavis LD, Grimm JB,
1203 Viswanathan S, Looger LL, Lionnet T, Stasevich TJ. 2016. Real-time quantification of
1204 single RNA translation dynamics in living cells. *Science (80-)* **0899**:aaf0899.
1205 doi:10.1126/science.aaf0899
1206 Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, Reese AL,
1207 McAnally JR, Chen X, Kavalali ET, Cannon SC, Houser SR, Bassel-Duby R, Olson EN.
1208 2016. A peptide encoded by a transcript annotated as long noncoding RNA enhances
1209 SERCA activity in muscle. *Science (80-)* **351**:271–275. doi:10.1126/science.aad4076
1210 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for
1211 differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–
1212 40. doi:10.1093/bioinformatics/btp616
1213 Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression
1214 analysis of RNA-seq data. *Genome Biol* **11**:R25. doi:10.1186/gb-2010-11-3-r25
1215 Rothenberg DA, Taliaferro JM, Huber SM, Begley TJ, Dedon PC, White FM. 2018. A
1216 Proteomics Approach to Profiling the Temporal Translational Response to Stress and
1217 Growth. *iScience* **9**:367–381. doi:10.1016/j.isci.2018.11.004
1218 Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. 2014. Long non-coding RNAs as a
1219 source of new peptides. *Elife* **3**:e03523. doi:10.7554/eLife.03523
1220 Saleh AM, Wilding KM, Calve S, Bundy BC, Kinzer-Ursem TL. 2019. Non-canonical amino
1221 acid labeling in proteomics and biotechnology. *J Biol Eng* **13**:43. doi:10.1186/s13036-
1222 019-0166-3
1223 Schwanhäusser B, Gossen M, Dittmar G, Selbach M. 2009. Global analysis of cellular protein
1224 translation by pulsed SILAC. *Proteomics* **9**:205–209. doi:10.1002/pmic.200800275
1225 Seedhom MO, Hickman HD, Wei J, David A, Yewdell JW. 2016. Protein Translation
1226 Activity: A New Measure of Host Immune Cell Activation. *J Immunol* **197**:1498–1506.
1227 doi:10.4049/jimmunol.1600088

Slavoff SA, Heo J, Budnik BA, Hanakahi LA, Saghatelian A. 2014. A Human Short Open Reading Frame (sORF)-encoded Polypeptide That Stimulates DNA End Joining. *J Biol Chem* **289**:10950–10957. doi:10.1074/jbc.C113.533968

Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A. 2013. Peptidomic discovery of short open reading frame–encoded peptides in human cells. *Nat Chem Biol* **9**:59–64. doi:10.1038/nchembio.1120

Sun Q, Hao Q, Prasanth K V. 2018. Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. *Trends Genet* **34**:142–157. doi:10.1016/j.tig.2017.11.005

Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, Emili A, Xie XS. 2010. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science (80-)* **329**:533–538. doi:10.1126/science.1188308

Tebaldi T, Zuccotti P, Peroni D, Köhn M, Gasperini L, Potrich V, Bonazza V, Dudnakova T, Rossi A, Sanguinetti G, Conti L, Macchi P, D'Agostino V, Viero G, Tollervey D, Hüttelmaier S, Quattrone A. 2018. HuD Is a Neural Translation Enhancer Acting on mTORC1-Responsive Genes and Counteracted by the Y3 Small Non-coding RNA. *Mol Cell* **71**:256-270.e10. doi:10.1016/j.molcel.2018.06.032

Tom Dieck S, Müller A, Nehring A, Hinz FI, Bartnik I, Schuman EM, Dieterich DC. 2012. Metabolic labeling with noncanonical amino acids and visualization by chemoselective fluorescent tagging. *Curr Protoc cell Biol* **Chapter 7**:Unit7.11. doi:10.1002/0471143030.cb0711s56

van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, Kirchner M, Maatz H, Blachut S, Sandmann C-L, Kanda M, Worth CL, Schafer S, Calviello L, Merriott R, Patone G, Hummel O, Wyler E, Obermayer B, Mücke MB, Lindberg EL, Trnka F, Memczak S, Schilling M, Felkin LE, Barton PJR, Quaife NM, Vanezis K, Diecke S, Mukai M, Mah N, Oh S-J, Kurtz A, Schramm C, Schwinge D, Sebode M, Harakalova M, Asselbergs FW, Vink A, de Weger RA, Viswanathan S, Widjaja AA, Gärtner-Rommel A, Milting H, dos Remedios C, Knosalla C, Mertins P, Landthaler M, Vingron M, Linke WA, Seidman JG, Seidman CE, Rajewsky N, Ohler U, Cook SA, Hubner N. 2019. The Translational Landscape of the Human Heart. *Cell* **178**:242-260.e29. doi:10.1016/j.cell.2019.05.010

van Solingen C, Sharma M, Bijkerk R, Afonso MS, Koelwyn GJ, Scacalossi KR, van Zonneveld AJ, Moore KJ. 2018. Abstract 027: A Micropeptide Concealed in a Putative Long Non-coding RNA Directs Inflammation. *Arterioscler Thromb Vasc Biol* **38**. doi:10.1161/atvb.38.suppl_1.027

Wang C, Han B, Zhou R, Zhuang X. 2016. Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. *Cell* **165**:990–1001. doi:10.1016/j.cell.2016.04.040

Wang T, Cui Y, Jin J, Guo J, Wang G, Yin X, He Q-Y, Zhang G. 2013. Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. *Nucleic Acids Res* **41**:4743–4754. doi:10.1093/nar/gkt178

Wild K, Aleksić M, Lapouge K, Juaire KD, Flemming D, Pfeffer S, Sinning I. 2020. MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. *Nat Commun* **11**:776. doi:10.1038/s41467-020-14603-7

Wu B, Eliscovich C, Yoon YJ, Singer RH. 2016. Translation dynamics of single mRNAs in live cells and neurons. *Science (80-)* **352**:1430–1435. doi:10.1126/science.aaf1084

Yan X, Hoek TA, Vale RD, Tanenbaum ME. 2016. Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* **165**:976–989. doi:10.1016/j.cell.2016.04.034

Ying Q-L, Stavridis M, Griffiths D, Li M, Smith A. 2003. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**:183–186. doi:10.1038/nbt780

Yoon J-H, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M, Becker KG, Gorospe M. 2012. LincRNA-p21 Suppresses Target mRNA Translation. *Mol Cell* **47**:648–655. doi:10.1016/j.molcel.2012.06.027

Zeng C, Fukunaga T, Hamada M. 2018. Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. *BMC Genomics* **19**:414. doi:10.1186/s12864-018-4765-z