

One-shot analysis of translated mammalian lncRNAs with AHARIBO

Luca Minati^{*1}, Claudia Firrito^{*1}, Alessia Del Piano^{*1}, Alberto Peretti^{*1}, Simone Sidoli², Daniele Peroni³, Romina Belli³, Francesco Gandolfi⁴, Alessandro Romanel⁴, Paola Bernabò¹, Jacopo Zasso⁵, Alessandro Quattrone⁵, Graziano Guella⁶, Fabio Lauria⁷, Gabriella Viero⁷ 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. * These authors contributed equally to this work. † Lead contact / Corresponding author

SUMMARY

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 uncharacterized proteins, 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 lncRNAs 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 polypeptides. 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, translatoe, 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 the lncRNAs are localized in the cytoplasm (Carlevaro-Fita et al., 2016), where they are supposed to be untranslated. Ribosome profiling (RIBO-seq), which provides positional information of ribosomes along transcripts (Clamer et al., 2018; Ingolia et al., 2012a), 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 the modulation of translation (Carrieri et al., 2012; Yoon et al., 2012), while others have been found to be potentially translated (Anderson et al., 2015; Aspden et al., 2014; Bazin et al., 2017; van Heesch et al., 2019; Ingolia et al., 2011; Nelson et al., 2016; Ruiz-Orera et al., 2014). As coding RNAs, lncRNAs can be associated with fully translating or with 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 translatoe and proteome. Unfortunately, classical RIBO-seq approaches cannot distinguish between lncRNAs producing peptides from those that are sequestering ribosomes (lncRNA bound to ribosomes without translation), thus acting as ribosome sponges. Complementary proteomics approaches, such as mass spectrometry, can help defining and quantitatively monitoring the production of peptides, but are less sensitive techniques than RNA sequencing (van Heesch et al., 2019; Slavoff et al., 2013). 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 RIBOsosome isolation), a combination of protocols to 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 means of the incorporation of 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) analyze co-purified 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 (Fig 1A). Briefly, the protocol consists in 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 by means of a small molecule, iv) cell lysis and AHA “copper-free click reaction” (Jewett and Bertozzi, 2010) for v) final ribosome capture with magnetic beads. We reasoned that the protocol for isolating ribosomes through AHA can be used to obtain information from nascent peptides, constitutive components of the ribosomes, mRNAs or lncRNAs associated with ribosomes. For this reason, we optimized several parameters along the entire protocol, from washing steps to nuclease treatments (Fig. 1A), to isolate (1) the full translational complex (ribosomes, ribosome-associated proteins, nascent peptides and RNAs; called AHARIBO-rC, Ribosomal Complex), (2) the *de novo* synthesized proteome (called AHARIBO-nP, Nascent Proteins) and (3) ribosome-protected fragments (RPFs) (called AHARIBO RIBO-seq, RIBOsome profiling by sequencing).

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

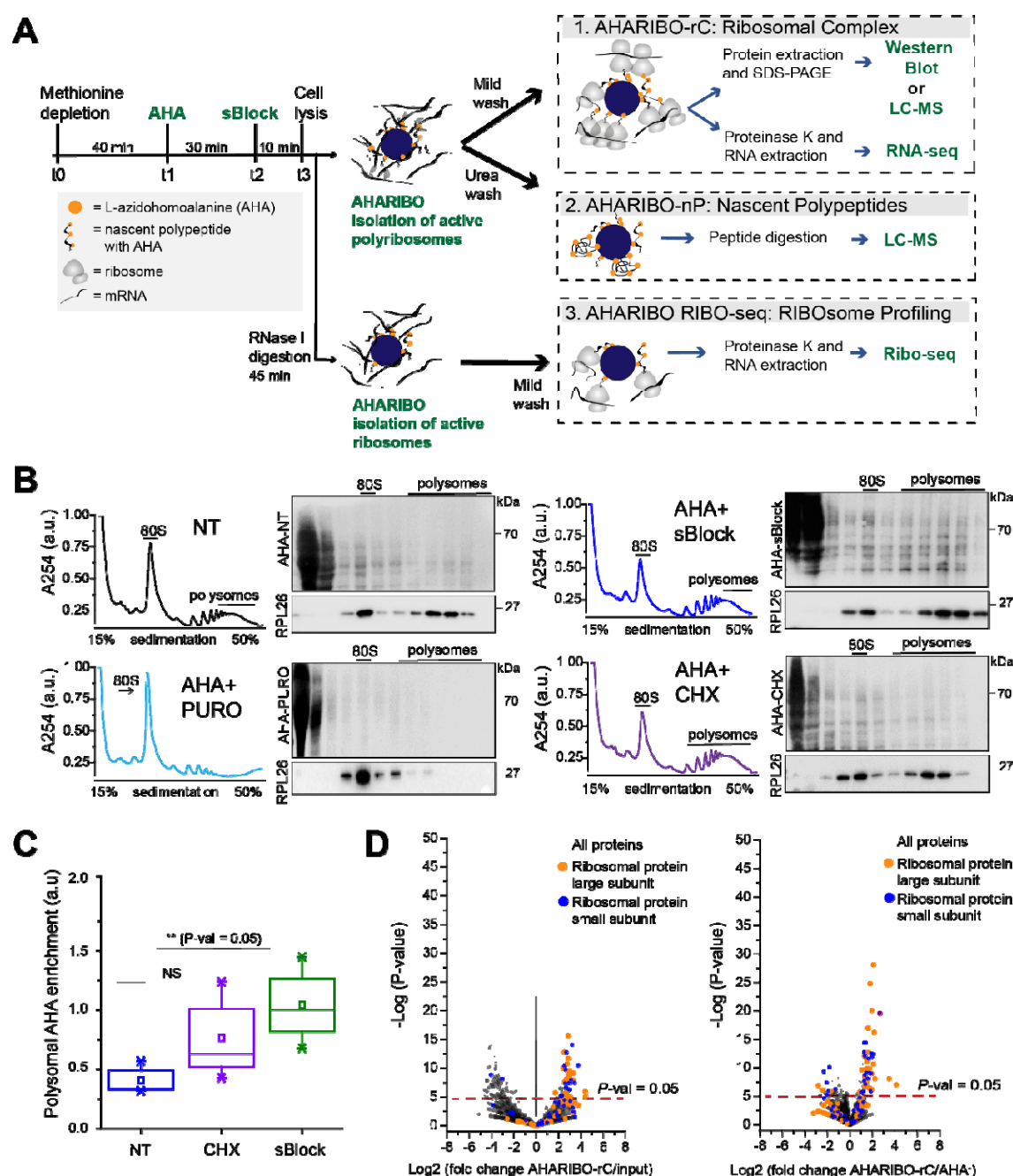


Figure 1. AHA labeling of nascent peptide chains and ribosome separation. A) Schematic representation of the AHARIBO experimental workflow. After methionine depletion, AHA incubation and sBlock treatment, cell lysates can be processed for (1) AHARIBO-rC: isolation of translational complex (ribosomes, ribosome-associated proteins, nascent peptides and RNAs); (2) AHARIBO-nP: isolation of *de novo* synthesized proteins; (3) AHARIBO RIBO-seq. B) Sucrose gradient profiles resulting from absorbance measurement at 254 nm. On the right of each profile, SDS-PAGE of protein extracts from each fraction of the profile. After ultracentrifugation on a 15-50% sucrose gradient, HeLa cell lysates were fractionated (1 mL fraction) and the total AHA protein content of each individual fraction was visualized after SDS-PAGE. Staining of the membrane was performed by biotin cycloaddition followed by streptavidin-HRP. RPL26 protein was used as ribosome marker in the different fractions. C) Box plot showing the AHA signal enrichment in the polysomal region of the profiles obtained from cells treated with either CHX or sBlock or no drugs (NT). Results are shown as a median (\pm SE) of 3 independent replicates. D) Volcano plots showing the $-\log(P\text{-value})$ versus the relative abundance of AHARIBO-rC-isolated

proteins. Data are compared with input (starting AHA-containing lysate, left) or with the non-specific signal from streptavidin-coated beads incubated with cell lysate not supplemented with AHA (AHA⁻).

We observed that 30 min is the optimal incubation time for sufficient AHA incorporation and maximum RNA recovery (Supplementary Fig. 1A-C). Then, we compared the drug efficiency of stabilization of the nascent peptide by co-sedimentation profiling of AHA-tagged polypeptides with ribosomes (Fig. 1B). As a control, cells were treated in parallel with puromycin to cause ribosome disassembly and release of the growing peptide chains (Fig. 1B) (Blobel and Sabatini, 1971). According to literature, we found that both CHX and sBlock are able to stabilize AHA-peptides on ribosomes and polysomes (i.e. actively translating complexes) (Biever et al., 2020; Mathias et al., 1964). The efficiency to anchor polypeptides on ribosomes in CHX and sBlock treated cells was 50% higher compared to untreated cells, confirming that the drugs effectively stabilize nascent polypeptides (Fig. 1C). The high signal observed in light fractions is caused by AHA-labeled proteins released from ribosomes. As expected, in puromycin-treated samples the AHA signal was mainly detected in the first two fractions, proving that the observed signal in the heavier fractions of CHX- and sBlock-treated cells is not caused by a diffusion of AHA-labeled peptides. Since sBlock outperformed CHX in anchoring efficiency (Fig. 1C), we used sBlock in all further experiments.

Prompted by the evidence that it is possible to stably anchor nascent peptides on ribosomes by means of a small molecule, we validated the ability of AHARIBO-rC to isolate RNAs and proteins associated with the translation complex. To this aim, we performed a label-free liquid chromatography-mass spectrometry (LC-MS) analysis on captured proteins. Enrichment analysis of AHARIBO-rC proteins relative to the input and to the non-specific signal (AHA⁻) showed that ribosomal proteins belonging to both the large and the small ribosome subunits are indeed more abundant in AHARIBO-rC samples than the controls (Fig. 1D). Gene ontology (GO) analysis of these proteins 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 (AHA⁻), confirming efficient pulldown of proteins involved in translation (Supplementary Tab. 1). LC-MS results were also confirmed by Western blot analysis on proteins involved in translation (RPS6, RPL26 and eEF2) (Supplementary Fig. 1D).

Then, we investigated the possibility to use AHARIBO-rC to determine the translational status of cultured cells. To address this question, we down-regulated protein synthesis by treating HeLa cells with arsenite (Ar), which induces translational inhibition and stress granules formation (Wang et al., 2016). Consistent with our previous evidence, we observed a reduction of RNA captured in Ar-treated cells relative to the control (Supplementary Fig. 1E-F). In line with this finding, qRT-PCR analysis showed about 50% reduction in 18S rRNA levels when translation is inhibited (i.e. in Ar AHARIBO-rC pulldown samples compare to not arsenite treated cells) (Supplementary Fig. 1G).

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

Motivated by the evidence that AHARIBO-rC can be used to isolate *bona-fide* translationally active ribosomes, we further tested the method genome-wide on mouse embryonic stem cells (mESCs) under a basal condition and mESCs differentiated into early neurons (EN) (Tebaldi et al., 2018) (Supplementary Fig. 2A). We analyzed 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 polypeptides (Supplementary Fig. 2B). In parallel, we isolated and analyzed the global transcriptome as the RNA extracted after 30% sucrose cushioning of cytoplasmatic lysates (Wang et al., 2013) and the global proteome, measured by pulsed SILAC (pSILAC) (Schwanhäusser et al., 2009) (Fig. 2A and 2B).

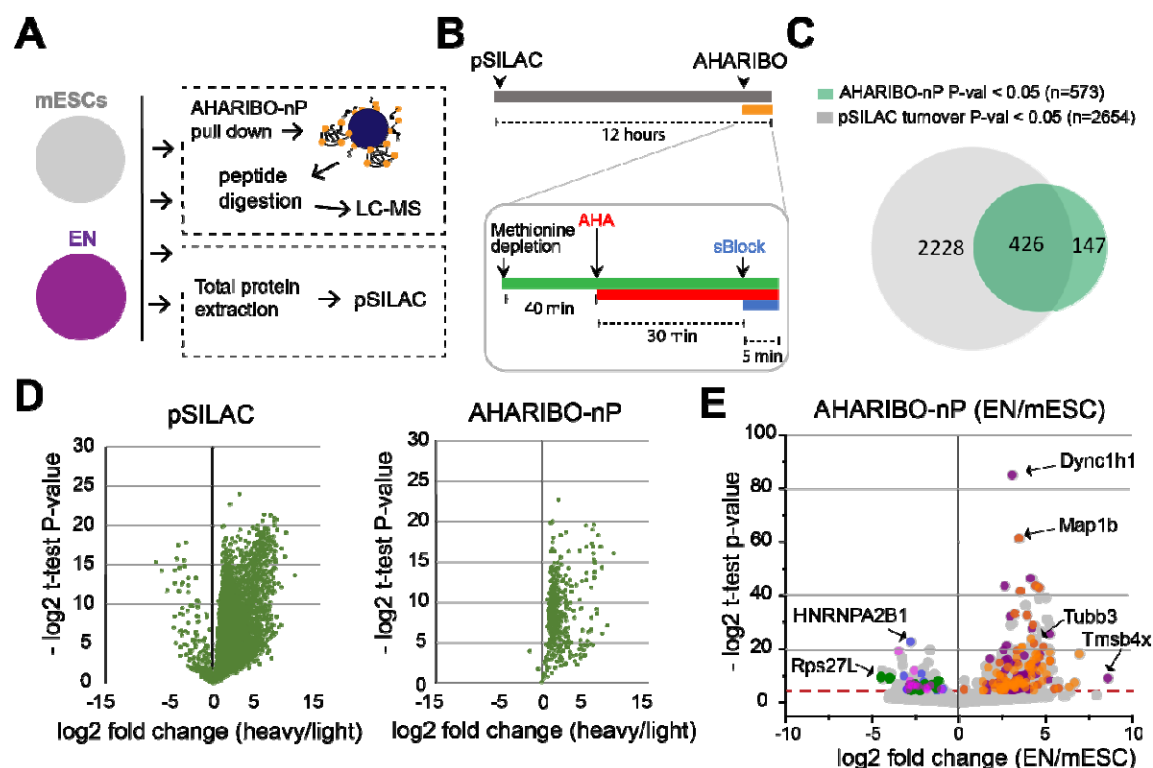


Figure 2. AHARIBO-nP and pSILAC approaches. A) Schematic representation of the experimental workflow. B) Schematic representation of combined cell treatments for pSILAC and AHARIBO-nP. C) Venn diagram representing the number of proteins identified by AHARIBO-nP and pSILAC. D) Volcano plots displaying for each protein the $-\log_2$ p-value against the fold changes of protein turnover (heavy/light) in pSILAC proteome (left) and AHARIBO-nP (right) for mESCs. E) Volcano plot displaying for each differentially expressed protein (EN/mESC) the AHARIBO-nP proteome versus $-\log_2$ (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) respectively. Grey dots, all proteins. Gene Ontology (GO) enrichment analyses were performed using PANTHER and Enrichr.

We identified 2417 newly synthesized proteins (heavy/light, P-val < 0.05) within the 24 hours of pSILAC treatment. As expected, differentiated cells (EN) showed a reduced turnover (Supplementary Fig. 2C). Then, we performed an enrichment analysis of ENs versus mESCs (EN/mESC) and identified 2667 and 573 differentially expressed proteins in pSILAC and AHARIBO-nP samples, respectively (Fig. 2C). The lower 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). Of note, 74 % of the AHARIBO-nP identified proteins is in common with the

pSILAC dataset (Fig. 2C). Accordingly, we observed high abundance of heavy amino acids in AHARIBO-nP (Fig. 2D) and a significantly higher heavy/light ratio in AHARIBO-nP compared to pSILAC (Supplementary Fig. 2D), suggesting that AHARIBO-nP is indeed able to capture the *de novo* synthesized proteome. We found positive fold changes for proteins captured by AHARIBO-nP during differentiation (Fig. 2E). Interestingly, among these proteins, several are known to be expressed during the early stages of development of the nervous system (e.g. Map1b, Tubb3 and, Dync1h1). The GO analysis of differentially expressed proteins showed that proteins involved in cytoskeleton organization and neurogenesis were upregulated (Fig. 2E), further confirming the reliability of AHARIBO-nP in monitoring *de novo* protein expression.

Collectively, these results show that *de novo* synthesized proteins isolated using AHARIBO-nP portray the phenotypic changes induced by cellular differentiation, and that 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 tested if mRNAs purified using AHARIBO-rC can be a good proxy of protein levels. To this aim, we compared AHARIBO-rC RNA and the global transcriptome with AHARIBO-nP in mESCs during differentiation.

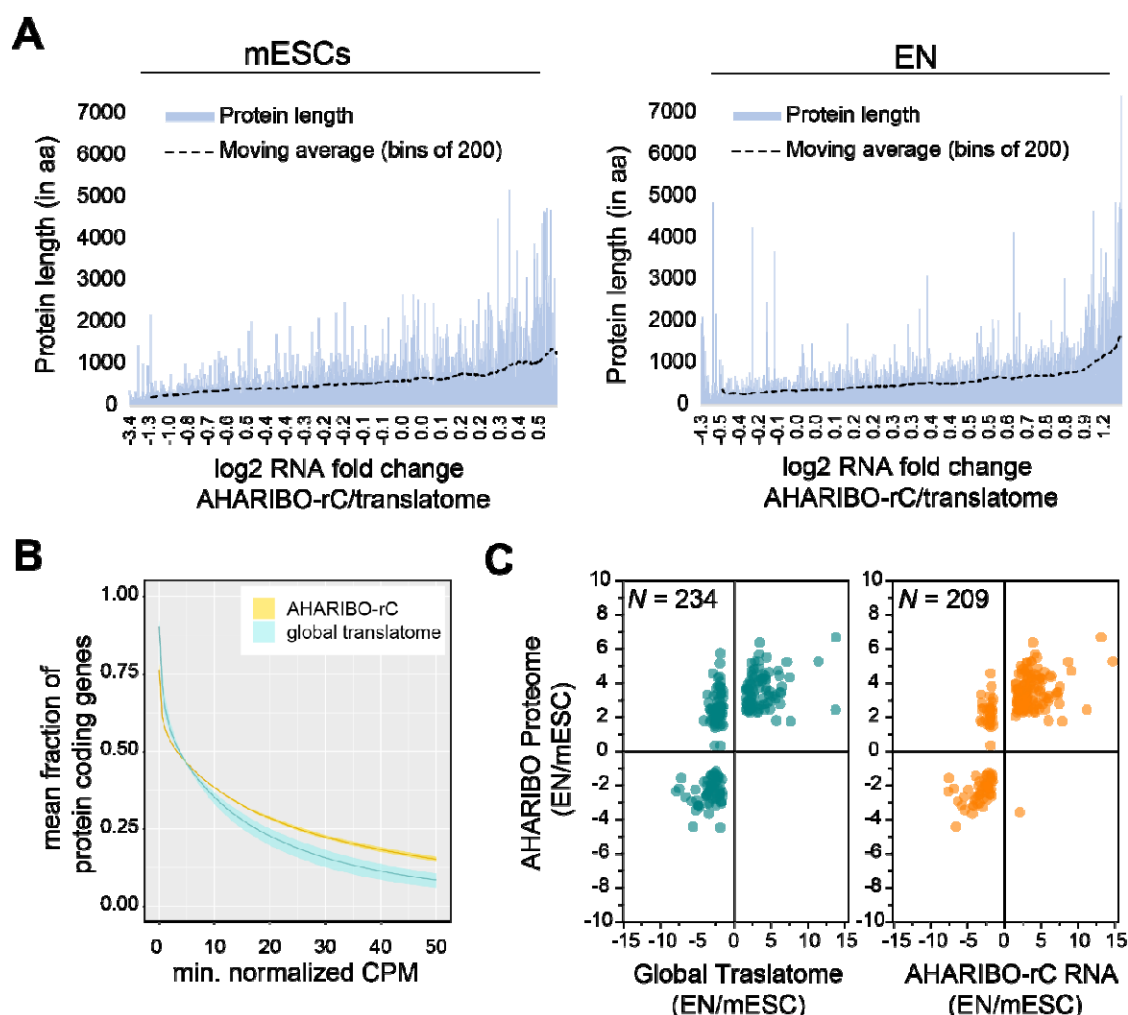


Figure 3. AHARIBO-rC RNA vs *de novo* proteome analysis. A) Protein length against increasing fold changes of AHARIBO-rC RNA /global translateome for mESCs (left) and ENs (right). B) Fraction of remaining protein-coding genes at increasing CPM thresholds in ENs. The mean of $n = 3$ (translateome) and $n = 3$ (AHARIBO-rCAHARIBO-rC) biological replicates is reported. Data are represented as mean \pm SD. (shadow). C) Scatter plot of RNA fold change (global translateome on the left, AHARIBO-rC on the right) compared to protein fold change (AHARIBO-rNP). N, number of DEGs with P-val < 0.05 .

To exclude any bias related to protein length, we checked whether AHARIBO-rNP preferentially captures either long or short proteins. We analyzed peptide size against the AHARIBO-rC RNA enrichment versus the global translateome (Fig. 3A), a value reflecting the extent to which AHARIBO-rC RNA differs from the standard analysis. Positive values were associated with long proteins (> 200 aa) in both mESCs and ENs. A 3.5 % of the proteins showed length of less than 100 aa. Proteins shorter than 100 aa account for approximately 10% of the total mouse proteome, with an average protein size in eukaryotes of about 300 aa (Frith et al., 2006). Since in all eukaryotes proteins are initiated with a methionine

residue, virtually any protein can be captured as soon as the nascent peptide exits the ribosome (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 good probability for at least one AHA residue to be available for each peptide when the blocking drug is added to the cell medium, enabling the capture of the polypeptide outside the ribosome exit tunnel. Our results confirm the effectiveness of the method and the efficiency in capturing polypeptides without major biases.

To further exclude any bias related to RNA abundance, we measured how efficiently AHARIBO-rC captures coding transcripts compared to the global transcriptome. We analysed the two datasets using increasing abundance thresholds in EN, and observed that AHARIBO-rC is comparable to the global transcriptome for low abundant transcripts and it can better capture more abundant transcripts (Fig 3B). In undifferentiated mESCs AHARIBO-rC performs as the global transcriptome (Supplementary Fig. 3A), confirming the efficiency of successful isolation of protein-coding RNAs genome-wide.

Finally, we tested whether the RNA isolated with AHARIBO-rC can predict the *de novo* synthesized proteome. After comparing differentially expressed genes (DEGs, with $P\text{-Val} < 0.05$) obtained by AHARIBO-rC RNA during mESC differentiation to the AHARIBO-nP proteome, we observed that AHARIBO-rC RNA is a good proxy of the newly synthesized proteome (Pearson's correlation $r = 0.75$) (Fig. 3C and Supplementary Fig. 3B). 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 transcriptome (Fig 3C and Supplementary Fig. 3C), thus recapitulating protein changes faithfully.

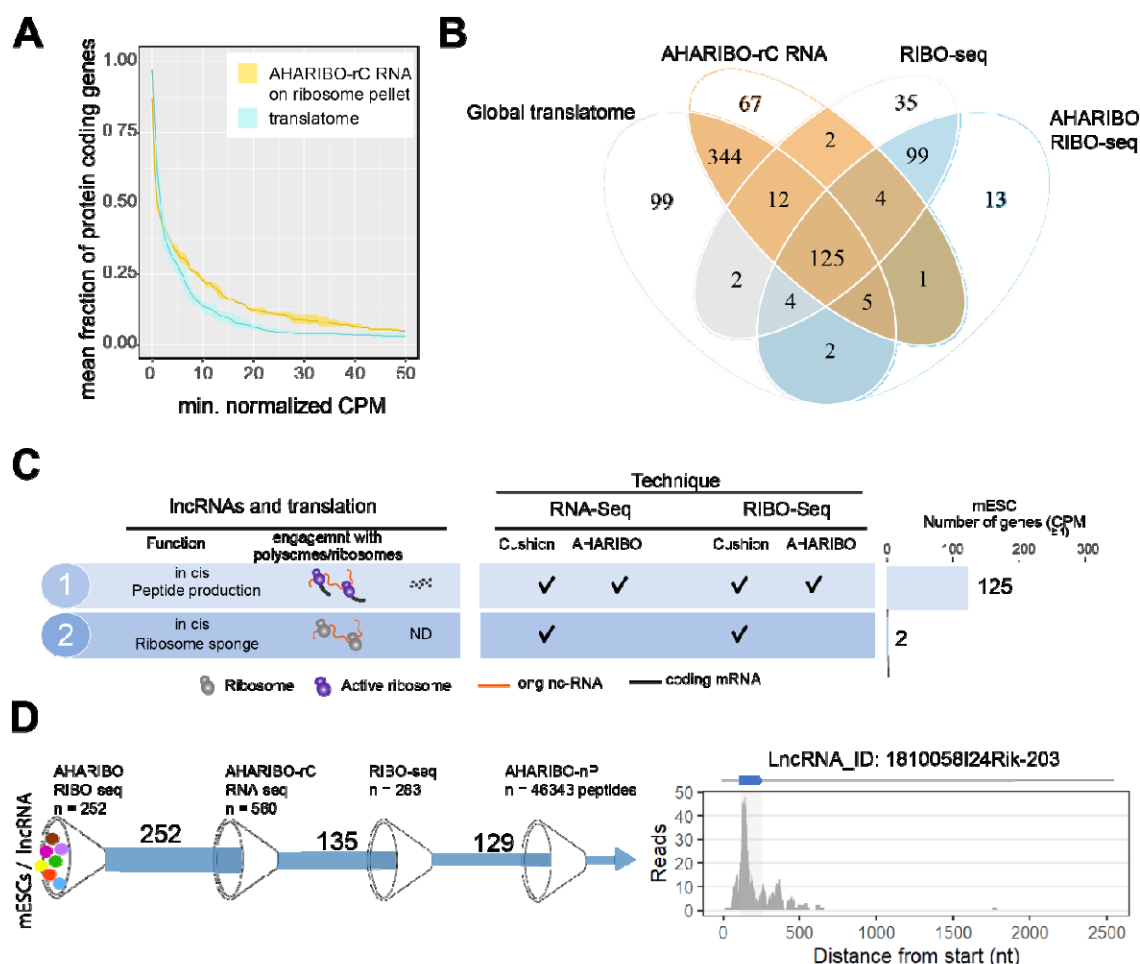


Figure 4. The AHARIBO platform can be used to detect ribosome-interacting lncRNAs. (A) Fraction of remaining non-coding genes at increasing CPM thresholds in EN. AHARIBO-rC was performed on the ribosome pellet after cushioning. The mean of $n = 3$ independent replicates is reported. Data are represented as mean \pm SD. (shadow) (B) Venn diagram with the number of lncRNAs (with at least 1 CPM) identified in each category (RNA-seq, AHARIBO-rC, RIBO-seq, AHARIBO RIBO-seq). (C) Left, Classification of lncRNAs interacting with ribosomes and relative detection through the different AHARIBO and standard approaches. (D) Left, schematic representation of the number of mESC lncRNAs in common between AHARIBO RIBO-seq, AHARIBO-rC RNA, standard RIBO-seq, and finally validated by LC-MS. Right, example a AHARIBO RIBO-seq trace of the lncRNA 1810058I24Rik displaying the hallmarks of translation including reads distribution along the entire transcript and the accumulation of reads in the known short open reading frame (shadow area and blue arrow on top).

Combined AHARIBO approaches define the functional role of lncRNAs in translation

lncRNAs are known to be in some cases associated with ribosomes. Ribosome-associate lncRNAs play a role in the regulation of translation (Ingolia et al., 2011; Pircher et al., 2014; Zeng et al., 2018) or in the production of short peptides (Chen et al., 2020; van Heesch et al., 2019). Based on our 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 lncRNAs.

From the analysis of lncRNAs isolated by AHARIBO-rC, we identified about 400 differentially expressed (DE) lncRNAs during neuronal differentiation (Supplementary Fig. 4A). Among the top-5 DE lncRNAs, 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 with ribosome footprints (RPFs), identified by ribosome profiling in mESCs (Ingolia et al., 2011) (Supplementary Fig. 4B). The analysis of the 100 lncRNAs in common between the two datasets showed a stronger enrichment of AHARIBO-rC lncRNA ribosome footprints in the AHARIBO-rC than in the global translome (Fig. 4A and Supplementary Figure 4C), meaning that only a subset of lncRNAs are potentially translated. This result suggest that a fraction of non-coding transcripts efficiently isolated with AHARIBO-rC are potentially translated.

To understand if and how lncRNAs interact with ribosomes, we performed RIBO-seq after AHARIBO pulldown (named AHARIBO RIBO-seq), and standard RNA-seq and RIBO-seq (on inputs) in mESCs. With AHARIBO RIBO-seq samples we identified an absolute number of 252 lncRNAs within RPFs. For protein-coding genes, both standard and AHARIBO RIBO-seq enriched for RPFs. The two datasets correlates well ($R^2 = 0.96$) and the P-site density showed the expected 3nt periodicity in the coding sequence, thus confirming the capability of AHARIBO protocol in capturing ribosomes (Supplementary Figure 4D). By intersecting AHARIBO RIBO-seq data with those obtained from standard methods (RIBO-seq, RNA-seq after sucrose cushioning) and AHARIBO-rC, we identified 125 common putative translated lncRNAs (Fig. 4B). The vast majority of these lncRNAs have no already known function, demonstrating the our approach could be useful to unravel new and translation-related lncRNAs. Two lncRNAs (9330151L19Rik and Gm9776) were identified only by standard RIBO-seq and RNA-seq but not with AHARIBO (Fig. 4C), suggesting that these are not translated, but rather loaded with idle ribosomes.

To validate the coding potential of lncRNAs which are in common between AHARIBO and standard RIBO-seq (n = 228) (Fig. 4D), we first *in silico* translated the transcripts in all frames to find potential ORFs with a canonical start codon (AUG). Translated sequences

were then semi-trypsin digested *in silico*, screened to remove contaminants and then manually annotated to find confident matching spectra from the AHRIBO-nP protein dataset. MS-based proteomics detected five non-canonical peptides with corresponding ribosome footprints (Gm42743, Gm26518, B230354K17Rik, D030068K23Rik, 1810058I24Rik), out of the about 46,000 collected spectra. The low number of validated transcript could be due to challenges in detecting the trypsin-digested products from short, non-canonical coding sequences. From the short list of 129 in common with all AHARIBO protocols and standard RIBO-seq (Fig. 4D), we identified at a high degree of confidence a known micropeptide of 47 aa (Fig. 4D) derived from a lncRNA expressed in murine macrophages. This micropeptide was recently detected and characterized as a relevant protein in the modulation of the innate immunity in mice (Bhatta et al., 2020).

Altogether, our data confirm that the three diverse and complementary AHARIBO approaches represent a unique method to identify ribosome-associated and translated lncRNAs.

Discussion

LncRNAs can be localized in the nucleus or in the cytoplasm. When localizing in the nucleus they can modulate transcription, pre-mRNA splicing or act as scaffold for protein interaction during chromatin organization (Sun et al., 2018). When localizing in the cytoplasm, the majority is associated to polysomes (Carlevaro-Fita et al., 2016) where lncRNA can be used to produce or not to 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), bones development (Galindo et al., 2007), macrophage activity (van Solingen et al., 2018), metabolism (Magny et al., 2013; Nelson et al., 2016) and DNA repair (Slavoff et al., 2014). Different methodological approaches have been developed to quantify the variations of mRNA abundance (e.g., RNA-seq, RNA imaging) (Amit Blumberg et al., 2019; Jao and Salic, 2008; Morisaki et al., 2016; Wu et al., 2016), RNA engagement with the translational machinery (e.g., ribosomal profiling, polysomal profiling) (Arava et al., 2003; Clamer et al., 2018; Eden et al., 2011; Taniguchi et al., 2010), and protein synthesis (e.g., pSILAC, PUNCH-P, BONCAT) (Aviner et al., 2013; Dieterich et al., 2006; Schwanhäusser et al., 2009; Yan et al., 2016). Nevertheless, available technologies hardly capture in a single experimental setup the dynamics of translation across different biological conditions, the translation of unannotated coding transcripts and translation-related functions of lncRNAs. Now that is widely accepted that a portion of the genome annotated as non-coding (> 97 %), can result in a complex transcriptome partially engaged with ribosomes (Chen et al., 2020; Djebali et al., 2012; Iyer et al., 2015), exosome sequencing and ribosome profiling should include micropeptide detection.

With AHARIBO, we introduce a strategy for the selective isolation of active ribosomes via nascent peptide chain for more comprehensive interrogation of lncRNA biology and proteogenomic studies. 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 in a fast and accessible way. The intersection of standard RIBO-seq, RNA-seq and AHARIBO approaches allowed us to identify translated lncRNA as well as few example of lncRNA sequestering ribosomes without evidence of translation. Although RNA sequencing analysis is not supported by the same diversity of LC-MS data, we successfully

identify a mouse-specific micropeptide reported to originate from a lncRNA ORF as a case study, confirming the effectiveness of AHARIBO.

Overall, we provide evidence that AHARIBO is a comprehensive and reliable toolkit suitable for downstream orthogonal approaches (RNA-seq, RIBO-seq and LC-MS), empowering scientists to shed light onto the functional complexity of the mammalian genome and lncRNA translation.

SUPPLEMENTARY INFORMATION

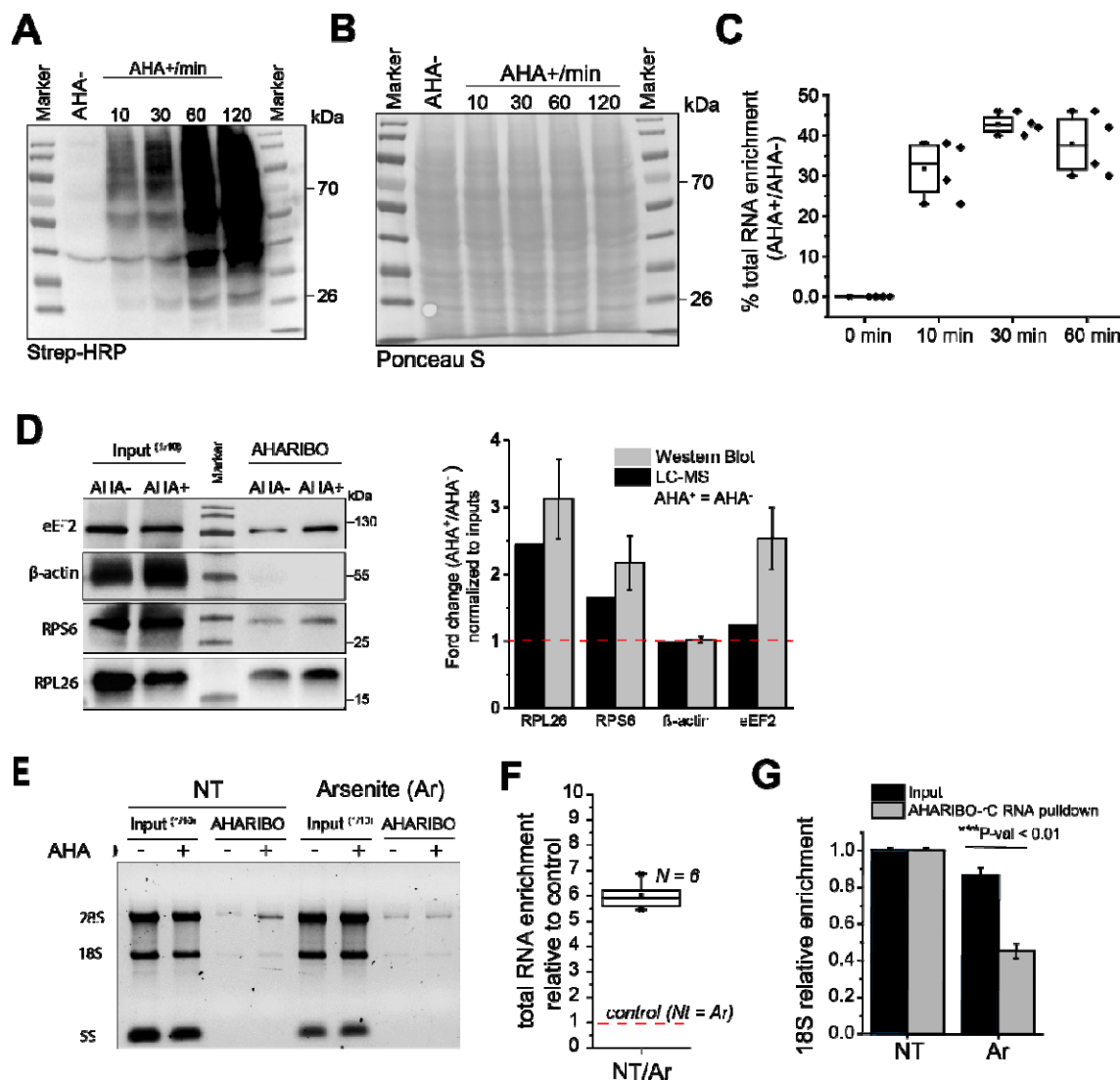
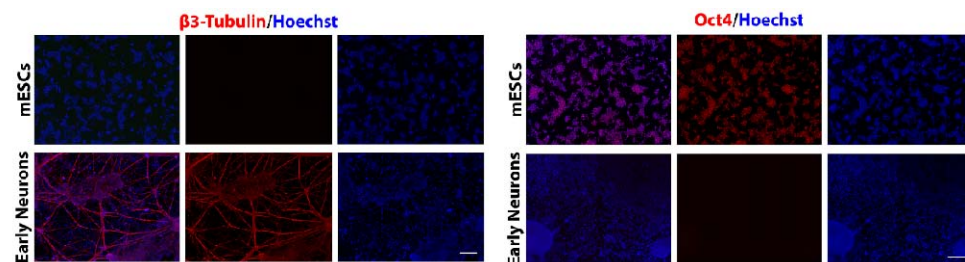


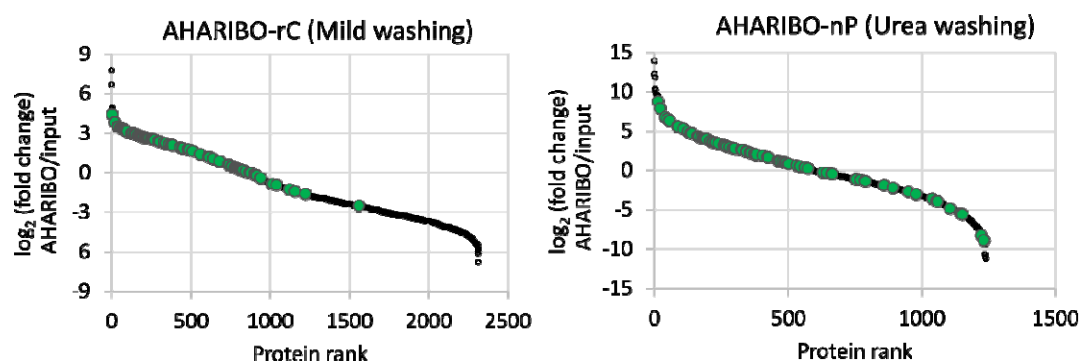
Figure S1. AHA incorporation, validation of AHA and RNA capture. A) Labeling of nascent peptides in cells treated with AHA (250 μ M) at different incubation times (10, 30, 60, 120 min). After SDS-PAGE of cell extracts, AHA residues were biotinylated by on-membrane cycloaddition based “click chemistry” and detected by streptavidin-HRP. B) Ponceau S staining of the membrane reported in A. C) RNA enrichment in AHARIBO-tC 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) Immunoblotting analysis of proteins captured by AHARIBO-rC and obtained from cells either treated or not treated with AHA. A lysate volume corresponding to 1/10th of the volume processed for pulldown was used as reference inputs. On the right, histogram reporting the relative abundance of RPL26, eEF2, beta-actin and RPS6 in LC-MS analysis and quantitative analysis of Western Blot data (n = 3; error bars, \pm SD). E) 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. NT, non-treated cells; Ar, arsenite-treated cells. F) Total RNA enrichment after AHARIBO-rC pulldown of lysates obtained from unstimulated cells over cells treated with arsenite. 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. Square box, mean; stars, 1-99% percentile. G) 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.

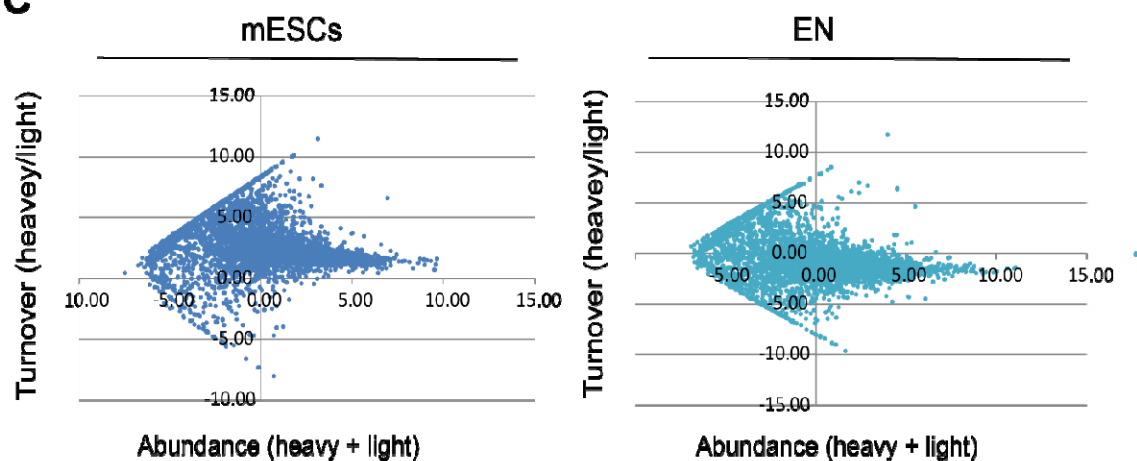
A



B



C



D

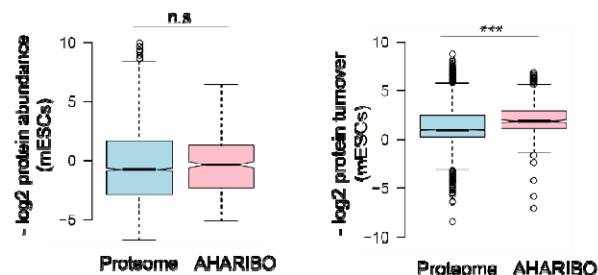


Figure S2. Proteomic analysis. A) Immunofluorescence for mESC (Oct4) and neuronal (β 3-tubulin) marker expression on self-renewing mESCs and 15DIV mESC-derived neurons. Scale bar: 200 μ m. B) Rank plot of fold change of full proteome (black dots) and ribosomal proteins (green dots), comparing AHARIBO 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 synthesized 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-Val < 0.001.

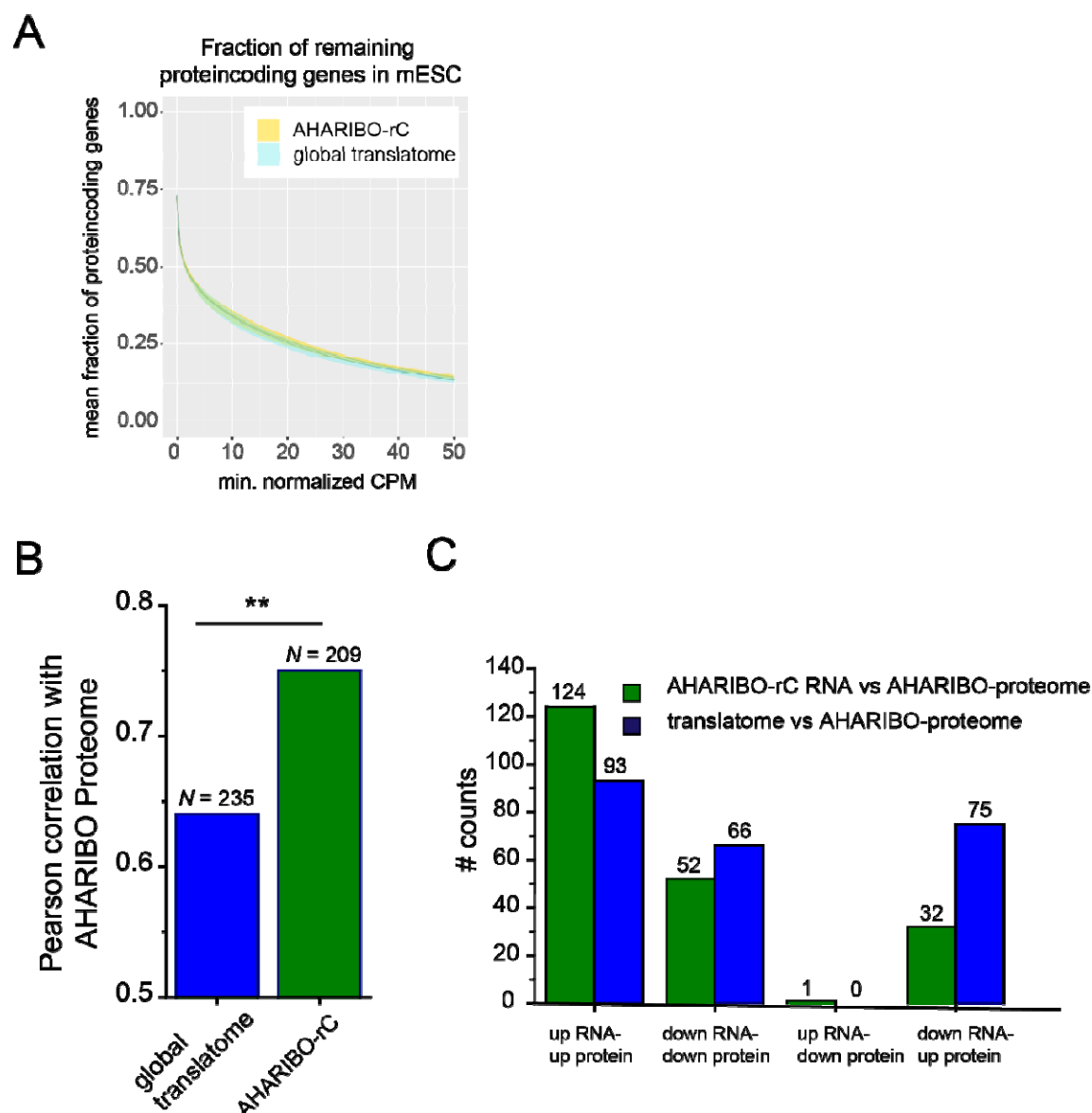


Figure S3. RNA-seq and protein coding RNA analysis. A) Fraction of remaining protein coding genes at increasing CPM thresholds in mESCs. The mean of $n = 3$ biological replicates is reported. Shadows: data are represented as mean \pm SD. B) Histogram showing Pearson's correlation analysis of AHARIBO-nP protein fold change (EN/mESC) determined by mass spectrometry versus global translome and AHARIBO-rC RNA fold change (EN/mESC) determined by RNA-seq. N, number of DEGs. P-val < 0.05 . C) Histogram of the number of DEGs (EN/mESCs) up- and down-regulated in AHARIBO-rC RNA or global translome relative to the AHARIBO-nP proteome.

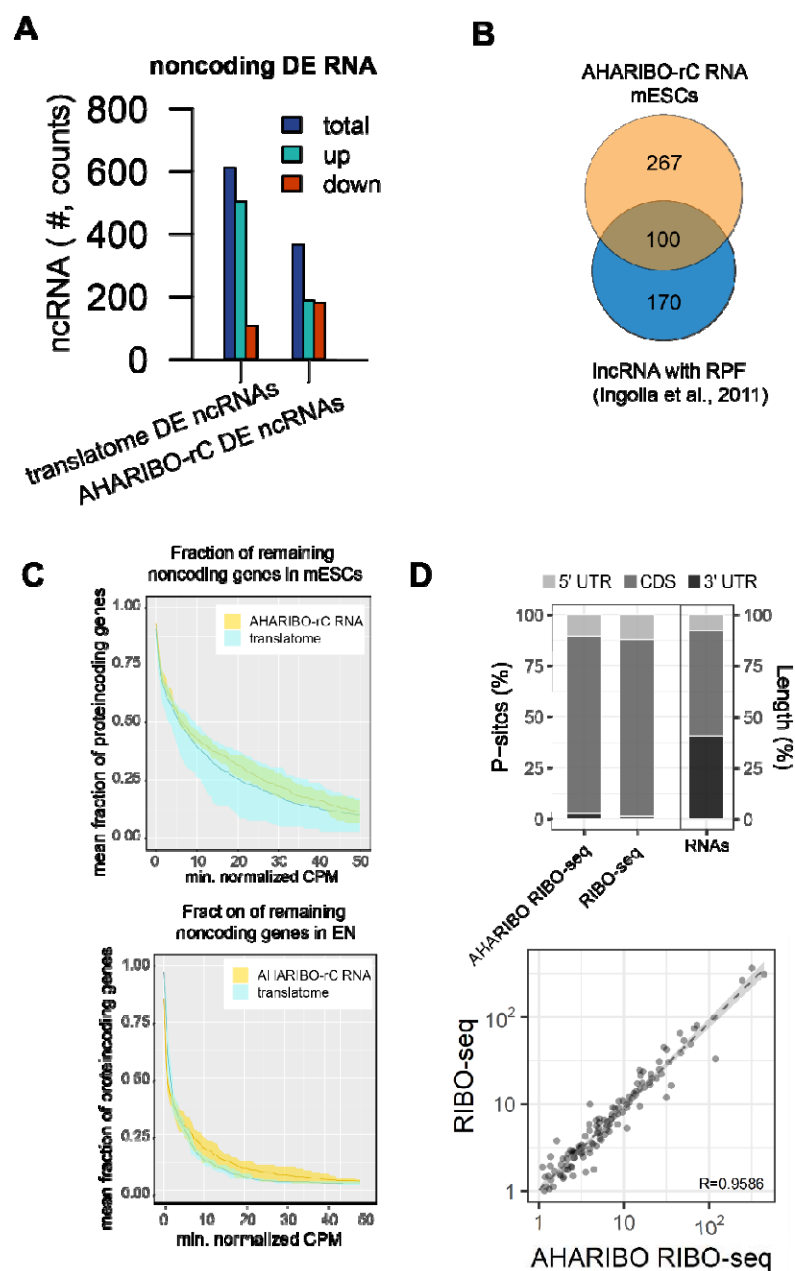


Figure S4. Isolation of lncRNAs with AHARIBO. A) Number of up- and down-regulated differentially expressed non-coding RNAs in the global translatome and in AHARIBO-rC RNA DE, differentially expressed. ncRNA, non-coding RNA. B) Venn diagram representing the number of differentially expressed lncRNAs identified by AHARIBO-rC (orange) and number of lncRNAs with at least 1 CPM in Ingolia et al., 2011 (blue). C) Fraction of remaining non-coding genes at increasing CPM thresholds in AHARIBO-rC versus global translatome RNA from both mESCs (top) and ENs (bottom). The mean of $n = 3$ biological replicates is reported. Shadows: data are represented as mean \pm SD. D) 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 (top). 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 (bottom).

Materials and Methods

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% CO₂. 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 (0.1M NaCl, 0.1 M MgCl₂, 0.1M 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 μ M 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, 5×10^5 mESCs/cm² 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 μ M 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/cm² 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 μ M AHA for 30 min. The remaining treatment steps and hypotonic lysis were performed as detailed above.

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- β 3-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 MgCl₂, 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; 50 μ M final concentration) in a reaction volume of 100 μ l for 1 h on a rotator in slow motion (9 rpm) at 4°C. Lysates were then incubated with 50 μ l 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 μ l 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 µl 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 µl. 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 mRNA 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.

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 36 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 analyzed 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 translationalome)

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 (Ingolia et al., 2012b). 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 μ l of 30% sucrose cushion (30 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM MgCl₂, 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). Supernatants were saved if needed for further analysis, and pellets were resuspended in 100 μ l of nuclease-free water supplemented with 30 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM MgCl₂.

Protein extraction from sucrose gradient fractions

Polysomal fractions (1 mL) or pellet/supernatant fractions from 30% sucrose cushioning (1/5th of total amount, adjusted to 260 μ l volume) were processed for methanol/chloroform protein extraction. Briefly, 600 μ l of methanol and 150 μ l of chloroform were added to each sample and samples were vortexed. Then, 450 μ l 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 μ l 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; 50µM 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

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-eEF2 (Abcam #ab33523; 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 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 analyzed 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 1x10⁶. MS/MS was performed in the ion trap using the top speed mode (3 secs), an AGC target of 5x10³ and an HCD collision energy of 30.

MS raw files were analyzed 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 and 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 etc 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 Superscript-III 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 waterbath for 75 min before acid phenol:chloroform:isoamyl alcohol (pH 4.5) RNA extraction.

AHARIBO RIBO-seq. The mixture was incubated with Dibenzocyclooctyne-PEG4-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 µL 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 phenol:chloroform:isoamyl alcohol 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 BioTechnology #RL001_mod2) and sequenced on an Illumina HiSeq 4000 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 INFORMATION

Funding

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

Author contributions

L.M. suggested the original workflow. M.C. conceived the experiments. A.P. prepared the RNA libraries for RNA-seq. C.F., L.M., A.P., A.D.P. performed the experiments. R.B., D.P., performed proteomics experiments. J.Z., cultured mESC cells, performed the 10-day neuronal differentiation and collected immunofluorescence images. A.Q. suggested the biological model of neuronal differentiation and made available mESC cells. M.C., S.S., R.B., D.P. analyzed proteomic data. S.S. identified lncRNA peptides from result spectra. F.G., A.R., M.C. analyzed RNA-seq data. G.V. and F.L. analyzed RIBO-seq and AHARIBO RIBO-seq data. L.M., A.P., M.C. drafted the manuscript. M.C. and G.V. reviewed the manuscript. M.C., A.P., P.B., L.M., S.S., A.R., C.F., A.D.P., J.Z., F.G., G.V., G.G. reviewed the final manuscript. All of the authors read and approved the final manuscript.

DECLARATION OF INTERESTS

M.C. is the founder of, director of, and a shareholder in IMMAGINA BioTechnology S.r.l., a company engaged in the development of new technologies for gene expression analysis at the ribosomal level. L.M., A.P., C.F., A.D.P., P.B. are employees of IMMAGINA BioTechnology S.r.l. A.Q. and G.G. are shareholders of IMMAGINA BioTechnology S.r.l. G.V. is a scientific advisor of IMMAGINA BioTechnology S.r.l. All of the other authors declare no competing interests.

References

- Amit Blumberg, Yixin Zhao, Yi-Fei Huang, Noah Dukler, Edward J. Rice, Katie Krumholz, Charles G. Danko, V.O.P.S. (2019). Characterizing RNA stability genome-wide through combined analysis of PRO-seq and RNA-seq data. *BioRxiv*.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Anderson, D.M., Anderson, K.M., Chang, C.-L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R., et al. (2015). A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell* 160, 595–606.
- Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 3889–3894.
- Aspden, J.L., Eyre-Walker, Y.C., Phillips, R.J., Amin, U., Mumtaz, M.A.S., Brocard, M., and Couso, J.-P. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *Elife* 3.
- Aviner, R., Geiger, T., and Elroy-Stein, O. (2013). PUNCH-P for global translome profiling. *Translation* 1, e27516.
- Bagert, J.D., Xie, Y.J., Sweredoski, M.J., Qi, Y., Hess, S., Schuman, E.M., and Tirrell, D.A. (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.
- Bazin, J., Baerenfaller, K., Gosai, S.J., Gregory, B.D., Crespi, M., and Bailey-Serres, J. (2017). Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. *Proc. Natl. Acad. Sci.* 114, E10018–E10027.
- Bazzini, A.A., Johnstone, T.G., Christiano, R., Mackowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., et al. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* 33, 981–993.
- Bhatta, A., Atianand, M., Jiang, Z., Crabtree, J., Blin, J., and Fitzgerald, K.A. (2020). A Mitochondrial Micropeptide Is Required for Activation of the Nlrp3 Inflammasome. *J. Immunol.* 204, 428–437.
- Biever, A., Glock, C., Tushev, G., Ciirdaeva, E., Dalmay, T., Langer, J.D., and Schuman,

- E.M. (2020). Monosomes actively translate synaptic mRNAs in neuronal processes. *Science* (80-.). 367, eaay4991.
- Biscarini, S., Caputo, D., Peruzzi, G., Lu, L., Colantoni, A., Santini, T., Shneider, N.A., Caffarelli, E., Laneve, P., and Bozzoni, I. (2018). Characterization of the lncRNA transcriptome in mESC-derived motor neurons: Implications for FUS-ALS. *Stem Cell Res.* 27, 172–179.
- Blobel, G., and Sabatini, D. (1971). Dissociation of Mammalian Polyribosomes into Subunits by Puromycin. *Proc. Natl. Acad. Sci.* 68, 390–394.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Carelli, S., Giallongo, T., Rey, F., Latorre, E., Bordoni, M., Mazzucchelli, S., Gorio, M.C., Pansarasa, O., Provenzani, A., Cereda, C., et al. (2019). HuR interacts with lincBRN1a and lincBRN1b during neuronal stem cells differentiation. *RNA Biol.* 16, 1471–1485.
- Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L.A., and Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* 22, 867–882.
- Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491, 454–457.
- Chandrasekaran, V., Juskiewicz, S., Choi, J., Puglisi, J.D., Brown, A., Shao, S., Ramakrishnan, V., and Hegde, R.S. (2019). Mechanism of ribosome stalling during translation of a poly(A) tail. *Nat. Struct. Mol. Biol.* 26, 1132–1140.
- Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., et al. (2020). Pervasive functional translation of noncanonical human open reading frames. *Science* (80-.). 367, 1140–1146.
- Clamer, M., Tebaldi, T., Lauria, F., Bernabò, P., Gómez-Biagi, R.F., Marchioretto, M., Kandala, D.T., Minati, L., Perenthaler, E., Gubert, D., et al. (2018). Active Ribosome Profiling with RiboLace. *Cell Rep.* 25, 1097-1108.e5.
- D’Lima, N.G., Ma, J., Winkler, L., Chu, Q., Loh, K.H., Corpuz, E.O., Budnik, B.A., Lykke-Andersen, J., Saghatelian, A., and Slavoff, S.A. (2017). A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* 13, 174–180.
- Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A., and Schuman, E.M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci.* 103, 9482–9487.

- Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Eden, E., Geva-Zatorsky, N., Issaeva, I., Cohen, A., Dekel, E., Danon, T., Cohen, L., Mayo, A., and Alon, U. (2011). Proteome Half-Life Dynamics in Living Human Cells. *Science* (80-.). 331, 764–768.
- Frith, M.C., Forrest, A.R., Nourbakhsh, E., Pang, K.C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Bailey, T.L., and Grimmond, S.M. (2006). The Abundance of Short Proteins in the Mammalian Proteome. *PLoS Genet.* 2, e52.
- Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A., and Couso, J.P. (2007). Peptides Encoded by Short ORFs Control Development and Define a New Eukaryotic Gene Family. *PLoS Biol.* 5, e106.
- Garreau de Loubresse, N., Prokhorova, I., Holtkamp, W., Rodnina, M. V., Yusupova, G., and Yusupov, M. (2014). Structural basis for the inhibition of the eukaryotic ribosome. *Nature* 513, 517–522.
- Grollman, A.P. (1967). Inhibitors of protein biosynthesis. II. Mode of action of anisomycin. *J. Biol. Chem.* 242, 3226–3233.
- Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482, 339–346.
- van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J.F., Adami, E., Faber, A.B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.-L., et al. (2019). The Translational Landscape of the Human Heart. *Cell* 178, 242-260.e29.
- Huang, J.-Z., Chen, M., Chen, D., Gao, X.-C., Zhu, S., Huang, H., Hu, M., Zhu, H., and Yan, G.-R. (2017). A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol. Cell* 68, 171-184.e6.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012a). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550.

- Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012b). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550.
- Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208.
- Jao, C.Y., and Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U. S. A.* 105, 15779–15784.
- Jewett, J.C., and Bertozzi, C.R. (2010). Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* 39, 1272–1279.
- Jiao, Y., and Meyerowitz, E.M. (2010). Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol. Syst. Biol.* 6, 419.
- Kapur, M., Monaghan, C.E., and Ackerman, S.L. (2017). Regulation of mRNA Translation in Neurons—A Matter of Life and Death. *Neuron* 96, 616–637.
- Kleaveland, B., Shi, C.Y., Stefano, J., and Bartel, D.P. (2018). A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. *Cell* 174, 350-362.e17.
- Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., and Qian, S.-B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci.* 109, E2424–E2432.
- Magny, E.G., Pueyo, J.I., Pearl, F.M.G., Cespedes, M.A., Niven, J.E., Bishop, S.A., and Couso, J.P. (2013). Conserved Regulation of Cardiac Calcium Uptake by Peptides Encoded in Small Open Reading Frames. *Science* (80-.). 341, 1116–1120.
- Martinez, A., Traverso, J.A., Valot, B., Ferro, M., Espagne, C., Ephritikhine, G., Zivy, M., Giglione, C., and Meinel, T. (2008). Extent of N-terminal modifications in cytosolic proteins from eukaryotes. *Proteomics* 8, 2809–2831.
- Mathias, A.P., Williamson, R., Huxley, H.E., and Page, S. (1964). Occurrence and function of polysomes in rabbit reticulocytes. *J. Mol. Biol.* 9, 154-IN11.
- Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., et al. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* (80-.). 349, aab2276–aab2276.
- Morisaki, T., Lyon, K., DeLuca, K.F., DeLuca, J.G., English, B.P., Zhang, Z., Lavis, L.D., Grimm, J.B., Viswanathan, S., Looger, L.L., et al. (2016). Real-time quantification of single RNA translation dynamics in living cells. *Science* (80-.). 0899, aaf0899.

- Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R., Chen, X., Kavalali, E.T., et al. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* (80-.). 351, 271–275.
- Pircher, A., Gebetsberger, J., and Polacek, N. (2014). Ribosome-associated ncRNAs: An emerging class of translation regulators. *RNA Biol.* 11, 1335–1339.
- Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11, R25.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Rothenberg, D.A., Taliaferro, J.M., Huber, S.M., Begley, T.J., Dedon, P.C., and White, F.M. (2018). A Proteomics Approach to Profiling the Temporal Translational Response to Stress and Growth. *IScience* 9, 367–381.
- Ruiz-Orera, J., Messeguer, X., Subirana, J.A., and Alba, M.M. (2014). Long non-coding RNAs as a source of new peptides. *Elife* 3, e03523.
- Schwanhäusser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009). Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* 9, 205–209.
- Seedhom, M.O., Hickman, H.D., Wei, J., David, A., and Yewdell, J.W. (2016). Protein Translation Activity: A New Measure of Host Immune Cell Activation. *J. Immunol.* 197, 1498–1506.
- Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L., and Saghatelian, A. (2013). Peptidomic discovery of short open reading frame–encoded peptides in human cells. *Nat. Chem. Biol.* 9, 59–64.
- Slavoff, S.A., Heo, J., Budnik, B.A., Hanakahi, L.A., and Saghatelian, A. (2014). A Human Short Open Reading Frame (sORF)-encoded Polypeptide That Stimulates DNA End Joining. *J. Biol. Chem.* 289, 10950–10957.
- van Solingen, C., Sharma, M., Bijkerk, R., Afonso, M.S., Koelwyn, G.J., Scacalossi, K.R., van Zonneveld, A.J., and Moore, K.J. (2018). Abstract 027: A Micropeptide Concealed in a Putative Long Non-coding RNA Directs Inflammation. *Arterioscler. Thromb. Vasc. Biol.* 38.
- Sun, Q., Hao, Q., and Prasanth, K. V (2018). Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. *Trends Genet.* 34, 142–157.
- Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in

Single Cells. *Science* (80-.). 329, 533–538.

Tebaldi, T., Zuccotti, P., Peroni, D., Köhn, M., Gasperini, L., Potrich, V., Bonazza, V., Dudnakova, T., Rossi, A., Sanguinetti, G., et al. (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.

Wang, C., Han, B., Zhou, R., and Zhuang, X. (2016). Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. *Cell* 165, 990–1001.

Wang, T., Cui, Y., Jin, J., Guo, J., Wang, G., Yin, X., He, Q.-Y., and 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.

Wild, K., Aleksić, M., Lapouge, K., Juaire, K.D., Flemming, D., Pfeffer, S., and Sinning, I. (2020). MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. *Nat. Commun.* 11, 776.

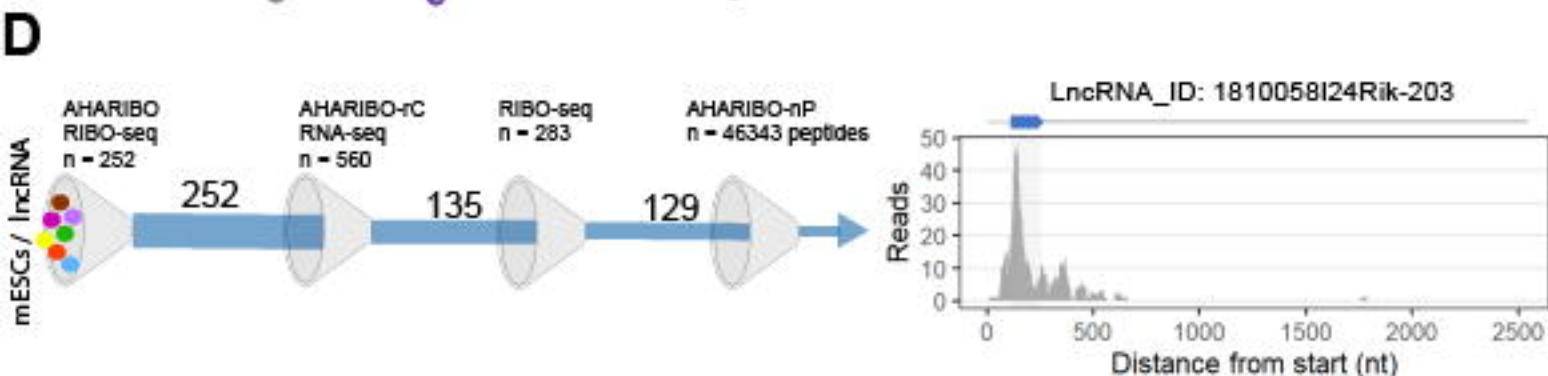
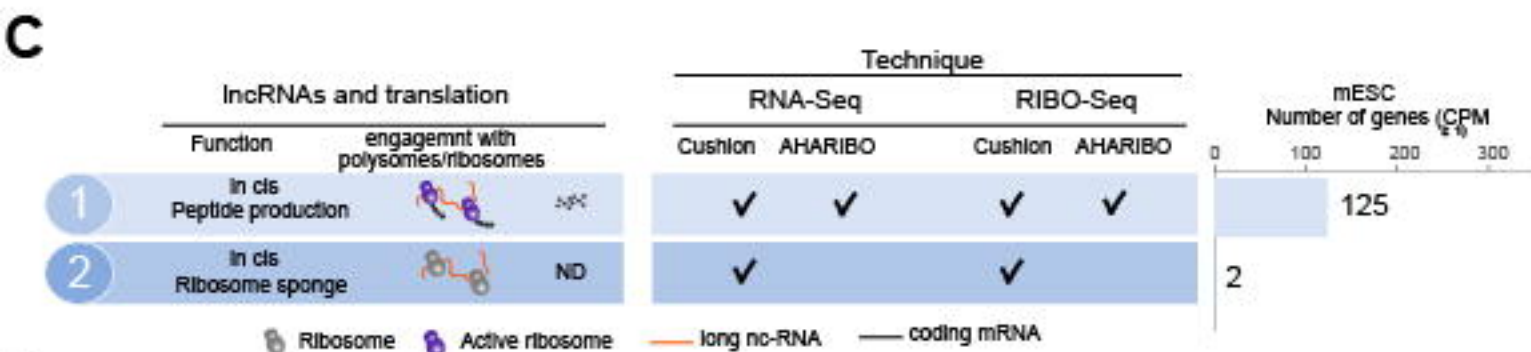
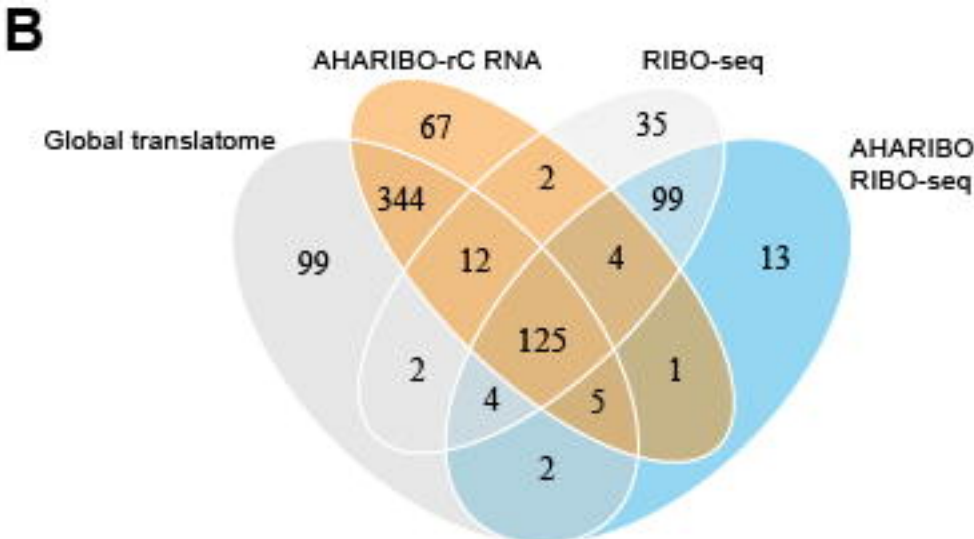
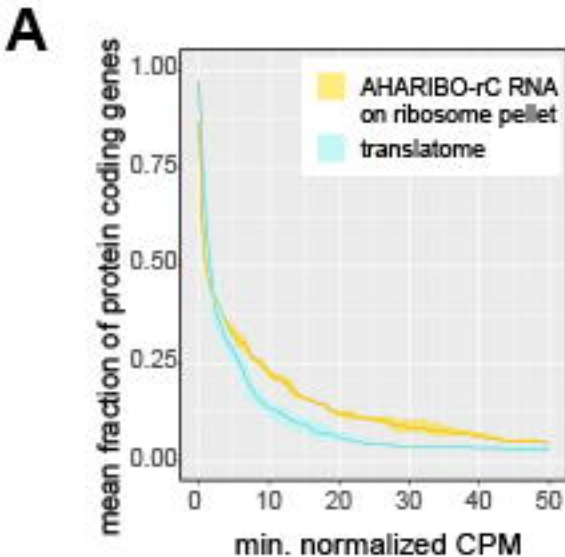
Wu, B., Eliscovich, C., Yoon, Y.J., and Singer, R.H. (2016). Translation dynamics of single mRNAs in live cells and neurons. *Science* (80-.). 352, 1430–1435.

Yan, X., Hoek, T.A., Vale, R.D., and Tanenbaum, M.E. (2016). Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* 165, 976–989.

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

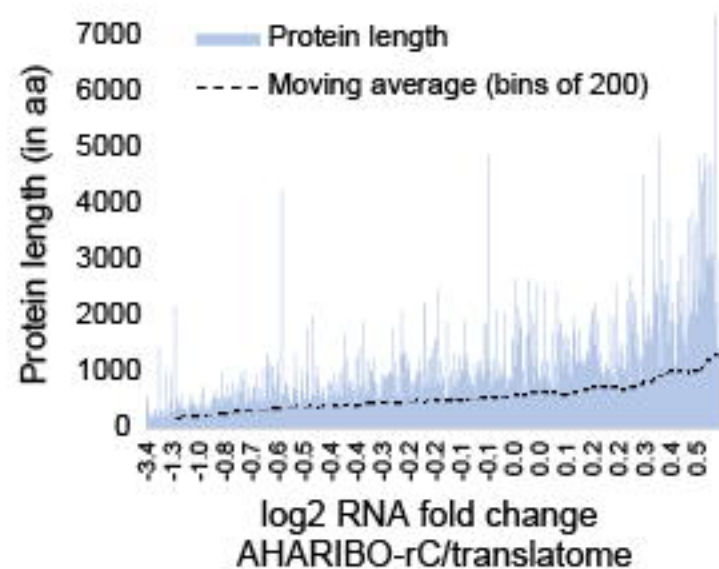
Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 Suppresses Target mRNA Translation. *Mol. Cell* 47, 648–655.

Zeng, C., Fukunaga, T., and Hamada, M. (2018). Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. *BMC Genomics* 19, 414.

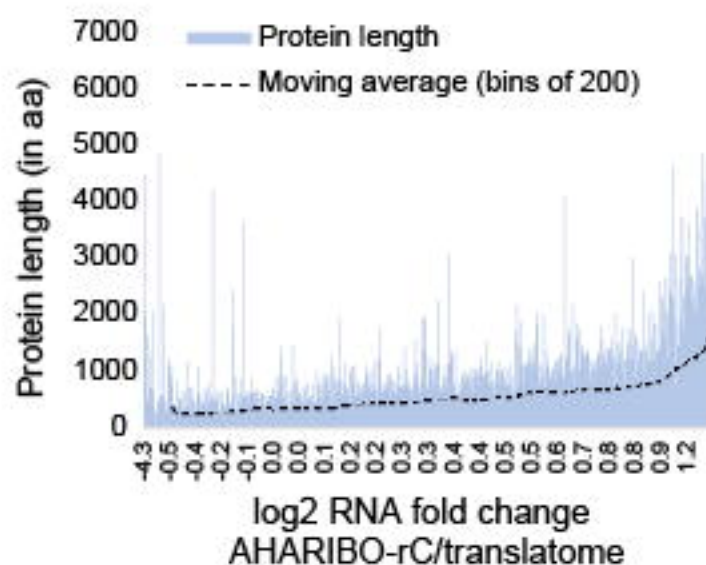
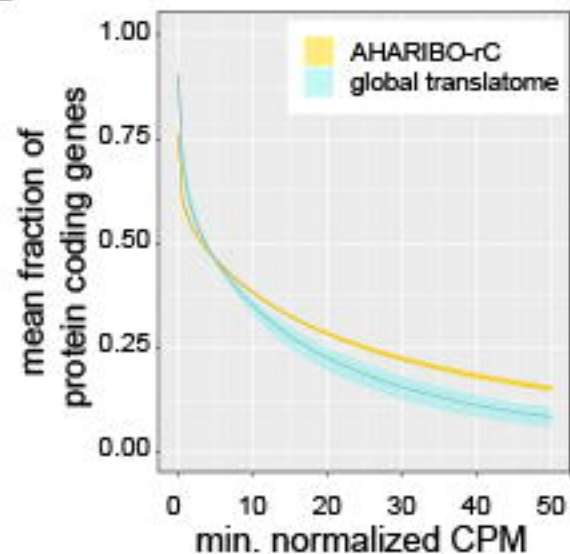
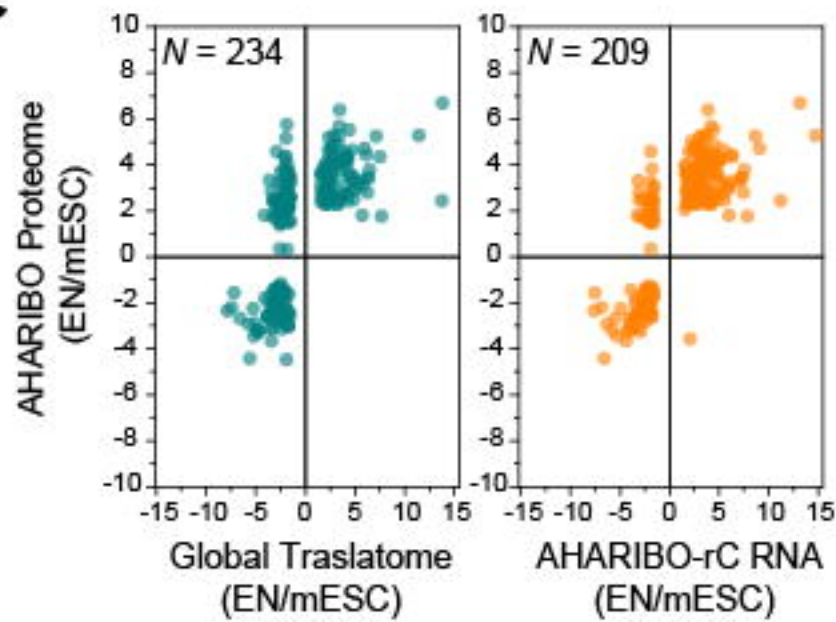


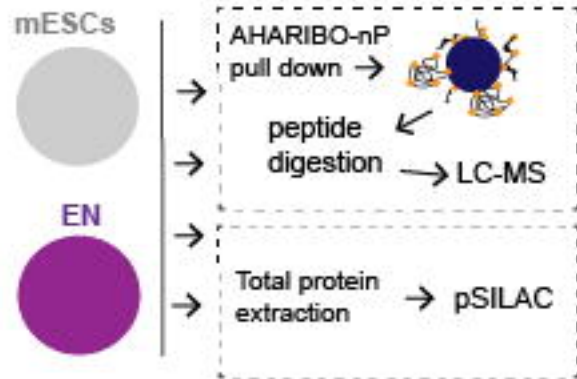
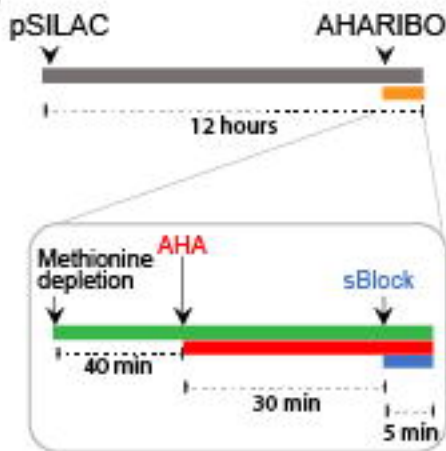
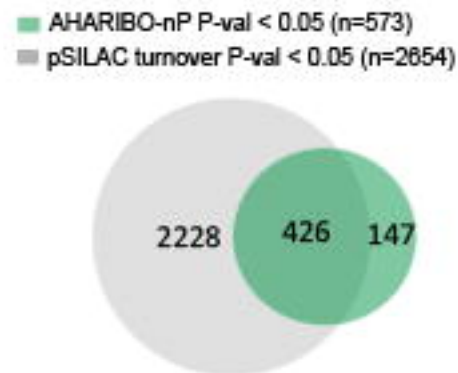
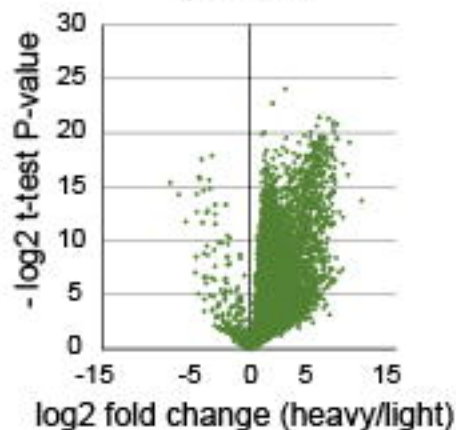
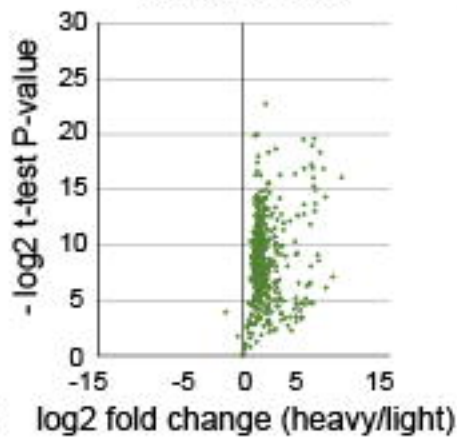
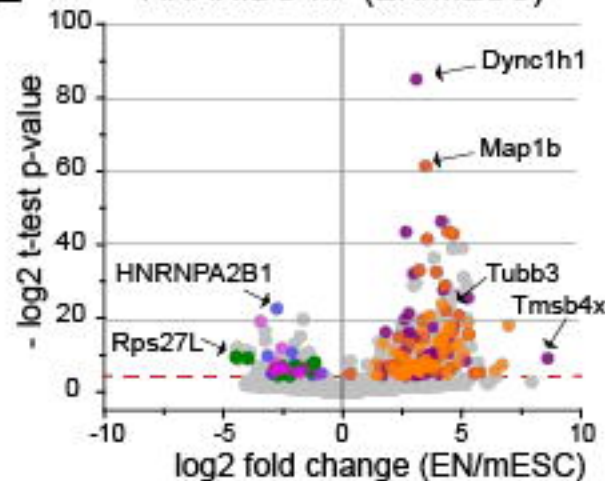
A

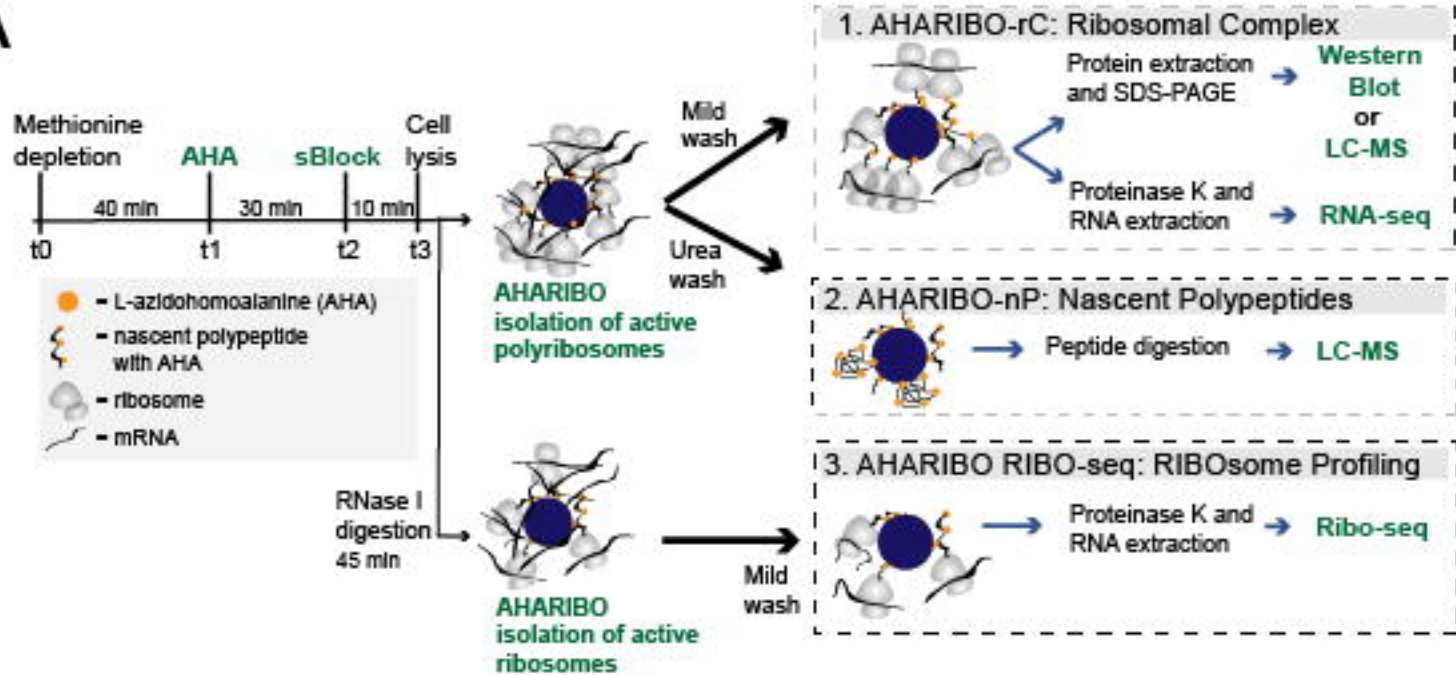
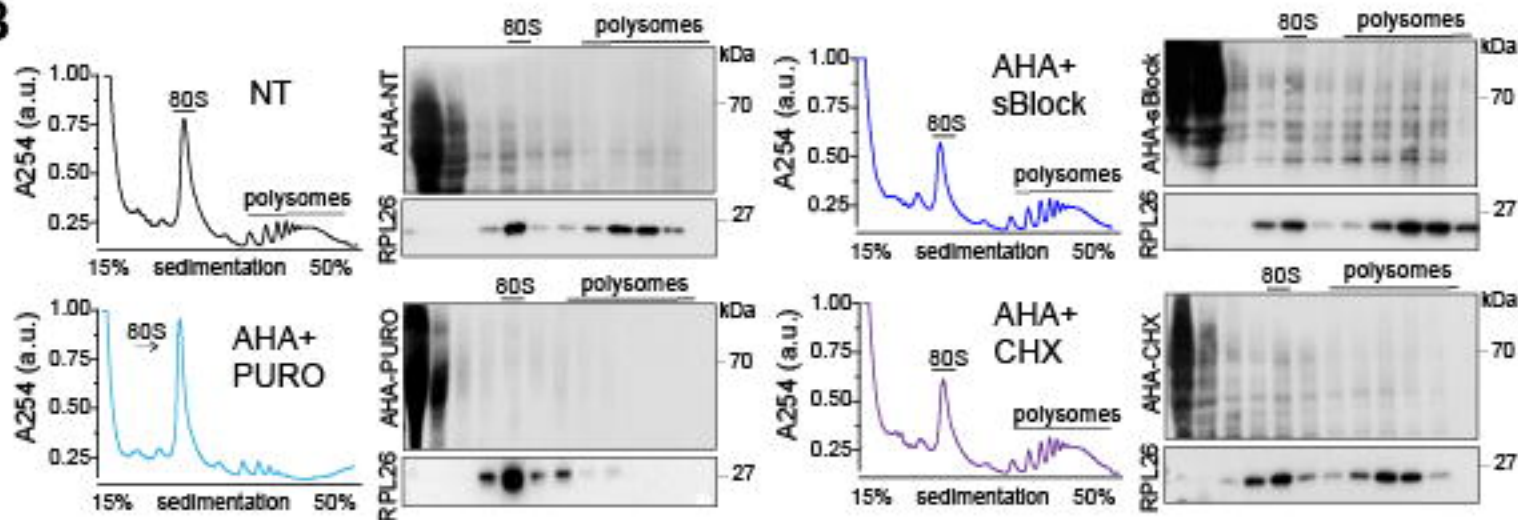
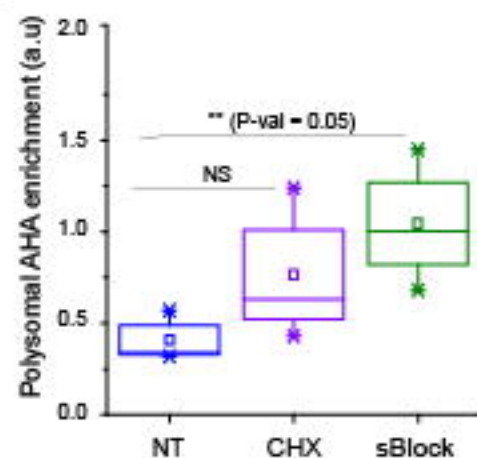
mESCs



EN

**B****C**

A**B****C****D****pSILAC****AHARIBO-nP****E****AHARIBO-nP (EN/mESC)**

A**B****C****D**