A Sequence determinant in 3'UTR of mRNAs for Nuclear Retention by Paraspeckles

Audrey Jacq[1], Denis Becquet[1], Bénédicte Boyer[1], Séverine Guillen[1], Maria-Montserrat Bello-Goutierrez[1], Marie-Pierre Blanchard[1], Claude Villard[1], Maya Belghazi[1], Manon Torres[1], Jean-Louis Franc[1] and Anne-Marie François-Bellan[1,*]

[1] INP, AMU-CNRS UMR 7051, 27 Bd Jean Moulin Faculté des Sciences Médicales et Paramédicales, Campus Timone, 13385 Marseille cedex 05

* To whom correspondence should be addressed. Tel: 33 491 698 807; Fax: 33 491 698 920; Email: anne-marie.francois@univ-amu.fr

Present Address:

Marie-Pierre Blanchard, Montpellier Ressources Imagerie, BioCampus, University of Montpellier, CNRS, INSERM, Montpellier, France;

Manon Torres, Charité–Universitätsmedizin Berlin Campus Charité Mitte Chronobiology in the Charité CrossOver (CCO) Virchowweg 6 Charitéplatz 1 10117 Berlin Germany.

ABSTRACT

Paraspeckles are nuclear membraneless structures composed of a long non-coding RNA, Nuclear-Enriched-Abundant-Transcript-1 and RNA binding proteins, which associate numerous mRNAs. It is therefore believed that their cellular function is to sequester in the nucleus their associated proteins and/or target mRNAs. However, little is known about the molecular determinant in mRNA targets that allow their association to paraspeckles except that inverted repeats of Alu sequences (IRAlu) present in 3'UTR of mRNAs may allow this association. However, we didn't find IRAlus in 3'UTR of the paraspeckle target RNAs, we previously identified in a pituitary cell line. Since 3'UTR of mRNA are known to contain sequence recognition for nuclear localization, we search for other potential RNA recruitment elements in the 3'UTR of a candidate paraspeckle target gene, namely calreticulin mRNA. We found a 15 nucleotides sequence, present as a tandem repeat in 3'UTR of this mRNA, which is involved in the nuclear retention by paraspeckles. While being not overrepresented in the 3'UTR of the paraspeckle target RNAs, this recruitment element was present in near 30% of 3'UTR mRNAs. In addition, since this sequence binds the paraspeckle protein component HNRNPK, it is proposed that HNRNPK constitute a bridging protein between paraspeckles and target mRNAs.

INTRODUCTION

The paraspeckles, discovered in 2002 (1) are nuclear bodies found in almost all of the cultured cell lines and primary cultures from tissues (2), except for embryonic stem cells (3). They are usually detected as a variable number of discrete dots found in close proximity to nuclear speckles (2). The structural element of paraspeckles is a long non-coding RNA, nuclear-enriched abundant transcript one (Neat1) which exists under a short and a long transcript generated from the same promoter, previously identified as MENε (Neat1–1) and MENβ (Neat1–2), respectively (4)(5). More than 60 proteins have been identified thus far to be associated with paraspeckles (6), among which 4 RNA-binding proteins, including 3 members of the Drosophila Melanogaster Behavior Human Splicing (DBHS) family proteins (NONO, PSPC1 and SFPQ) and RNA-binding motif protein 14 (RBM14) are considered as core paraspeckle proteins (2)(7)(8). Another crucial paraspeckle protein, the heterogeneous nuclear ribonucleoprotein K (HNRNPK), appeared to affect the production of the essential NEAT1-2 isoform by negatively regulating the 3'-end polyadenylation of the NEAT1-1 isoform (9). Paraspeckles have been also shown to associate numerous mRNA. It is therefore believed that the cellular function of paraspeckles is to sequester in the nucleus their associated proteins and/or their target mRNAs. In a rat pituitary cell line, the GH4C1 cells, we identified using a Neat1 RNA pull-down procedure (10)

followed by RNA sequencing, 3928 RNAs that are paraspeckle's targets (11). In these cells, we have shown that the expression of all major components of paraspeckles including Neat1 and core paraspeckle proteins followed a circadian pattern that leads to rhythmic variations in paraspeckle number within these cells (11). Thanks to their circadian expression pattern and given their functions in mRNA nuclear retention, it was shown that paraspeckles rhythmically retain target RNAs in the nucleus of the cells and that this rhythmic nuclear retention leads to the rhythmic expression of the corresponding genes (11)(12). However, how mRNAs are bound to paraspeckles is largely unknown. In particular, the molecular sequence determinants that allowed mRNAs to be paraspeckle's targets are poorly understood even if it has been shown that inverted repeats of Alu sequences (IRAlu) present in 3'UTR of mRNAs allow these mRNAs to associated with paraspeckles (3)(7)(13). Indeed the insertion of IRAlus in the 3'-UTR of an Egfp reporter gene induced both a nuclear retention of Egfp mRNA and a reduction of EGFP protein expression in the cytoplasm (13)(11). Furthermore, endogenous mRNAs containing IRAlus in their 3'-UTRs like Nicolin 1 (NICN1) or Lin 28 have been shown to be retained in the nucleus by paraspeckles (13)(14) providing conclusive evidence that the presence of IRAlus in 3'-UTRs of genes allows paraspeckles to sequester mature mRNAs within the nucleus.

In human cells, hundreds of genes contain inverted repeated short interspersed nuclear elements (SINEs) (mainly Alu elements) in their 3'-UTRs. Alu elements are unique to primates and account for almost all of the human SINEs and >10% of the genome. Their abundance leads to the frequent occurrence of IRAlus in gene regions. Unexpectedly however, in 3'-UTR of the 3928 RNAs that are Neat1 targets in rat GH4C1 cells, we didn't find inverted repeated SINEs, the rat equivalent of primate IRAlus (unpublished results). The question whether other sequence determinants than IRAlus in 3'UTR of mRNAs allow them to be paraspeckle's targets remains then unanswered. To identify potential RNA recruitment elements, we chose to adopt a gene candidate strategy based on the analysis of the properties of the candidate gene 3'-UTR to associate to paraspeckles. To this end, we selected a candidate gene, namely calreticulin (Calr) mRNA, that was previously shown to be both a post-transcriptional cycling transcript in the liver (15) and a mRNA paraspeckle target in our cell line (11). In the 3'UTR of Calr mRNA, we identified a 15 nucleotides sequence, found as a tandem repeat, which was involved in the nuclear retention by paraspeckles. This sequence determinant was further shown to bind the paraspeckle protein component HNRNPK and to be present in near 30% of the 3'UTR of the 3928 paraspeckle target RNAs. Accordingly, HNRNPK may constitute a bridging protein between paraspeckles and target mRNAs through binding to this 15 nucleotides sequence.

MATERIAL AND METHODS

Cell line culture and preparation of stably transfected cell lines

GH4C1 cells, a rat pituitary somatolactotroph line, were obtained in 2012 from ATCC (CCL-82.2, lot number: 58945448, Molsheim, France) with certificate analysis and were confirmed to be free of mycoplasma (MycoAlert, Lonza, Levallois-Perret, France). They were grown in HamF10 medium supplemented with 15% horse serum and 2% fetal calf serum. GH4C1 cells were synchronized between themselves by a replacement of fresh medium. For the generation of stable GH4C1 cell lines, cells were transfected with the plasmid constructs expressing a neomycin resistance gene by Lipofectamine 3000 (Invitrogen, Cergy Pontoise, F). Cells were selected with 250 mg/ml G418 (Invitrogen) beginning 48 h after transfection.

Plasmid constructs

3'UTR of Calr (3'UTR-Calr) mRNA (574 bases) was amplified using classic 3' RACE by dual PCR from total RNA obtained from GH4C1 cells (see Supplemental Table 1 for primers). Sequence was then inserted in the plasmid pEGFP-C1 (Clontech, Mountain View, CA) at Sma1 site. Using this 3'UTR construction, PCR were performed in order to obtain three overlapping subsequences named C1, C2 and C3 (see Supplemental Fig.2). These sequences were inserted using SmaI site in pEGFP-C1. Primers (IDT, Coralville, Iowa, USA) used for these constructions were listed in Supplemental Table 1. All sequences were controlled by sequencing.

Mutagenesis

The reporter plasmid pEGFP-C1 containing the entire 3'UTR-Calr mRNA was mutated in the 15 nucleotides sequence present in position 37-51 ($C1_M$) or in position 466-480 ($C3_M$) or in both positions ($C1/3_M$) by inverse PCR using the oligonucleotides listed in Supplemental Table 1. The mutant cDNA sequences were controlled by sequencing.

RNA expression analysis

From GH4C1 cells, nuclear and cytoplasmic RNA were prepared using the Nucleospin RNA XS kit (Macherey Nagel, Hoerdt, France) and total RNA was prepared using the Nucleospin RNA kit (Macherey Nagel). Nuclear and cytoplasmic RNA isolation was performed using 10 cm cell dishes that were rinsed twice with ice-cold PBS, incubated in 1 ml of ice-cold cell lysis buffer A (10 mmol/L Tris pH 7.4, 3 mmol/L MgCl2, 10 mmol/L NaCl and 0.5% NP-40). Nuclei and cytoplasma were

separated by centrifugation (500 g for 5 min). One-sixth of the supernatant was used to prepare cytoplasmic RNA. To obtain pure nuclear RNA, the nuclear pellets were subjected to two additional washes with 1 ml lysis buffer A and were then extracted with XS kit reagent. Total RNA (500 ng) was then used for cDNA synthesis performed with a High Capacity RNA to cDNA kit (Applied Biosystem, Courtaboeuf, France). Real time PCR was performed on a 7500 fast Real-Time qPCR system (Applied Biosystems) using Fast SYBR Green mix (Applied Biosystems). The sequences of the primers used in qPCR are given in Supplemental Table 1. mRNA accumulation given as the relative ratio of nuclear versus cytoplasmic mRNA was normalized relative to Gapdh mRNA levels.

RNA immunoprecipitation (RIP) experiments

GH4C1 cells grown in 10 cm dishes were rinsed two times with 5 ml cold phosphate buffer saline (PBS). Cells were then harvested by scraping in ice-cold PBS and transferred to a centrifuge tube. After centrifugation (2500 g for 5 min) cells were pelleted and suspended in 100 µl of Polysome lysis buffer (PLB; 10 mM HEPES, pH 7.0, 0.1M KCl, 5 mM MgCl2, 0.5% NP40, 1 mM DTT, 100 U/ml RNAse OUT and complete protease inhibitor cocktail). After mixing by pipetting up and down, cells were kept on ice for 5 min to allow the hypotonic PLB buffer to swell the cells. The cell lysate was then aliquoted and stored at -80°C.

Cell lysate was centrifuged at 14,000 g for 10 min at 4°C and diluted 1/100 in NET2 buffer (NET2 buffer corresponded to NT2 buffer: 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM MgCl2 and 0.05% NP40 added with 1 mM DTT, 20 mM EDTA, 200 U/ml RNAse Out). An aliquot of diluted cell lysate was removed and represented the starting material or 'input' which was processed alongside the immunoprecipitation to compare with immunoprecipitated mRNAs at the end. RIP experiments were performed overnight at 4°C on diluted cell lysate with antibodies to NONO (ab45359, Abcam), SFPQ (ab38148, Abcam), PSPC1 (SAB4200068, Sigma-Aldrich, Saint-Quentin Fallavier, France), RBM14 (ab70636, Abcam) and HNRNPK (sc-28380, Santa-Cruz Biotechnology, Heidelberg, Germany) or non-specific rabbit polyclonal antibody (anti-Furin, sc-20801). After incubation was completed, 15 µl of Magna ChIP protein A magnetic beads (16–661, Millipore, Molsheim, France) were added for 1h at 4°C. Beads were washed 6 times with cold NT2 buffer and treated by proteinase K for 30 min at 55°C. RNA eluted from beads was purified using Nucleospin RNA XS (Macherey-Nagel) and processed for cDNA synthesis using a High Capacity RNA to cDNA kit (Applied Biosystems).

Neat1 RNA pull-down

Neat1 RNA pull-down is a hybridization-based strategy we previously published (10), that uses

complementary oligonucleotides to purify Neat1 RNA together with its targets from reversibly cross-linked extracts. Two antisense DNA oligonucleotide probes that target accessible regions of the lncRNA Neat1 were used for Neat1 RNA specific pull-down (SO1 and SO2) and one biotynylated irrelevant probe (NSO) was used for Neat1 RNA non-specific pull-down (Supplemental Table 1). All three probes were biotinylated at the 3' end.

Briefly, 10 cm cell dishes were incubated in 1 ml of ice-cold cell lysis buffer A as described above. Nuclei were scraped and separated by centrifugation (500 g for 5 min). The nuclear pellets were then fixed with 1% paraformaldehyde in PBS for 10 min at room temperature. Crosslinking was then quenched with 1.25 M glycine for 5 min. Cross-linked nuclei were rinsed two times again with PBS and pelleted at 500 g. Nuclear pellets were stored at -80°C. To prepare lysates, nuclear pellets were suspended in lysis buffer (50 mM Tris, pH 7.0, 10 mM EDTA, 1% SDS added with a protease inhibitor cocktail and RNAse-Out), and sonicated using BioruptorPlus (Diagenode, Seraing, Belgium) by 2 pulses of 30 s allowing complete lysate solubilization. RNA was in the size range of 400 to 2000 nucleotides. Nuclear lysates were diluted V/V in hybridization buffer (750 mM NaCl, 1% SDS, 50 mM Tris, pH 7.0, 1 mM EDTA, 15% formamide). The two specific or the non-specific probes (100 pmol) were added to 1.2 ml of diluted lysate, which was mixed by end-to-end rotation at room temperature 4 to 6 h. Streptavidin-magnetic C1 beads (Dynabeads MyOne Streptavidin C1– Invitrogen Life Technologies) were added to hybridization reaction (50 µl per 100 pmol of probes) and the whole reaction was mixed overnight at room temperature. Beads–biotin-probes–RNA adducts were captured by magnets (Invitrogen) and washed five times with a wash buffer (2 SSC, 0.5% SDS). After the last wash, buffer was removed carefully. For RNA elution, beads were suspended in 100 µl RNA proteinase K buffer (100 mM NaCl, 10 mM Tris, pH 7.0, 1 mM EDTA, 0.5% SDS) and 100 µg proteinase K (Ambion). After incubation at 45°C for 45 min, RNA was isolated using NucleoSpin RNA XS (Macherey-Nagel). Eluted RNA was subject to RT–qPCR for the detection of enriched transcripts.

In vitro transcription/translation

In vitro transcription/translation assays were carried out in a S30 T7 HighYield Protein Expression System (Promega). The full lenght Hrnpk cDNA template was cloned from total RNA obtained from GH4C1 cells and transferred to a pET28a *Escherichia coli* T7 expression vector (Novagen, Nottingham, UK). An empty pET28a DNA template was used as control. Reactions were carried out in a total volume of 50 µl with 0.5µg plasmid DNA.

RNA-Protein pull-down

50 pmol of a specific 3' biotin-TEG oligonucleotide RNA probe (Supplemental Table 1) (IDT,) or a non specific RNA probe (Negative RNA Control [poly$(A)_{25}$ RNA] from the Pierce™ Magnetic RNA-Protein Pull-Down Kit) were incubated for 15-30 minutes at room temperature under agitation with 50µl of Streptavidin Magnetic Beads. 100µl of a Master Mix containing 15% glycerol and nuclear GH4C1 protein extracts or proteins translated in vitro (20-200 µg) were incubated 30-60 minutes at 4°C with agitation. After three washes, protein complexes were eluted from RNA by adding 50µl of Elution Buffer to the beads and incubating 15-30 minutes at 37°C with agitation. Samples were then loaded onto a NuPAGE 4-12% Bis-Tris gel (Invitrogen) for mass spectrometry analysis or a 10% acrylamide Tris-Glycine gel for western blotting.

MS analysis

Samples were allowed to migrate 5 min at 200 V. The gel was then silver stained and cut into small pieces. Before trypsin digestion, the gel pieces were destained, rinsed and then reduced with dithiothreitol 20mM and alkylated with iodoacetamide 10mM in 100mM NH4HCO3.

Samples were then incubated overnight at 37°C with 12.5 ng/µL trypsin (sequencing grade; Promega) in 25 mM NH4HCO3. Gel pieces were then extracted 3 times with 50% acetonitrile 0.1% formic acid and evaporated to dryness using a speed vac.

The dried samples were resuspended in 8 µl of 2% acetonitrile and 0.1% formic acid and analyzed by mass spectrometry using a hybrid Q-Orbitrap mass spectrometer (Q-Exactive, Thermo Fisher Scientific, United States) coupled to a nanoliquid chromatography (LC) Dionex RSLC Ultimate 3000 system (Thermo Fisher Scientific, United States). Peptide separation was performed on an Acclaim PepMap RSLC capillary column (75 µm × 15 cm, nanoViper C18, 2 µm, 100 Å) at a flow rate of 300 nl/min. Solvant A was 0.1% formic acid. The analytical gradient was run with various percentages of solvent B (acetonitrile with 0.1% formic acid) in the following manner: (1) 2.5–25% for 57 min, (2) 25–50% for 6 min, (3) 50-90% for 1 min, and (4) 90% for 10 min. MS spectra were acquired at a resolution of 35,000 within a mass range of 400–1,800 m/z. Fragmentation spectra of the 10 most abundant peaks (Top10 method) were acquired with high-energy collision dissociation (HCD).

Bioinformatics

Mass spectrometry data analysis

All mass spectrometry RAW files were uploaded into MaxQuant version 1.5.1.0 (https://maxquant.net/maxquant/, (16)) and searched against a rat SwissProt protein database (December 2018 release). The following parameters were used for the search: trypsin/P enzyme with up to 3 missed cleavages allowed; carbamidomethylation of cysteine was set to fixed modification; with oxidation of methionine, N-terminal protein acetylation set as variable modifications; first search peptide tolerance was set to 20 ppm against a small 'mouse-first-search' database for the purpose of mass recalibration and main search was performed at 4.5 ppm; contaminants were included in the search; database was reversed for the purpose of calculating the peptide and protein level false-discover rate (FDR) at 1%; Label Free Quantification (LFQ) minimum-ratio count was set to 1.

For statistical comparison a dataset of 3 Specific versus 3 Non-Specific was imported into Perseus software platform version 1.6.2.3 (https://maxquant.net/perseus/, (17)). Contaminants proteins, proteins only identified by site and reverse identifications were filtered out of the dataset. LFQ values were Log2 transformed and samples were filtered with a minimum of 2 for valid values in at least one of the two conditions (Specific and Non-Specific samples).

Results are summarized as a heat Map showing the clustering of the different experiments for the validated proteins.

Nucleotide sequence analysis

This analysis was conducted using the bioconductor Biostrings package (v2.52) of R (v 3.6) software.

- Consensus motif

Local alignment of the 3'UTR-C1 and the 3'UTR-C3 fragments was performed using the pairwiseAlignment function.

- 3' UTR sequences

The rat DNA sequences and the coordinates of the 3'UTR were obtained from the Rattus_norvegicus.Rnor_6.0.fa and Rattus_norvegicus.Rnor_6.0.100.gtf in Ensembl, respectively. For each gene, only the sequence of the longest 3'UTR described, when available, was taken into account, leading to a list of 15693 sequences, which was considered as the background file. From this file, a subset of sequences was extracted, those belonging to the 3928

Neat1 RNA targets we previously established (11), leading to a list of 3533 sequences considered as the specific file.

- Consensus matrix

The specific and the background files were scanned for the TGCCYCCAGGRCTGG motif using the vmatchPattern function, allowing 4 mismatches.

A consensus matrix was generated from the 1753 hits obtained, using the consensusMatrix function.

A logo was drawn from this matrix, using the seqLogo package (v 1.50) from bioconductor.

Western-blot analysis

Eluate samples from RNA-protein pull-down experiments were submitted to Western-blot analysis as previously described (18) with polyclonal primary antibodies raised against HNRNPK (1:1000, sc-28380, Santa Cruz Biotechnology).

RNA-FISH

To detect Neat1 or/and Calr RNA, GH4C1 cells grown on glass coverslips coated with poly-ornithine were fixed in 3.6% formaldehyde. Hybridization was carried out using Custom Stellaris FISH Probes (Biosearch Technologies, Novato, USA). Probes used for dual FISH are Calr probes labeled with Quasar 670 Dye and Neat1 probes labeled with Quasar 570 Dye. Nuclei of the cells were counterstained by Hoechst solution (1µmol/ml).

Cosinor and statistical analysis

Cosinor and statistical analyses were performed using Prism4 software (GraphPad Software, Inc.).

For cosinor analysis, mean experimental values (± SEM), expressed as a percent of initial value, were fitted using Prism4 software (GraphPad Software, Inc.) by a non-linear sine wave equation: Y = Baseline + Amplitude * sin (Frequency*X + Phase-shift), where Frequency= 2pi/period. Goodness-of-fit was quantified using R squared, experimental values being considered well fitted by cosinor regression when the R squared was higher than 0.55. A statistically significant circadian oscillation was considered if the 95% confidence interval for the amplitude did not include the zero value (zero-amplitude test)(19)(20).

One-way ANOVA followed by Dunnett's multiple comparison test was used to test for significant differences between groups. Values were considered significantly different for $p < 0.05$ (*), $p < 0.01$ (**) or $p<0.001$ (****).

RESULTS

1/ Post-transcriptional Calr mRNA circadian expression

Calr mRNA displayed a circadian expression pattern in GH4C1 cells. Experimental mRNA values could indeed be fitted by the following equations: Calr: $y= 111.8 – 21.35*\sin (0.261x + 0.938)$ with a $R^2= 0.688$. As it was previously shown to be the case in the liver (15), this rhythmic pattern was further shown to be generated at a post-transcriptional step since the corresponding pre-mRNA did not display a rhythmic expression pattern as illustrated in Fig. 1A. Indeed, Calr pre-mRNA levels could not be significantly fitted by a sine wave equation with a period of 24h ($R^2<0.55$) (Fig. 1A). As expected for a mRNA paraspeckle target, the control of its circadian expression pattern took place at a post-transcriptional level.

2/ Association of Calr mRNA with paraspeckles in GH4C1 cells

Using a Neat1 RNA pull down procedure, as described in a previous paper (10), Calr mRNA was shown to be significantly enriched relative to input when specific oligonucleotide (SO) probes directed against Neat1 were used as compared to a non specific oligonucleotide (NSO) probe (Fig. 1B). While Calr mRNA enrichment was statistically significant with the two SO used (SO1 and SO2) compared to NSO, showing clearly that Calr mRNA was associated with Neat1, SO2 induced a significantly higher enrichment (Fig 1B); this prompted us to use this oligonucleotide probe in our further experiments. Using RNA-immunoprecipitation (RIP) experiments with antibodies directed against NONO, SFPQ, RBM14 or PSPC1, Calr mRNA was further shown to be associated with the four core protein components of paraspeckles (Fig. 1C). A 3- to 8.5-fold enrichment in Calr mRNA was obtained with these antibodies as compared to an irrelevant non-specific antibody (Fig. 1C). The association between the four major proteins of paraspeckles with Calr mRNA reinforced the view that Calr mRNA bound paraspeckles.

Dual FISH experiments provided additional morphological evidence for Calr mRNA association to paraspeckles visualized by Neat1 RNA staining. As already described, Neat1 RNA staining under confocal laser scanning microscope appeared as regular punctates within the boundaries of the nucleus (Green Fig.1D) (11). By contrast, Calr mRNA staining was mainly localized in the cytoplasm while some punctates also appeared in the nucleus (Red Fig. 1D). Among these nuclear punctates, some merged with Neat1 RNA staining (Fig. 1D arrow) showing that Calr mRNA could be associated with Neat1 RNA. By contrast, some Neat1 RNA stained punctate did not merge with Calr mRNA staining (Fig. 1D arrow head).

All together, these results confirmed that Calr mRNA was strongly associated with paraspeckles in

GH4C1 cells and validated its choice as a gene candidate to search for molecular mechanism of this association.

3/ Involvement of 3'UTR in the binding of Calr mRNA to paraspeckles

Paraspeckles are known to retain in the nucleus RNAs containing duplex structures from inverted repeats of the conserved Alu sequences (IRAlu elements) within their 3'-UTR as shown for Nicolin 1 (NICNI) gene (14). We had previously generated a GH4C1 cell line stably transfected by a construct in which an IRAlu sequence from the 3'-UTR of NICN1 gene was inserted between the EGFP cDNA 3'-UTR region and the SV40 polyadenylation signal of the expression vector pEGFP-C1 and we had previously shown that this IRAlu element could bind paraspeckles and induce Egfp mRNA nuclear retention (11). The same strategy was used to determine whether paraspeckle nuclear bodies could bind Calr mRNA through its 3'UTR. 3'UTR of Calr mRNA was cloned by PCR and inserted in pEGFP-C1 vector and the construct was stably transfected into GH4C1 cells. The efficacy of 3'UTR-Calr mRNA to enrich Egfp mRNA after Neat1 RNA pull-down with SO2 compared to NSO was compared to that of IRAlu-containing Egfp mRNA. As shown in Fig. 2A, no statistical difference in Egfp enrichment after Neat1 RNA pull-down was obtained between cell lines expressing either 3'UTR-Calr- or IRAlu-containing Egfp mRNA (Fig. 2A). Furthermore, there was no significant difference in nuclear/cytoplasmic distribution of 3'UTR-Calr- or IRAlu-containing Egfp mRNA, consistent with the assumption that 3'UTR-Calr sequence retained Egfp mRNA in the nucleus as efficiently as IRAlu sequence did (Fig. 2B). Taken together, these results demonstrated that binding of Calr mRNA to paraspeckle nuclear bodies involved the 3'UTR of Calr mRNA.

4/ Delineation of sub-regions in 3'UTR-Calr mRNA engaged in paraspeckle binding

To further characterize the sub-regions of the 3'UTR of Calr mRNA that are involved in the binding to paraspeckles, three fragments (3'UTR-C1 (166nt), 3'UTR-C2 (210nt), 3'UTR-C3 (184nt)) were generated by PCR from the entire 3'UTR region (Fig. 2C). Each of the three fragments were inserted between the Egfp cDNA 3'-UTR region and the SV40 polyadenylation signal of the expression vector pEGFP-C1 to generate constructs that were then stably transfected into GH4C1 cells. The efficacy of each fragment to enrich Egfp mRNA after Neat1 RNA pull-down with SO2 versus NSO was compared to the efficacy of the entire 3'UTR of Calr mRNA. Egfp enrichment obtained with 3'UTR-C1 and 3'UTR-C3 fragments did not differ from that obtained with the entire 3'UTR-Calr; by contrast Egfp enrichment obtained with 3'UTR-C2 was reduced and significantly lower than that obtained with 3'UTR-C1 (Fig. 2D). All three fragments were less able than the entire 3'UTR of Calr mRNA to retain Egfp mRNA in the nucleus; however while 3'UTR-C1 and 3'UTR-C3 induced a same retention of

about 30% compared to the entire 3'UTR, 3'UTR-C2 was significantly less efficient, inducing a 20% nuclear retention compared to that obtained with the entire 3'UTR (Fig. 2E). It then appeared that while 3'UTR-C1 and 3'UTR-C3 were as efficient as the entire 3'UTR-Calr to induce the binding of Egfp mRNA to paraspeckles, they were not as efficient as the entire 3'UTR-Calr to retain Egfp mRNA in the nucleus. In addition, 3'UTR-C2 was less efficient both to induce its binding to paraspeckles and to retain Egfp mRNA in the nucleus.

5/ Involvement of a tandem sequence of 15 nucleotides in 3'UTR-Calr mRNA in the binding to paraspeckles

Through alignment of the oligonucleotide sequences of the fragments C1 and C3 using the Biostring pairwiesAlignment function (Smith-Waterman local alignment), a common sequence of 15 bases was identified in these two fragments, with a T or C (Y) in 5th and a G or A (R) in 11th position (Fig. 2F). The 3'UTR of Calr mRNA exhibited then a tandem of this 15 nucleotides sequence (Supplemental Fig 2). Mutations were introduced in the 15 nucleotides sequence present in C1 (3'UTR-$C1_M$) , or in C3 (3'UTR-$C3_M$) or in both (3'UTR-$C1_M$/$C3_M$) by replacement of every odd base by its complementary inverse base. Enrichment in Egfp mRNA after Neat1 RNA pull-down with SO2 versus NSO obtained after these mutations was compared to that obtained with the native 3'UTR-Calr mRNA. Mutation in the sequence present in C3 significantly decreased Egfp mRNA enrichment (Fig. 2G); by contrast, while the mutation in the sequence present in C1 did not by itself reach a statistically significance in reducing Egfp mRNA enrichment, the dual mutation of the two sequences amplified the reduction in Egfp mRNA enrichment induced by C3 mutation (Fig. 2G). Mutation in the sequence present in C1 or in C3 significantly reduced the relative ratio of nuclear versus cytoplasmic Egfp mRNA distribution when compared to the native 3'UTR-Calr (Fig. 2H). Mutation in C3 was however more efficient than mutation in C1. However dual mutation of the two sequences did not display additive or synergic effect (Fig. 2H).

6/ Rate of occurrence of the 15 nucleotides sequence in the 3'UTR of the Neat1 mRNA targets

The prevalence of the 15 nucleotides sequence identified in the 3'UTR of the Calr mRNA was determined in the 3'UTR of the 3928 Neat1 RNA targets we previously established (11). Since a same mRNA can display different length of 3'UTR, the list of the Neat1 RNA targets was filtered to retain for each mRNA only the longest 3'UTR referenced. This led to a list of 3533 3'UTR. For this analysis four mismatches were allowed in the 15 nucleotides sequence. In the list of 3533 3'UTR, we found 1753 times this specific sequence included in 1093 mRNA (Table 2). It then appeared that near 30% of

the Neat1 RNA targets have at least one identified 15 nucleotides sequence in their 3'UTR, with a mean of 1.6 sequences per gene. However when considering the background defined as total rat RNAs, the 15693 3'UTR retained for the analysis (after filtering for the longest 3'UTR of a single mRNA), 7440 sequences were identified and found to correspond to 4678 mRNA; near 30% of total RNAs displayed then at least one identified 15 nucleotides sequence in their 3'UTR, as Neat1 RNA targets did (Table 2). Accordingly, while the 15 nucleotides sequence we described in the present paper could allow a mRNA to bind to paraspeckles, this sequence was not enriched in the 3'UTR of mRNA that are paraspeckle's target, as it was also the case for IRAlu sequences.

From the 1753 sequences found in 3'UTR of Neat1 mRNA targets, a table of consensus matrix probabilities was built (Supplemental Table 2), allowing the design of a sequence logo (https://rdrr.io/bioc/seqLogo/) (Fig. 2I).

7/ Analysis by mass spectrometry of proteins bound to the 15 nucleotides sequence

To identify proteins able to bind the 15 nucleotides sequence, RNA-protein pull-down was performed using nuclear protein extracts from GH4C1 cells and a biotinylated 30 nucleotides RNA probe corresponding to the 15 nucleotides sequence present in C3 in the 3'UTR-Calr mRNA surrounded in 5' by 7 nucleotides and in 3' by 8 nucleotides (Supplemental Table 1). Three replicates using the specific probe and three replicates using a non-specific probe (Negative RNA Control [poly$(A)_{25}$ RNA] from the PierceTM Magnetic RNA-Protein Pull-Down Kit) were performed. Label free mass spectrometry analysis gave a list of 45 proteins that were identified at least in two biological replicates using the same probe (Specific or Non-Specific) (Supplemental Fig.1). Finally, 24 from this pool of 45 proteins displayed higher expression in Specific as compared to Non-Specific samples. These proteins were thus considered as specifically bound to the RNA probe containing the 15 nucleotides sequence (Table 1). String analysis (21) of these proteins showed that 21 of them were RNA-binding proteins (Fig. 3A, red color) and that a dense network connected 21 of each to each other (Fig. 3A). HNRNPK was the only protein in our list to have been previously described as a paraspeckle protein component (9).

8/ Evidence for the direct binding of the protein HRNPK to the 15 nucleotides RNA sequence

In order to determine whether HRNPK could connect the 15 nucleotides sequence identified here to paraspeckles, we tested whether HNRNPK can bind directly to this sequence. Rattus HNRNPK was then transcribed *in vitro* and incubated with the specific 30 nucleotides RNA probe containing the 15 nucleotides sequence identified in 3'UTR-Calr mRNA. Used as a control, an empty pET28a DNA lysate was also incubated with the specific RNA probe. After elution, samples were processed for

western blot with an antibody directed against HRNPK. Western blot analysis of nuclear protein lysates from GH4C1 cells showed that HNRNPK corresponded to a prime band of 65 kDa (Fig. 3B Left panel). As shown in Fig. 3B right panel, a prime band of 65 kDa was also visible in elution fractions from samples where HNRNPK had been transcribed from Hnrnpk-pET28a (lanes 2 and 3) in contrast to samples transcribed with an empty pET28a (lanes 4 and 5).

DISCUSSION

Although the physiological functions of paraspeckles remain only partly understood, two known functions of paraspeckles are the sequestration of specific transcription factors and/or RNA-binding proteins and the regulation of the expression of specific transcripts via their retention in the nucleus (22). Thus, Neat1 can regulate expression of genes transcriptionally by sequestering proteins involved in the regulation of gene promoters, and co-transcriptionally by binding to mRNAs. While some studies have been undertaken to determine the sequence elements in Neat1 responsible for the recruitment of proteins (23), little is known about potential specific elements in paraspeckle mRNA targets except that several mRNA targets contain inverted repeats derived from SINE repeats in their 3' UTRs (7)(3) an element we didn't find in 3'UTRs of paraspeckle mRNA targets in rat GH4C1 cells. Two main strategies can be used to identify recurring sequence patterns or motifs, namely a global computational motif discovery approach or a gene candidate strategy. The first one adopts either traditional and popular methods such as MEME (Multiple Expression motifs for Motif Elicitation) (24) or more recent one such as Amadeus (A Motif Algorithm for Detecting Enrichment in mUltiple Species) (25) to discover the sequence consensus of RNA Binding Protein's binding sites or used Z-score statistics to search for the overrepresented nucleotides of a certain size. In the case of our issue, the main practical limitations of these genone-scale detections of known and/or novel motifs lies in the fact that the known or novel motif has to be overrepresented to be identified and the size of the searched sequence has to be defined. However, a nucleotide sequence may be involved in paraspeckle binding without being overrepresented in paraspeckle mRNA targets, as shown here to be the case for IRAlu sequences and the size of the potential sequence cannot be presupposed a priori. For all these reasons, a gene candidate strategy appeared more suitable to reach our goal.

Among the 3928 paraspeckle RNA targets we reported in GH4C1 cells (11), our choice of gene candidate fell on Calr mRNA, which is shown here to be closely associated not only with Neat1 but also with every major core-protein component of paraspeckles. Furthermore since a circadian post-transcriptional regulation is shown here to account for its circadian expression pattern as previously shown in the liver (15) and since its circadian nuclear retention is dependent on Neat1, as previously

shown (11) it may be assumed that paraspeckle binding is crucial for Calr mRNA expression and circadian pattern (11). Finally, the 3'UTR of Calr mRNA appears as efficient as an IRAlu sequence to bind Neat1 and to induce a nuclear retention of a reporter gene, leading to the conclusion that the dissection of 3'UTR Calr mRNA may allow to uncover sequence elements involved in paraspeckle binding.

The most proximal and the most distal regions of the 3'UTR Calr mRNA are then identified as two regions able to associate with Neat1. However in spite of their efficient association with paraspeckles, these two regions are far less efficient than the entire 3'UTR to induce a nuclear retention of the reporter gene. It then seems that each of these regions can bind to paraspeckles with a binding strength that is insufficient to allow for their nuclear retention. In addition to show that Calr mRNA needs to be anchored by two sites to be efficiently retained in the nucleus by paraspeckles, more generally, these results suggest that cooperative effects between different parts of 3'UTR may be necessary for an efficient nuclear retention of the mRNA. This assumption is further confirmed by mutation experiments, in which dual mutations in the two parts of 3'UTR Calr reinforces the decrease in paraspeckle binding and reporter gene nuclear retention induced by a single mutation. It is tempting to speculate that the cooperative effects between different parts of the 3'UTR are sustained by secondary structures carried by mRNA and as previously suggested are determinant for nuclear retention (26). In any case, the efficiency of the two regions of 3'UTR Calr to associate with paraspeckles prompted us to further analyze their molecular characteristics.

Using RNA-RNA interaction prediction algorithms (IntaRNA, http://rna.informatik.uni-freiburg.de/) (27) we were able to ruled out that the two identified regions of 3'UTR Calr can directly hybridize with Neat1. Indeed, the only part of 3'UTR Calr that offered a significant prediction score for Neat1 interaction is located between the two regions of interest (see Supplemental Fig 2). While devoid of Neat1 direct interaction sequence, the two regions were shown to display after alignment a common sequence of 15 nucleotides with Y in 5th and R in 11th position. Interestingly 30% of paraspeckle mRNA targets arbor a sequence close to this one in their 3'UTR, with an average of 1.6 sequences per 3'UTR. However, the occurrence of this sequence in 3'UTR is not overrepresented in mRNA that are paraspeckle targets as compared to all expressed mRNA in the cells. As indicated for IRAlu sequences, it then appears that while the 15 nucleotides sequence identified here allows the 3'UTR of mRNAs to associate with paraspeckles, its occurrence by itself in 3'UTR of mRNAs does not make them paraspeckle targets.

A mass spectrometry analysis allows the identification of 24 proteins specifically bound to an oligonucleotide containing the 15 nucleotides sequence. As expected, a majority of them are RNA binding proteins that form a dense plexus of interactions between each other. Among them, only one namely HNRNPK, has been previously described as paraspeckle protein component (22). Focusing on this protein, it was found that HNRNPK is able to directly bind the 15 nucleotides sequence identified. Interestingly, the motif enriched in the HNRNPK binding peaks throughout the transcriptome, as identified by GraphProt (28) has been shown to correspond to a pyrimidine-rich sequence (26), consistent with the known preferences of HNRNPK for C-rich sequences. This appears in accordance with the 15 nucleotides sequence we identified here in which more than half of the nucleotides are pyrimidines (29). Moreover, the three KH RNA-binding domains of HNRNPK were previously shown to act cooperatively in binding sequences with triplets of C/T-rich regions (30) fitting the sequence architecture of the entire 3'UTR Calr mRNA. It has also been suggested that the C/T-rich motif is preferentially bound in a structured context consistent with our observation that a single sequence was less efficient than dual repeat of the sequence.
Finally, HNRNPK has also been shown to bind a 42 nucleotides fragment that contained three stretches of at least six pyrimidines (C/T) which has been named SIRLOIN (SINE-derived nuclear RNA LOcalizatIoN) (26). It has been shown that HNRNPK binding to this 42 nucleotides fragment can direct nuclear enrichment of mainly lncRNAs but also some mRNAs. As suggested by these authors and in view of our present results, multiple independent pathways are likely responsible for nuclear enrichment in lncRNAs and mRNAs, recognizing specific RNA sequences and/or other features of the ribonucleoproteins.

The present results allow from the 1753 sequences found in 3'UTR of paraspeckle mRNA targets to design a motif, revealing a novel matrix in 3'UTR of mRNAs that allows them through association with paraspeckles to be retained in the nucleus. Furthermore, these results present evidence that the specific recognition of this sequence by HNRNPK is important to bridge paraspeckles with their mRNA targets. HNRNPK appears then not only essential for paraspeckle formation (9) but also for nuclear retention of paraspeckle mRNA target. However the sequence we described here is widely present in 3'UTR of mRNAs whether they are or they aren't paraspeckle targets and it thus may be assumed that other specific RNA sequences and/or other associated proteins are involved in mRNA paraspeckle nuclear retention. Nevertheless, based on these findings, the novel matrix proposed can be presented as a promising candidate for the elusive recruiting elements of mRNAs by paraspeckles and lay the groundwork for future studies to interrogate the role of cooperative other sequences in paraspeckle mRNA nuclear retention.

SUPPLEMENTARY DATA

Supplementary Data are available online.

FUNDING

This work was supported by Aix-Marseille University and Centre National Recherche Scientifique and funded by a grant from Sandoz Laboratories.
The mass spectrometer was obtained using financial support of the "Fédération de Recherche pour le Cerveau" (FRC) through the Rotary operation "Espoir en tête."
Funding for open access charge: Centre National Recherche Scientifique

CONFLICT OF INTEREST

The authors declare that no competing interests exist.

REFERENCES

1. Fox,A.H., Lam,Y.W., Leung,A.K., Lyon,C.E., Andersen,J., Mann,M. and Lamond,A.I. (2002) Paraspeckles: a novel nuclear domain. *Current Biology*, **12**, 13–25.

2. Bond,C.S. and Fox,A.H. (2009) Paraspeckles: nuclear bodies built on long noncoding RNA. *The Journal of Cell Biology*, **186**, 637–644.

3. Chen,L.-L. and Carmichael,G.G. (2009) Altered Nuclear Retention of mRNAs Containing Inverted Repeats in Human Embryonic Stem Cells: Functional Role of a Nuclear Noncoding RNA. *Molecular Cell*, **35**, 467–478.

4. Sasaki,Y.T. and Hirose,T. (2009) How to build a paraspeckle. *Genome biology*, **10**, 1–5.

5. Guru,S.C., Agarwal,S.K., Manickam,P., Olufemi,S.-E., Crabtree,J.S., Weisemann,J.M., Kester,M.B., Kim,Y.S., Wang,Y., Emmert-Buck,M.R., *et al.* (1997) A transcript map for the 2.8-Mb region containing the multiple endocrine neoplasia type 1 locus. *Genome research*, **7**, 725–735.

6. Naganuma,T. and Hirose,T. (2013) Paraspeckle formation during the biogenesis of long non-coding RNAs. *RNA Biology*, **10**, 456–461.

7. Prasanth,K.V., Prasanth,S.G., Xuan,Z., Hearn,S., Freier,S.M., Bennett,C.F., Zhang,M.Q. and Spector,D.L. (2005) Regulating Gene Expression through RNA Nuclear Retention. *Cell*, **123**, 249–263.

8. Nakagawa,S. and Hirose,T. (2012) Paraspeckle nuclear bodies—useful uselessness? *Cellular and Molecular Life Sciences*, **69**, 3027–3036.

9. Naganuma,T., Nakagawa,S., Tanigawa,A., Sasaki,Y.F., Goshima,N. and Hirose,T. (2012) Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles: LncRNA processing for nuclear body architecture. *The EMBO Journal*, **31**, 4020–4034.

10. Torres,M., Becquet,D., Guillen,S., Boyer, Bénédicte,B., Moreno,M., Blanchard,M.-P., Franc,J.-L. and François-Bellan,A.-M. (2018) RNA Pull-down Procedure to Identify RNA Targets of a Long Non-coding RNA. *JoVE*, 10.3791/57379.

11. Torres,M., Becquet,D., Blanchard,M.-P., Guillen,S., Boyer,B., Moreno,M., Franc,J.-L. and François-Bellan,A.-M. (2016) Circadian RNA expression elicited by 3'-UTR IRAlu-paraspeckle associated elements. *eLife*, **5**, e14837.

12. Torres,M., Becquet,D., Blanchard,M.-P., Guillen,S., Moreno,M., Franc,J.-L., François-Bellan,A.-M. and François-Bellan,A.-M. (2017) Paraspeckles as rhythmic nuclear mRNA anchorages responsible for circadian gene expression. *Nucleus*, **8**, 249–254.

13. Chen,L.-L., DeCerbo,J.N. and Carmichael,G.G. (2008) Alu element-mediated gene silencing. *The EMBO Journal*, **27**, 1694–1705.

14. Chen,L.-L. and Carmichael,G.G. (2008) Gene regulation by SINES and inosines: biological consequences of A-to-I editing of Alu element inverted repeats. *Cell Cycle*, **7**, 3294–3301.

15. Menet,J.S., Rodriguez,J., Abruzzi,K.C. and Rosbash,M. (2012) Nascent-Seq reveals novel features

of mouse circadian transcriptional regulation. *Elife*, **1**, e00011.

16. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.

17. Tyanova,S., Temu,T., Sinitcyn,P., Carlson,A., Hein,M.Y., Geiger,T., Mann,M. and Cox,J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*, **13**, 731–740.

18. Guillaumond,F., Boyer,B., Becquet,D., Guillen,S., Kuhn,L., Garin,J., Belghazi,M., Bosler,O., Franc,J.-L. and Francois-Bellan,A.-M. (2011) Chromatin remodeling as a mechanism for circadian prolactin transcription: rhythmic NONO and SFPQ recruitment to HLTF. *The FASEB Journal*, **25**, 2740–2756.

19. Yang,X., Zhang,Y.-K.J., Esterly,N., Klaassen,C.D. and Wan,Y.-J.Y. (2009) Gender Disparity of Hepatic Lipid Homoeostasis Regulated by the Circadian Clock. *Journal of Biochemistry*, **145**, 609–623.

20. Kavčič,P., Rojc,B., Dolenc-Grošelj,L., Claustrat,B., Fujs,K. and Poljak,M. (2011) The impact of sleep deprivation and nighttime light exposure on clock gene expression in humans. *Croatian Medical Journal*, **52**, 594–603.

21. Szklarczyk,D., Franceschini,A., Wyder,S., Forslund,K., Heller,D., Huerta-Cepas,J., Simonovic,M., Roth,A., Santos,A., Tsafou,K.P., *et al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, **43**, D447–D452.

22. Yamazaki,T. and Hirose,T. (2015) The building process of the functional paraspeckle with long non-coding RNAs. *Front Biosci (Elite Ed)*, **7**, 1–41.

23. Simko,E.A.J., Liu,H., Zhang,T., Velasquez,A., Teli,S., Haeusler,A.R. and Wang,J. (2020) G-quadruplexes offer a conserved structural motif for NONO recruitment to NEAT1 architectural lncRNA. *Nucleic Acids Research*, 10.1093/nar/gkaa475.

24. Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, **2**, 28–36.

25. Linhart,C., Halperin,Y. and Shamir,R. (2008) Transcription factor and microRNA motif discovery: The Amadeus platform and a compendium of metazoan target sets. *Genome Research*, **18**, 1180–1189.

26. Lubelsky,Y. and Ulitsky,I. (2018) Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature*, **555**, 107–111.

27. Mann,M., Wright,P.R. and Backofen,R. (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Research*, **45**, W435–W439.

28. Maticzka,D., Lange,S.J., Costa,F. and Backofen,R. (2014) GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biol.*, **15**, R17.

29. Choi,H.S., Hwang,C.K., Song,K.Y., Law,P.-Y., Wei,L.-N. and Loh,H.H. (2009) Poly(C)-binding proteins as transcriptional regulators of gene expression. *Biochemical and Biophysical*

Research Communications, **380**, 431–436.

30. Paziewska,A., Wyrwicz,L.S., Bujnicki,J.M., Bomsztyk,K. and Ostrowski,J. (2004) Cooperative binding of the hnRNP K three KH domains to mRNA targets. *FEBS Lett.*, **577**, 134–140.

TABLE AND FIGURES LEGENDS

Table 1: Mass spectrometry identification of 24 proteins specifically bound to an oligonucleotide containing the 15 nucleotides sequence.

Table 2: Prevalence of the 15 nucleotides sequence in the 3'UTR of Neat1 RNA targets and in the 3'UTR of all rat RNAs

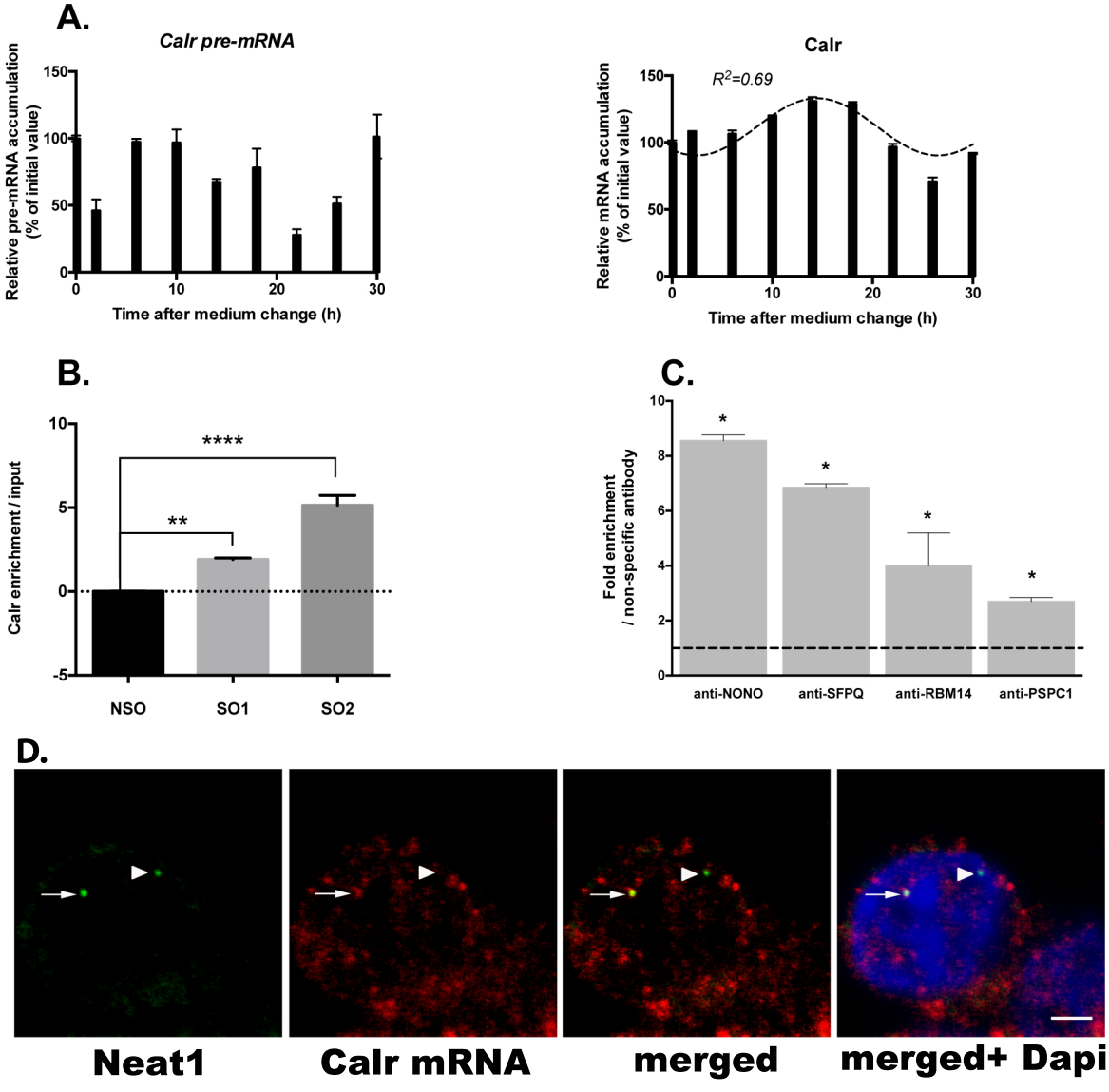
Fig. 1: Post-transcriptional control of Calr mRNA circadian expression and association of Calr mRNA with paraspeckle components
A. **Rhythmic expression of Calr mRNA.** The expression of Calr mRNA is determined by RT-qPCR over a 30 h time period. Experimental values expressed as a percent of the initial value obtained at ZT 0 can be adequately fitted (R^2>0.55) with a non-linear sine wave equation in which the period value is set to 24 h. **Expression pattern of the corresponding pre-mRNA** Experimental values expressed as a percent of the initial value obtained at ZT 0 can not be adequately fitted (R^2<0.55) with a non-linear sine wave equation in which the period value is set to 24 h. **B. Association of Calr mRNA with Neat1** Enrichment in Calr mRNA relative to input after Neat1 RNA pull-down with two different specific biotinylated oligonucleotides, Specific Oligonucleotide 1 (SO1) and Specific Oligonucleotide 2 (SO2) and a non-specific oligonucleotide (NSO); **p<0.01 ****p<0.001. **C. Association of Calr mRNA with four core paraspeckle proteins** Enrichment in Calr mRNA after RNA Immuno-Precipitation (RIP) with antibodies directed against NONO, SFPQ, RBM14 and PSPC1 relative to an irrelevant antibody; *p<0.05 **D. Visualization of Calr mRNA localizing to paraspeckles** Dual Calr and Neat1 RNA-FISH to pituitary GH4C1 cells show **A.** the nuclear distribution of Neat1 RNA in a few distinct foci (arrow and arrow head). **B.** the cytoplasmic and diffuse nuclear localization of Calr mRNA and its distribution in some nuclear distinct foci. **C.** foci in which Calr and Neat1 RNA overlap (arrow) indicating the paraspeckle localization of Calr mRNA and foci without overlap (arrow head). **D.** Nuclear staining by Hoechst is added to C. Scale bars equal 5 mm.

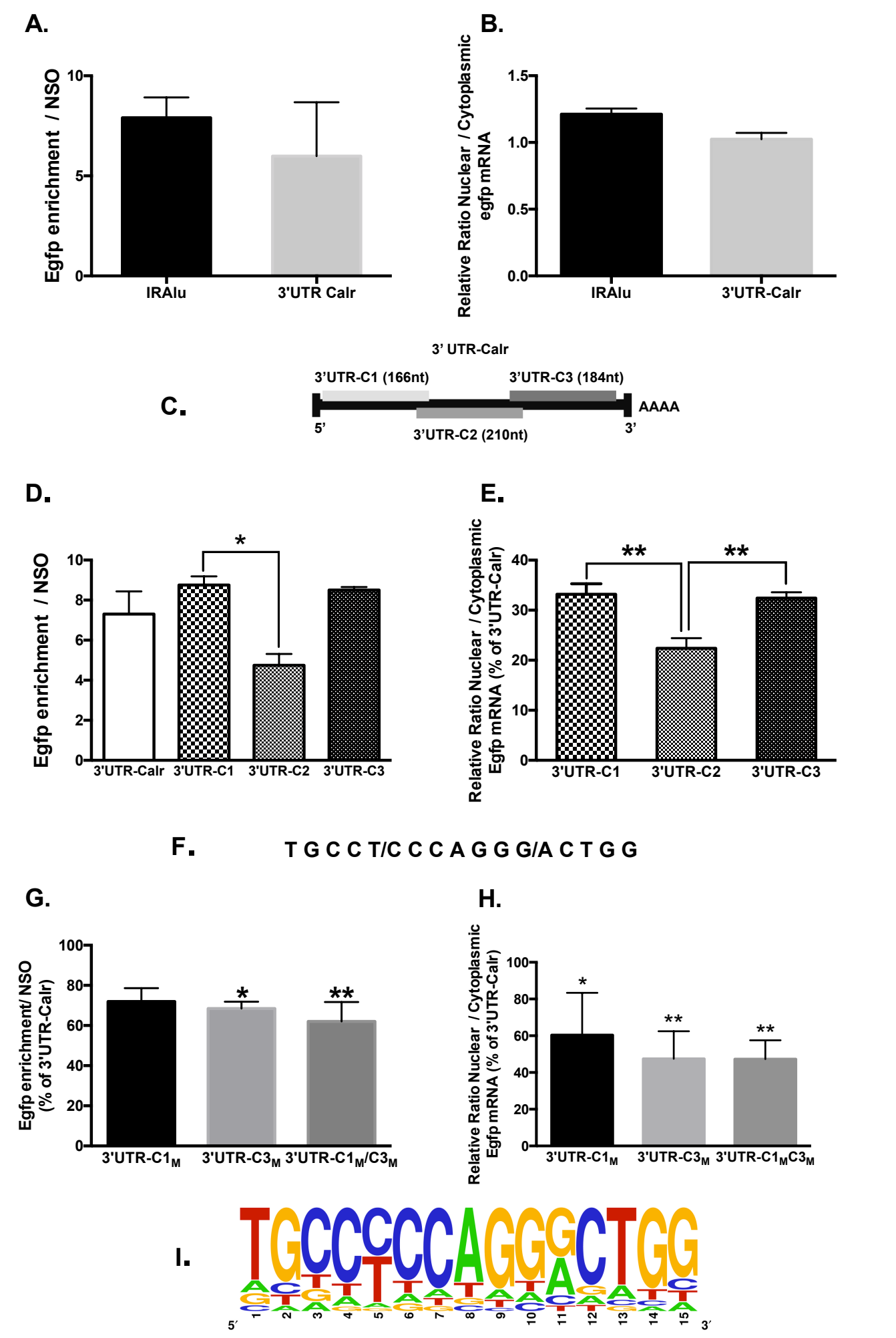
Fig. 2: A 15 nucleotides sequence in 3'UTR involved in the nuclear retention of Calr mRNA by paraspeckles: A-B: Involvement of 3'UTR in Calr mRNA association with paraspeckles
A. Enrichment in Egfp mRNA after Neat1 RNA pull-down with Specific biotinylated Oligonucleotide 2 relative to a non-specific oligonucleotide. The relative enrichment in egfp mRNA obtained after RNA pull-down (n=6 for each cell line) is not statistically different in IRAlu-egfp versus 3'UTR-Calr--egfp cell lines. **B.** Nuclear and cytoplasmic Egfp mRNA were quantified by qPCR in IRAlu-egfp and 3'UTR-Calr-egfp cell lines and normalized to the relative amount of gapdh mRNA (n=3 for each cell line). Ratio of nuclear versus cytoplasmic Egfp mRNA levels is compared between IRAlu-egfp and 3'UTR-Calr-egfp cell lines. **C-E: Subregions of 3'UTR-Calr mRNA associated with paraspeckles**
C. Schematic representation of the three fragments generated from 3'UTR Calr mRNA. 3'UTR Calr mRNA was fragmented in three parts of 166 (3'UTR-C1), 210 (3'UTR-C2) or 184 (3'UTR-C3) nucleotides, respectively. **D.** Enrichment in Egfp mRNA after Neat1 RNA pull-down with Specific biotinylated Oligonucleotide 2 relative to a non-specific oligonucleotide (NSO). The relative enrichment in egfp mRNA obtained after RNA pull-down (n=3 for each cell line) is statistically different in 3'UTR-C1-egfp versus 3'UTR-C2-egfp cell lines. **E.** Nuclear and cytoplasmic Egfp mRNA were quantified by qPCR and normalized to the relative amount of gapdh mRNA (n=3 for each cell line). Ratio of nuclear versus cytoplasmic Egfp mRNA levels obtained in 3'UTR-C1-egfp, 3'UTR-C2-egfp and 3'UTR-C3-egfp cell lines are expressed as a percent of the ratio obtained in 3'UTR-Calr-egfp cell line; *p<0.05 **p<0.01. **F-I: Identification of a 15 nucleotides sequence involved in paraspeckle association F.** Common sequence of 15 nucleotides found in 3'UTR-C1 and 3'UTR-C3 fragments of the 3'UTR-Calr. **G-H:** Every odd base was replaced by its complementary

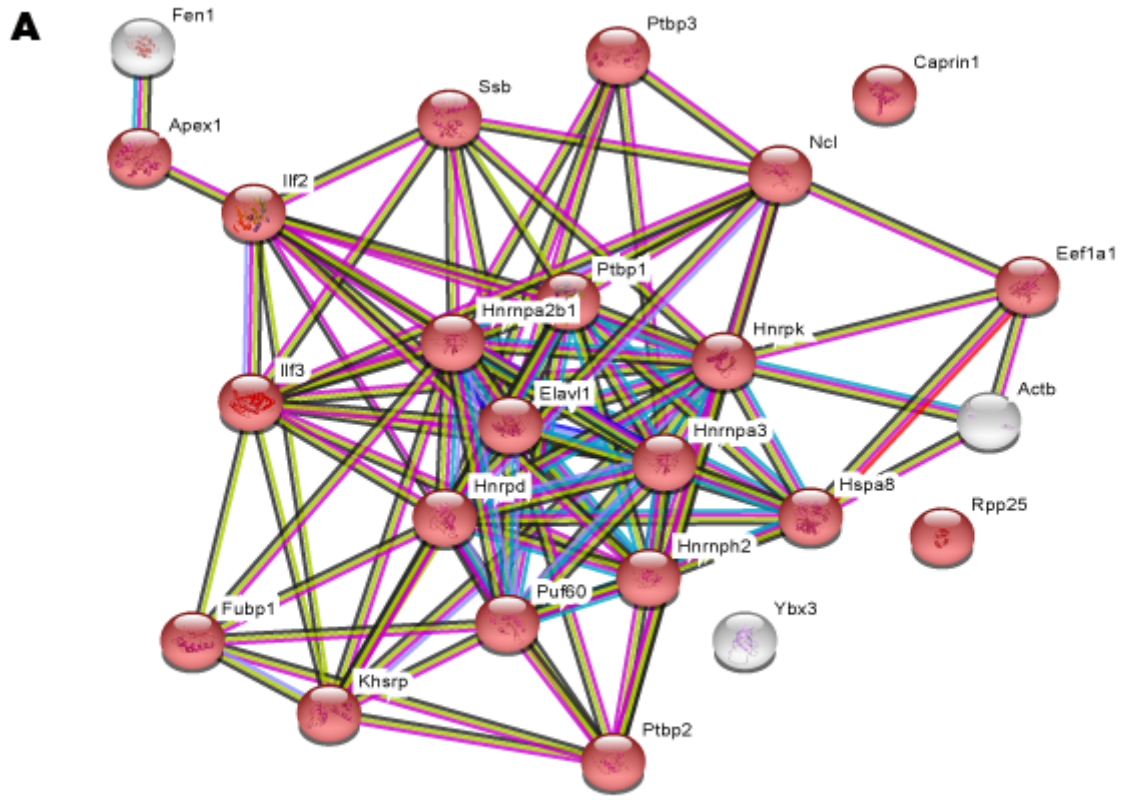
inverse base in the 15 nucleotides sequence present in position 37-51 (3'UTR-$C1_M$), in position 466-480 (3'UTR-$C3_M$) or in both position (3'UTR-$C1_M$/$C3_M$). **G.** Enrichment in Egfp mRNA after Neat1 RNA pull-down with Specific biotinylated Oligonucleotide 2 relative to a non-specific oligonucleotide (NSO) obtained in 3'UTR-$C1_M$-egfp, 3'UTR-$C3_M$-egfp and in 3'UTR-$C1_M$/$C3_M$-egfp cell lines is compared to that obtained in 3'UTR-Calr-egfp cell line. **H.** Ratio of nuclear versus cytoplasmic Egfp mRNA levels obtained in 3'UTR-$C1_M$-egfp, 3'UTR-$C3_M$-egfp and in 3'UTR-$C1_M$/$C3_M$-egfp cell lines are expressed as a percent of the ratio obtained in 3'UTR-Calr-egfp cell line. *$p<0.05$ **$p<0.01$ **I.** Sequence logo that can be designed from the 1753 sequences found in 3'UTR of mRNA paraspeckle targets**.**

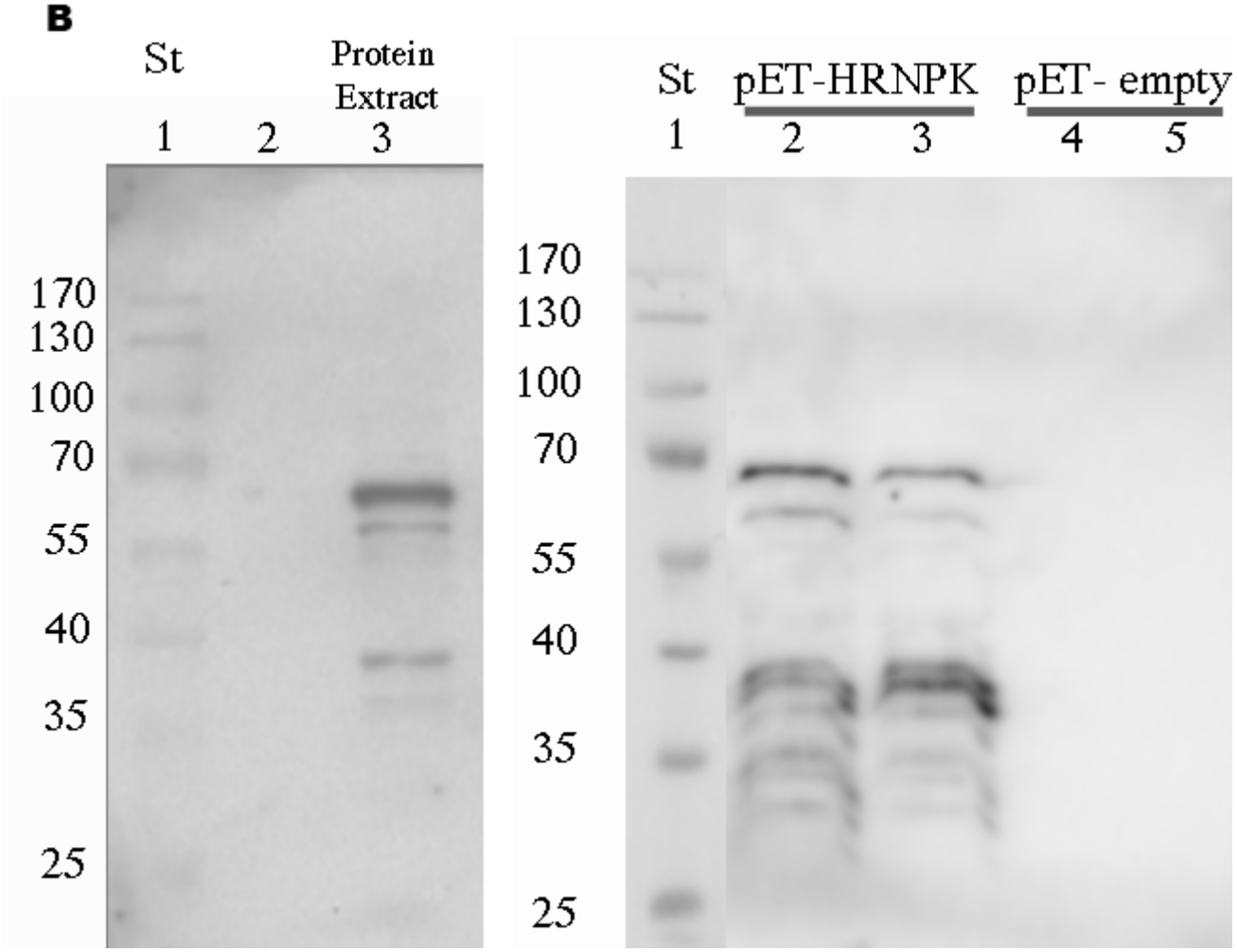
Fig. 3: Analysis of proteins associated with the 15 nucleotides sequence identified in 3'UTR-Calr mRNA: Direct binding of HNRNK

A. String analysis of the 24 proteins identified by mass spectrometry in eluates from binding to biotinylated RNA containing the 15 nucleotides sequence. Red color: RNA binding protein. **B. Left Panel:** Western blot analysis of HNRNPK in GH4C1 nuclear protein extracts (Lane 3). Lane 1: molecular weight. **Right Panel:** Western blot analysis of the binding of *in vitro* transcribed HNRNPK. Lane 2 and 3: identification of HNRNPK *in vitro* transcribed after RNA-protein pull down using biotinylated RNA specific oligonucleotide containing the 15 nucleotides sequence. Lanes 4 and 5: absence of immunolabeling in the presence of control pET28a lysate. Lane 1: molecular weight


A.
Calr pre-mRNA
Relative pre-mRNA accumulation (% of initial value)
150
100
50
0
0
10
20
30
Time after medium change (h)
Calr
R^2=0.69
Relative mRNA accumulation (% of initial value)
150
100
50
0
0
10
20
30
Time after medium change (h)
B.
Calr enrichment / input
10
5
0
-5
****
**
NSO
SO1
SO2
C.
Fold enrichment / non-specific antibody
10
8
6
4
2
0
*
*
*
*
anti-NONO
anti-SFPQ
anti-RBM14
anti-PSPC1
D.
Neat1
Calr mRNA
merged
merged+ Dapi


A.
Egfp enrichment / NSO
10
5
0
IRAlu
3'UTR Calr
B.
Relative Ratio Nuclear / Cytoplasmic egfp mRNA
1.5
1.0
0.5
0.0
IRAlu
3'UTR-Calr
3' UTR-Calr
C.
3'UTR-C1 (166nt)
3'UTR-C3 (184nt)
AAAA
5'
3'
3'UTR-C2 (210nt)
D.
Egfp enrichment / NSO
*
10
8
6
4
2
0
3'UTR-Calr
3'UTR-C1
3'UTR-C2
3'UTR-C3
E.
Relative Ratio Nuclear / Cytoplasmic Egfp mRNA (% of 3'UTR-Calr)
**
**
40
30
20
10
0
3'UTR-C1
3'UTR-C2
3'UTR-C3
F.
T G C C T/C C C A G G G/A C T G G
G.
Egfp enrichment/ NSO (% of 3'UTR-Calr)
100
80
60
40
20
0
*
**
3'UTR-C1M
3'UTR-C3M
3'UTR-C1M/C3M
H.
Relative Ratio Nuclear / Cytoplasmic Egfp mRNA (% of 3'UTR-Calr)
*
**
**
100
80
60
40
20
0
3'UTR-C1M
3'UTR-C3M
3'UTR-C1MC3M
I.
5′
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
3′


A
Fen1
Apex1
Ssb
Ptbp3
Caprin1
Ncl
Eef1a1
Ilf2
Ptbp1
Hnrnpa2b1
Hnrpk
Ilf3
Elavl1
Actb
Hnrnpa3
Hnrpd
Hspa8
Rpp25
Hnrnph2
Puf60
Fubp1
Ybx3
Khsrp
Ptbp2


B
St
Protein Extract
1
2
3
170
130
100
70
55
40
35
25
St
pET-HRNPK
pET- empty
1
2
3
4
5
170
130
100
70
55
40
35
25

Gene names	Peptides	Sequence coverage [%]	Score
Hnrnpa2b1	8	22,9	191,04
Elavl1	3	10,7	33,948
Ncl	9	13,7	74,054
Ssb	2	5,3	24,37
Apex1	5	19,6	38,015
Actb	4	14,4	33,149
Hnrnpk	8	21,8	163,5
Eef1a1	4	8	25,485
Hspa8	11	18,9	183,02
Ptbp1	11	25,8	323,31
Fubp1	7	13,5	63,543
Caprin1	4	7,6	69,329
Rpp25	3	19,1	22,532
Fen1	2	7,1	12,445
Ybx3	14	57,3	323,31
Ptbp2	7	16,9	86,032
Hnrnph2	2	5,8	51,4
Hnrnpa3	2	10	31,933
Ilf2	4	10,2	28,089
Khsrp	4	7,9	38,154
Ilf3	6	7,2	45,276
Puf60	9	19,9	104,4
Ptbp3	3	8,2	17,253
Hnrnpd	12	37.4	215.08

	Number of 3'UTR	Number of sequences	Number of genes	Number of sequences/genes
Neat1 RNA targets	3533	1753	1093	1.6
Background (total RNAs)	15693	7440	4678	1.59