RAN translation of the expanded CAG repeats in the SCA3 disease context

Magdalena Jazurek-Ciesiolka^{1,*}, Adam Ciesiolka¹, Alicja A. Komur^{1,#}, Martyna O. Urbanek-Trzeciak^{1,#} and Agnieszka Fiszer^{1,*}

¹ Department of Medical Biotechnology, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

* To whom correspondence should be addressed. Tel: +48 6185203; Ext. 135; Fax: +48 618520532; Email: agnieszka.fiszer@ibch.poznan.pl; ORCID: 0000-0001-7829-1926

Correspondence may also be addressed to Magdalena Jazurek-Ciesiolka. Tel: +48 6185203; Ext. 135; Fax: +48 618520532; Email: mjazurek@ibch.poznan.pl; ORCID: 0000-0001-7403-7218

#Present Address: Alicja A. Komur, Department of Integrative Biology, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland. Martyna O. Urbanek-Trzeciak, Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

ACKNOWLEDGEMENTS

This work is dedicated to the memory of Professor Włodzimierz J. Krzyzosiak. We thank Professor Marta Olejniczak from Department of Genome Engineering for reviewing the manuscript and every day stimulating discussions. We thank Professor Jerzy Ciesiolka from Department of RNA Biochemistry for helpful suggestions. The confocal microscopy analyses were performed in the Laboratory of Subcellular Structures Analysis, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan. Proteomic mass spectrometry analyses were performed in the Mass Spectrometry Laboratory, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw. This work was supported by the National Science Centre, Poland [2015/19/D/NZ5/02183 to M.J.-C., 2012/06/A/NZ1/00094 to W.J.K., and 2015/19/B/NZ2/02453 to A.C.]; the Polish Ministry of Science and Higher Education (statutory funds for young investigators to M.J.-C.); and the Polish Ministry of Science and Higher Education under the KNOW program [01/KNOW2/2014].

ABSTRACT

Spinocerebellar ataxia type 3 (SCA3) is a progressive neurodegenerative disorder caused by a CAG repeat expansion in the *ATXN3* gene encoding the ataxin-3 protein. Despite extensive research the exact pathogenic mechanisms of SCA3 are still not understood in depth. In the present study, to gain insight into the toxicity induced by the expanded CAG repeats in SCA3, we comprehensively investigated repeat-associated non-ATG (RAN) translation in various cellular models expressing translated or non-canonically translated *ATXN3* sequences with an increasing number of CAG repeats. We demonstrate that two SCA3 RAN proteins, polyglutamine (polyQ) and polyalanine (polyA), are found only in the case of CAG repeats of pathogenic length. Despite having distinct cellular localization, RAN polyQ and RAN polyA proteins are very often coexpressed in the same cell, impairing nuclear integrity and inducing apoptosis. We provide for the first time mechanistic insights into SCA3 RAN translation indicating that *ATXN3* sequences surrounding the repeat region have an impact on SCA3 RAN translation initiation and efficiency. We revealed that RAN translation of polyQ proteins starts at non-cognate codons upstream of the CAG repeats, whereas RAN polyA proteins are likely translated within repeats. Furthermore, integrated stress response activation enhances SCA3 RAN translation. We suggest that RAN translation in SCA3 is a common event substantially contributing to SCA3 pathogenesis and that the *ATXN3* sequence context plays an important role in triggering this unconventional translation.

KEYWORDS

SCA3, CAG repeat expansion, RAN translation, translation initiation, non-cognate initiation codon, integrated stress response

ABBREVIATIONS

ATXN3 - ataxin-3, c9ALS/FTD - c9orf72 amyotrophic lateral sclerosis/frontotemporal dementia, CHOP - C/EBP homologous protein, DM1 - myotonic dystrophy type 1 and type 2, ER - endoplasmic reticulum, FECD - Fuchs' endothelial corneal dystrophy, FXTAS - fragile X tremor ataxia syndrome, iMet-tRNA - initiator Met-tRNA, ISR - integrated stress response, MetAPs - methionine aminopeptidases, NAT - N-terminal acetyltransferase, NPC - nuclear pore complex, polyA – polyalanine, polyQ – polyglutamine,

polyS – polyserine, RAN translation - repeat associated non-ATG translation, SA - sodium arsenite, SCA3 - spinocerebellar ataxia type 3, TG - thapsigargin

INTRODUCTION

Spinocerebellar ataxia type 3 (SCA3) is the most common form of dominantly inherited ataxia in the world and one of nine polyglutamine (polyQ) neurodegenerative diseases. This progressive and fatal disorder is primarily characterized by neuronal dysfunction and degeneration of the cerebellum and functionally related brain regions. SCA3 is caused by an expanded stretch of CAG repeats in exon 10 of the *ATXN3* gene, which encodes the deubiquitinating enzyme ataxin-3. These repeat tracts range in *ATXN3* from 12 to 44 triplets in healthy individuals and from 56 to 87 in SCA3 patients [1–4]. Until recently, it was believed that the expanded CAG repeats in SCA3 exerted their pathogenic effects at only the protein level. The presence of an abnormally long tract of glutamine in the C-terminal region of ataxin-3 leads to a toxic gain-of-function of the protein, which manifests mainly as characteristic neuronal cytoplasmic and nuclear aggregates [1–4]. However, there is also mounting evidence to suggest that mutant transcripts contribute to the pathogenesis of SCA3 *via* an RNA gain-of-function mechanism [1, 5, 6]. RNA toxicity in this polyQ disease can be induced by aberrant splicing of MBNL1-dependent genes due to the sequestration of this alternative splicing factor by mRNAs containing expanded CAG repeats [7, 8]. Moreover, an abnormal interaction between mutant *ATXN3* transcripts and nucleolin, resulting in the induction of nucleolar stress, was reported in a SCA3 model [9].

The pathogenesis of SCA3 is expected to be more complex, as additional mechanisms are being discovered in other repeat expansion diseases that may also function in SCA3. One such example is repeat-associated non-ATG (RAN) translation, which was described for repeat tracts of CAG, CUG, CGG, CCG, GGGGCC, and CCCGG and, recently, for CCUG, CAGG and UGGAA units [10–19]. These expanded repeat tracts, located mainly in the non-coding regions of specific genes, can trigger non-canonical translation, i.e., in the absence of an AUG initiation codon induce the synthesis of various sets of potentially toxic proteins composed of homopolymeric, dipeptide, tetrapeptide or even pentapeptide repeated tracts of amino acids. RAN translation can be initiated in multiple reading frames in both sense and antisense directions, leading to the expression of distinct proteins [10–19]. The precise mechanism of RAN translation initiation is still poorly understood; however, it has been shown that the repeat tract length and RNA structure formed by repeats play crucial roles [20–23]. More

recently, it was revealed that RAN translation of expanded CGG and GGGGCC repeats might utilize a cap-dependent and/or cap-independent ribosomal scanning mechanism for the initiation at near-cognate codons upstream of the repeats [24–30]. Furthermore, among the potential regulators of RAN translation process identified i.a. by genetic screens are translation initiation factors, DDX3X helicase and ribosomal protein S25 [31–35]. Additionally, the cellular stress response promotes this non-canonical mode of protein synthesis from CGG and GGGGCC repeats [26, 27, 29, 36]. A large body of evidence shows the high relevance of RAN proteins for neurodegeneration; these proteins form aggregates that accumulate in both the affected tissues of patients and animal models of SCA8, myotonic dystrophy type 1 (DM1), *c9orf72* amyotrophic lateral sclerosis/frontotemporal dementia (c9ALS/FTD), fragile X tremor ataxia syndrome (FXTAS), Huntington's disease (HD), DM2, SCA31 and Fuchs' endothelial corneal dystrophy (FECD) [37–39]. Depending on the type of repeats and the experimental model used, RAN protein toxicity is attributed to the impaired function of the ubiquitin proteasome system, aberrant nucleocytoplasmic transport or the induction of nucleolar and endoplasmic reticulum stress and usually results in enhanced cell death [37–41].

RAN translation is mainly described for diseases in which repeat expansion is located in non-coding regions of the associated genes. In polyQ disorders, in which repeat regions are present in open reading frames, RAN translation has not been extensively studied. There has been one report describing the importance of RAN translation in HD patients, and the initial results of another study suggested that this phenomenon might also occur in SCA2 [15, 42].

In the present study, we show that RAN translation is initiated at CAG repeats located in the sequence context of the *ATXN3* gene and might represent an additional toxic pathway involved in the pathology of SCA3. To show this, we generated multiple translated or non-canonically translated *ATXN3* constructs with different numbers of CAG repeats and various lengths of surrounding sequences and utilized them to analyze RAN translation in transfected cells. We demonstrate that SCA3 RAN translation occurs exclusively in cells expressing constructs containing pathogenic CAG repeats, and the efficiency of this process greatly depends on the *ATXN3* sequence flanking the repeat region. Interestingly, the resulting RAN proteins, i.e., polyQ and polyalanine (polyA) products exhibit different cytoplasmic and nuclear localization; however they very often coexist within the same cell. Importantly, in cells expressing these unusual proteins, the nuclear envelope structure is disturbed. Proteomic analysis revealed that RAN translation in the glutamine frame is initiated from non-cognate codons

upstream of the repeat region. We also found that various cellular stress conditions upregulate CAG RAN translation through eIF2 α phosphorylation.

MATERIALS AND METHODS

Generation of the *ATXN3* constructs

Full-length human *ATXN3* cDNA was previously cloned from total mRNA isolated from SCA3 patient-derived fibroblasts (GM06153, obtained from NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research) [43]. The cloning strategy used for the constructs expressing translated or untranslated *ATXN3* cDNA containing different lengths of CAG tracts was previously described [8, 44]. These constructs were prepared in a pCDH-EF1 α -MCS-BGH-PGK-GFP-T2A-Puro vector (System Biosciences). To monitor the expression of every possible RAN protein, these constructs were modified by cloning into an Eco0109I site, which is located downstream of the repeat region, a DNA oligo encoding three different protein tags (HA-tag for glutamine, Myc-tag for alanine and His-tag for serine reading frames). In the case of constructs containing shortened 5' and 3' flanking sequences, first, a sequence containing the same three tags and two stop codons in each frame was cloned into a pcDNA3.1 (Invitrogen) expression vector using EcoRI and Eco0109I restriction sites. Second, inserts containing fragments of the *ATXN3* sequence with various lengths of CAG repeats were amplified using the full-length cDNA of *ATXN3* as template constructs and specific primers that also introduced BamHI and EcoRI restriction sites at the 5' and 3' ends, respectively. Additionally, a forward primer was designed to add two stop codons before the 5' flanking sequence in each frame. PCR products were cut with BamHI and EcoRI and cloned into a pcDNA3.1 vector. Mutation of the near-cognate codon AAG to CAG was performed using a QuikChange II XL Site-Directed Mutagenesis Kit (Agilent). All generated constructs were verified by Sanger sequencing.

Cell culture

HEK293T and HeLa cells (ATCC) were grown in DMEM (Gibco) supplemented with 8% FBS (Biowest), 2 mM L-glutamine (Gibco) and 100 U/ml penicillin/100 μ g/ml streptomycin (Gibco). Human neuroblastoma SH-SY5Y cells (ATCC) were grown in DMEM/F12 medium (Gibco) supplemented with 10% FBS, 100 U/ml penicillin/100 μ g/ml streptomycin (Gibco), and 1x non-essential amino acids

(Sigma-Aldrich). The cells were maintained under standard conditions of temperature (37°C), humidity (95%) and carbon dioxide (5%). For longer storage periods, the cells were frozen in freezing medium and stored in liquid nitrogen.

Plasmid transfection and drug treatments

Routine cell transfection of HEK293T and HeLa cells with plasmids was conducted using Lipofectamine 2000 and Lipofectamine LTX (Life Technologies), respectively, according to the manufacturers' instructions. Prior to transfection, cells were plated at the required confluence on 12-well plates and harvested 48 h after transfection. For stress induction, HEK293T cells were transfected with the generated constructs for 20 h, followed by 8 h of treatment with 1 μ M thapsigargin (Enzo Life Technologies) or 50 μ M sodium arsenite (Sigma-Aldrich). SH-SY5Y cells were electroporated with a Neon™ Transfection System (Invitrogen) in 100- μ l tips using the following parameters: 1.100 V, 50 ms, and 1 pulse.

Western blotting

Routinely, cells were sonicated in lysis buffer containing a 60 mM TRIS-base, 2% SDS, 10% sucrose, 2 mM PMSF and phosphatase inhibitor cocktail (PhosSTOP; Roche). For the analysis of RAN translation after cellular stress induction, the cells were lysed in RIPA buffer (Sigma-Aldrich) containing protease inhibitor cocktail (cOmplete, EDTA-free; Roche) and phosphatase inhibitor cocktail. The protein concentration was estimated using a NanoDrop spectrophotometer. A total of 50 μ g of protein was diluted in sample buffer containing 2-mercaptoethanol, denatured at 95°C for 5 min, and separated on a 12% SDS-PAGE gel. After electrophoresis, the proteins were transferred onto a 0.1 μ m nitrocellulose membrane (Amersham), blocked with 5% non-fat dry milk in TBS-0.1% Tween-20 for 1 h and incubated overnight with mouse anti-ataxin-3 (1:1000; Millipore), mouse anti-HA (1:1000; Sigma-Aldrich), rabbit anti-c-Myc (1:1000, Cell Signaling Technology, CST), rabbit anti-His (1:1000, CST), rabbit anti-phospho-eIF2 α (1:1000, CST), mouse anti-eIF2 α (1:1000, CST) or mouse anti-GAPDH (1:20000, Millipore) antibodies. Next, the membranes were incubated with secondary anti-mouse or anti-rabbit (Jackson ImmunoResearch) peroxidase-conjugated antibodies for 1 h, at room temperature (RT). The immunoreaction was detected using the WesternBright Quantum HRP Substrate (Advansta) and G:BOX chemiluminescent camera (Syngene). Band intensity was measured using ImageJ software.

Mass spectrometry analysis

To immunoprecipitate AUG-initiated and RAN proteins (polyQ and polyA), HEK293T cells transfected with AUG-78/32-117CAG, 78/32-115CAG or 33/32-119CAG constructs were lysed after 48 hours in RIPA buffer containing protease and phosphatase inhibitor cocktails. HA-tagged polyQ and Myc-tagged polyA proteins were enriched using Pierce™ Anti-HA Agarose (Thermo Fisher Scientific) and Myc-Trap®_A Kit (Chromotek) respectively, according to the manufacturer's instructions. After washing steps the agarose beads bound with polyQ and polyA proteins were further processed for LC-MS/MS analysis. Mass spectrometry experiments were performed and analyzed at the Mass Spectrometry Laboratory at the Institute of Biochemistry and Biophysics Polish Academy of Sciences. Agarose beads suspended in 40 µl of immunoprecipitation washing solution were filled up to 150 µl with 100 mM ammonium bicarbonate buffer. Cysteine bridges were reduced by 1 h incubation with 5 mM tris(2-carboxyethyl)phosphine (TCEP) at 60°C followed by 30 min alkylation at RT in the dark with 15 mM iodoacetamide (IAA). The samples were mixed thoroughly and divided into equal parts for digestion with various enzymes selected on the basis of the distribution of potential cleavage sites in the regions surrounding the polyQ and polyA sequences. All enzymes were of sequencing grade purity and were purchased from Promega Corporation. PolyQ samples were digested with 0.01 µg/µl of LysC or 0.005 µg/µl of trypsin. PolyA samples were digested with trypsin and elastase in a final concentration of 0.01 µg/µl. After overnight digestion, solutions were acidified with 0.1% trifluoroacetic acid and centrifuged (20000 g, 15 min, 4°C) to remove agarose beads from peptide solution. Next, 20 µl of each sample was analyzed using nanoACQUITY UPLC (Waters) directly coupled to a QExactive mass spectrometer (Thermo Fisher Scientific). Peptides were trapped and washed on C18 column (180 µm x 20 mm, Waters) with 0.1% formic acid (FA) in water as a mobile phase and transferred to a nanoACQUITY BEH C18 column (75 µm x 250 mm, 1.7 µm, Waters) using acetonitrile (ACN) gradient (0 – 35% ACN in 160 min) in the presence of 0.1% FA at a flow rate of 250 nl/min. Acquisition was performed in the data-dependent mode with the maximum of 12 precursors selected for MS2. Full MS scans covering the range of 300–1650 m/z were acquired at a resolution of 70000 with a maximum injection time of 60 ms and an AGC target of 1e6. MS2 scans were acquired at the resolution of 17500 and the AGC target of 5e5. Dynamic exclusion was set to 30 s. Acquired data was pre-processed with Mascot Distiller software (2.6, Matrix Science) and a protein identification was performed with Mascot Search Engine (Matrix Science). The searches were done against a homemade database of all potential polyQ and polyA

proteins translated from the transfected constructs and the most frequent contaminations (6555 sequences). To reduce mass errors, the peptide and fragment mass tolerance settings were established separately for individual LC-MS/MS runs after a measured mass recalibration, as described previously [45]. The rest of search parameters were as follows: enzyme, Trypsin or Lys-C; missed cleavages, 2 or 3; fixed modifications, Carbamidomethyl (C); variable modifications, Oxidation (M), Deamidation (NQ), Acetyl (N-term), Gln->pyro-Glu; instrument, HCD. Since there was not enough identifications to compute the FDR, peptides identified for polyQ and polyA sequences, for which ion scores were ≥ 20 , were further analyzed manually.

Immunofluorescence

HEK293T cells were seeded on 12-well plates with glass coverslips (Marienfeld Superior) coated with poly-L-lysine (Sigma-Aldrich) one day prior to transfection. Forty-eight hours post transfection, the cells were fixed with 4% paraformaldehyde for 30 min at RT, permeabilized with 0.5% Triton-X-100 in PBS for 10 min and blocked in 1% BSA for 1 h at RT. The coverslips were incubated overnight at 4°C with the following primary antibodies diluted in 1% BSA in PBS: anti-HA (1:100), anti-c-Myc (1:100), anti-His (1:100), mouse anti-nuclear pore complex (1:1000, BioLegend), and rabbit anti-cleaved-caspase 3 (1:100, CST). The next day, the coverslips were rinsed with PBS and incubated with anti-mouse and anti-rabbit antibodies labeled with Alexa488 and Alexa594, respectively (1:500, Jackson ImmunoResearch), for 1 h at RT. After rinsing the coverslips were mounted with SlowFade Diamond Antifade Mountant with DAPI (Life Technologies). Images were captured with a Leica SP5 confocal microscope and processed using Leica software.

Statistical analysis

Statistical significance was assessed by an unpaired *t* test and one-way ANOVA with Bonferroni correction for multiple comparisons. A p-value of <0.05 was considered significant. The data are presented as mean $\pm$ standard error of the mean (SEM). The results represent at least three independent biological replicates.

RESULTS

RAN translation occurs in the *ATXN3* context

To evaluate whether CAG repeats in the sequence context of *ATXN3* can support RAN translation, we developed a series of constructs in pcDNA3.1 plasmid, without a start codon, containing normal (21 CAG) or expanded repeat numbers (61 and 115 CAG) (Fig. 1a, upper scheme). These plasmids contained 78 bp and 32 bp of the *ATXN3* sequence upstream and downstream of the repeat region, respectively. The length of the 5' flanking sequence was limited by the many ATG codons within the *ATXN3* sequence upstream of the CAG repeat region that were undesired in the constructs, whereas the length of the 3' flanking sequence was determined with respect to the first stop codon for the alanine reading frame (Fig. S1a). Thus these specific *ATXN3* sequences are expected to be involved in SCA3 RAN translation in case of repeat tract expansion. The generated constructs also contained upstream and downstream stop codons and C-terminal tags located downstream of the 3' flanking sequence (Fig. 1a). As a control for AUG-initiated translation in the glutamine reading frame, we developed an additional construct with 117 CAG repeats containing a start codon positioned immediately before the 5' flanking sequence (Fig. 1a, lower scheme). All these constructs were transfected into HEK293T cells, and the potential RAN translation products were analyzed 48 h later.

In the absence of an AUG codon, RAN polyQ and RAN polyA proteins were detected for constructs containing expanded CAG repeat tracts (61 and 115 CAG) (Fig. 1b and c). We did not observe any specific signal from potential polyserine (polyS) proteins (Fig. 1d). In the case of the construct with unexpanded repeat tracts (21 CAG), RAN translation did not occur (Fig. 1b-d). RAN polyQ proteins migrated as two bands, for which the molecular weights and amounts increased with increasing CAG repeat tract length (Fig. 1b, Fig. S1b). The level of SCA3 RAN polyQ proteins from the 78/32-115CAG construct was significantly lower than that of the AUG-initiated polyQ proteins (Fig. S1c). Moreover, expression of the AUG-initiated polyQ proteins led to the formation of insoluble aggregates of high molecular weight that also remained in the stacking gel (Fig. 1b). RAN polyA proteins translated from the 61 CAG repeats migrated as one clear band below 17 kDa, whereas for the longer repeats (115 CAG), additional multiple proteins of even higher molecular weight appeared (Fig. 1c). Importantly, RAN polyA proteins were also observed in the cells expressing the AUG-containing construct, which indicates that the presence of the start codon did not inhibit RAN translation in other reading frames

(Fig. 1c). The obtained polyA proteins expression pattern was similar to that shown in previous results with respect to other repeat expansion diseases and might suggest that RAN translation in the alanine reading frame initiates at multiple sites within the repeats, leading to the generation of products of diverse molecular weights [10, 11, 15, 24].

SCA3 RAN polyQ and RAN polyA proteins exhibit different subcellular localization and aggregation properties

To determine the subcellular distribution of SCA3 RAN polyQ and RAN polyA proteins, we performed immunofluorescence analysis. RAN polyQ proteins were distributed throughout the cytoplasm and nucleus, whereas RAN polyA proteins were exclusively cytoplasmic (Fig. 2a, Fig. S2). Moreover, SCA3 RAN polyQ proteins exhibited aggregation properties; however, the number of aggregates was fewer compared to those formed by the AUG-initiated polyQ proteins and was dependent on the length of the repeat tract (Fig. 2a and b, Fig. S2). Approximately 3% and 22% of the cells contained aggregates formed by RAN polyQ proteins translated from constructs bearing 61 and 115 CAG repeats, respectively (Fig. 2b). We did not detect inclusions formed by RAN polyA proteins; however, we cannot exclude the possibility that these proteins exist as soluble oligomers (Fig. 2a, Fig. S2) [46–48]. Importantly, SCA3 RAN polyQ and RAN polyA proteins were very often coexpressed in the same cells, which might suggest that their translation was triggered by the same mechanism (Fig. 2a, Fig. S2). Moreover, RAN polyA proteins were also detected in cells containing AUG-initiated polyQ proteins (Fig. 2a, Fig. S2). As expected, we did not find RAN translation products in the cells expressing the nonpathological length of CAG repeats (21 CAG) (Fig. S2). Moreover, as revealed previously by western blot analysis, we did not detect any polyS proteins (data not shown).

SCA3 RAN proteins affect nuclear integrity and cause apoptosis

Many reports show that RAN proteins are toxic to cells [37–41]. During a microscopic analysis, we observed that, in the cells expressing SCA3 RAN proteins, the shape of the nucleus was disturbed (Fig. 2, Fig. S2). It had been previously shown that perinuclear aggregates formed by polyQ-expanded proteins induce focal distortion of the nuclear envelope [49]. More recently, it was reported that mutant huntingtin disrupts nuclear architecture, which results in aberrant nucleocytoplasmic transport [50, 51]. To analyze the nuclear membrane integrity in the cells expressing SCA3 RAN proteins,

immunofluorescence staining of nuclear pore complex (NPC) proteins was performed. As RAN polyQ and RAN polyA proteins frequently coexist in the same cell (Fig. 2a, Fig. S2), we studied the correlation of the nuclear morphology with the RAN polyA proteins only. We found that the detection of RAN polyA proteins was associated with altered NPC localization, which resulted in defects in nuclear envelope assembly (Fig. 3). In contrast, in the cells transfected with empty plasmid or expressing CAG repeats of normal length, the nuclear integrity was not disturbed (Fig. 3).

Disorganization of the nuclear envelope is often connected with the apoptosis process [52]. To address whether SCA3 RAN translation can trigger apoptosis, we additionally analyzed the presence of cleaved caspase-3, a well-known marker of apoptosis induction. As revealed by immunofluorescence, approximately 40% and 8% of the cells expressing AUG-containing and AUG-lacking constructs, respectively, were positive for both cleaved caspase-3 and polyQ proteins (Fig. S3a and b). This finding suggests a substantial impact of non-canonical translation on apoptosis induction, as the cumulative effect observed for the AUG construct was attributed to the presence of both canonically translated and RAN-translated proteins. In conclusion, our data suggest that the expression of SCA3 RAN proteins substantially affects nuclear integrity and promotes cell death.

SCA3 RAN translation is influenced by the sequences surrounding the expanded CAG repeat region

In the next step, we verified whether RAN translation could also be observed for the full-length *ATXN3* transcript in our experimental setup. We used constructs with various lengths of CAG repeats containing translated (20, 60, 85 and 122 CAG repeats) and non-canonically translated (20, 60, 91 and 126 CAG repeats) full-length *ATXN3* cDNA. Sequences encoding three epitope tags were added immediately downstream of the CAG repeats (Fig. S4a). These constructs were transfected into HEK293T cells, and the level of SCA3 RAN translation was analyzed 48 h after transfection. The cells expressing the translated *ATXN3* constructs produced one strong band corresponding to the overexpressed exogenous ataxin-3 with various lengths of polyQ tracts and additional multiple bands, which may have represented proteins produced from alternative initiation sites or post-translationally modified ataxin-3 (Fig. S4b and c). However, no specific RAN polyA or RAN polyS proteins were detected in cells transfected with these constructs (Fig. S4d and e). In the case of constructs without AUG codons, as

expected, the exogenous ataxin-3 expression was abolished (Fig. S4f and g). However, still SCA3 RAN translation was not observed in the cells with any of the three possible frames (Fig. S4g-i).

As we observed a significant difference in the occurrence of RAN translation in the full-length *ATXN3* constructs and their shortened versions, we decided to better recognize the sequence requirements for RAN translation initiation in mutant *ATXN3*. We modified the SCA3 RAN construct 78/32-115CAG by removing 33 nt or 60 nt of the 5' sequence and 15 nt of the 3' sequence from the repeats (Fig. S5a). HEK293T cells were transfected with these newly established vectors (45/32-110CAG, 45/17-117CAG, 18/32-119CAG and 18/17-116CAG), and 48 h after transfection, the SCA3 RAN translation was analyzed. Deleting 33 nt of the sequence upstream and/or 15 nt of the sequence downstream of the repeat tract (45/32-10CAG and 45/17-117CAG) did not affect RAN translation in the glutamine frame (Fig. S5b). In contrast, when we truncated the 5' sequence to 18 nt, the RAN translation in the glutamine frame was abolished (Fig. S5b). In the case of the RAN polyA proteins, we did not observe this high dependency between the occurrence of RAN translation and the length of the 5' flanking sequence (Fig. S5c). However, we found a significant effect of the 3' flanking sequence on the RAN translation efficiency in the alanine frame. Shortening of the 3' flanking sequence to 17 nt while maintaining the length of the 5' end (45/32-110CAG vs 45/17-117CAG and 18/32-119CAG vs 18/17-116CAG) led to a decrease in the level of RAN polyA proteins (Fig. S5c and d). These results show the profound involvement of the *ATXN3* 5' flanking sequence in the RAN translation initiation in the glutamine frame and the influence of the 3' flanking sequence on polyA translation efficiency.

To explain the high sequence dependence of SCA3 RAN translation in the glutamine frame, we analyzed the 5' flanking sequence for the presence of alternative start codons (Fig. S6a). It had been reported that RAN translation of CGG and GGGGCC repeats is initiated at near-cognate codons located upstream of the repeat tract [11, 24–26, 28–30]. Alternative start codons differ from AUG codon by only one nucleotide and generally are less efficient than standard AUG codon [53, 54]. Two near-cognate AAG codons were found in the glutamine frame in the *ATXN3* sequence in 78/32-115CAG, 45/32-110CAG and 45/17-117CAG constructs (Fig. S6a). In the constructs with the *ATXN3* 5' flanking sequence truncated to 18 nt (18/32-119CAG and 18/17-116CAG), one of these codons was absent, which suggested the potential involvement of this AAG codon in RAN translation initiation (Fig. S6a). To examine this possibility, in the 45/32-110CAG and 45/17-117CAG constructs, we mutated this AAG codon to CAG codon using site-directed mutagenesis. Western blot analysis of the cells transfected

with these mutated constructs did not show AAG codon requirements for the initiation of RAN polyQ proteins (Fig. S6b). Additionally, the introduced mutation did not affect RAN translation of RAN polyA proteins (Fig. S6c)

Thus, to determine the minimal sequence requirements for translation initiation in the glutamine frame, we designed an additional set of constructs (33/32-119CAG, 33/17-118CAG, 24/32-116CAG and 24/17-119CAG), in which the 5' flanking sequence was shortened to 33 nt and 24 nt, whereas the 3' flanking sequence was shortened to 17 nt (Fig. 4a). When we truncated the 5' flanking sequence to 33 nt, we were surprised to observe enhanced RAN translation for the polyQ proteins compared to that of the longer constructs, whereas 3' shortening did not affect translation (Fig. 4b). Interestingly, the resultant RAN polyQ proteins more often formed aggregates, as determined by western blotting and confocal microscopy (Fig. 4b and e). In contrast, when the 5' flanking sequence was shortened to 24 nt, RAN translation in the glutamine frame was abolished (Fig. 4b). Similar to previous experiments, truncation of the 3' flanking sequence to 17 nt resulted in decreased RAN translation efficiency in the alanine frame (Fig. 4c and d).

We also examined the influence of the 5' flanking sequence on SCA3 RAN translation initiation in the glutamine frame in other cell lines. Transfection of HeLa and SH-SY5Y cells with constructs differing only in the length of the 5' sequence upstream of the CAG region (78/32-115CAG, 45/32-110CAG, 33/32-119CAG, 24/32-116CAG and 18/32-119CAG) demonstrated that SCA3 RAN translation can proceed in different cellular backgrounds (Fig. S7). However, in contrast to that in the HEK293T cells, RAN translation in the other cell lines, particularly in neuroblastoma cells, was less efficient. This effect can also be explained by differences in the transfection efficiency between the studied cell lines. In the case of the HeLa cells, we were able to detect the signal for AUG-initiated polyQ proteins and RAN polyQ proteins produced from the 33/32-119CAG construct and a very weak signal for RAN polyA proteins expressed from studied RAN constructs (Fig. S7a and 7b). We also observed the same polyQ proteins in SH-SY5Y cells as in the HeLa cells; however, in contrast to the HeLa cells, there was no specific signal for RAN polyA proteins in the SH-SY5Y cells (Fig. S7c and d). Taken together, these data seem to indicate that the 5' flanking sequence, particularly that located 24 nt upstream of the CAG repeats, plays a significant role in RAN translation initiation in the glutamine frame, and the efficiency of RAN translation might be cell-type dependent.

SCA3 RAN translation in the glutamine frame is initiated at upstream non-cognate codons

In order to characterize translation initiation sites solely dependent on RAN translation, we used immunoprecipitation of RAN polyQ and RAN polyA proteins from the HEK293T cells transfected with 78/32-115CAG and 33/32-119CAG constructs and analyzed specific peptides by mass spectrometry. Additionally, we immunoprecipitated the AUG-initiated polyQ proteins from the cells transfected with the AUG-78/32-117CAG construct as a control to show the events dependent on canonical translation initiation. To map the initiation sites of RAN and canonical translation products with high confidence, as well as to exclude any bias caused by digestion, we digested immunopurified AUG-initiated and RAN polyQ proteins with two different proteases, trypsin and LysC (Fig. S8a, Fig. S9a, Fig. S10a). Moreover, the RAN polyA proteins were digested with trypsin and with an unspecific protease, i.e., elastase (Fig. S8b, Fig. S9b, Fig. S10b). These proteolytic digestion selections enabled us to identify many unique peptides from both polyQ and RAN polyA proteins, many of which originated from regions downstream of the CAG repeats (Fig. 5, Fig. S8, Fig. S9, Fig. S10). Moreover, the analysis of our LC-MS/MS data compared with that of the human proteome (UniProtKB Release 2020_01) resulted in no unique peptides that could have originated from the proteolytic digestion of the endogenous, full-length ATXN3, confirming that all the analyzed peptides originated from immunoprecipitated AUG-initiated polyQ, RAN polyQ and polyA proteins.

Interestingly, our proteomic analysis revealed that the canonically translated polyQ proteins from AUG-78/32-117CAG control construct had an initial acetylated serine, as revealed by both LysC and trypsin digestion, i.e., the obtained peptide sequence was ^{AC}SGTNLTSEELRK (Fig. 5a and b, Fig. S8a). No peptides from polyQ proteins initiated from the canonical AUG codon contained N-terminal methionine, indicating that, although the first AUG codon is indeed decoded by an initiator Met-tRNA (iMet-tRNA), the incorporated methionine is removed rapidly by methionine aminopeptidases (MetAPs) followed by the efficient acetylation of the novel protein N-terminus by an N-terminal acetyltransferase (NAT) complex (Fig. 5b, Fig. S8a). Our observation is consistent with finding from previous studies showing that the complete cleavage of iMet by MetAPs is achievable for the following substrates: Met-Ser-, Met-Thr-, Met-Gly-, Met-Ala-, Met-Val-, Met-Cys- and Met-Pro- [55, 56]. Furthermore, the resulting new N-terminus (excluding Pro) was shown to be very efficiently Nt-acetylated by a major NAT complex – the NatA complex [57–59]. Since the Nt-acetylome is estimated to include as many as 80–90% of soluble human proteins, it provides a reliable indication of where translation initiation starts [59, 60].

In the case of the RAN polyQ proteins expressed in the HEK293T cells transfected with the 78/32-115CAG construct, we identified unique N-terminally acetylated peptides, ^{AC}TSEELRK (2 peptides) and ^{AC}SEELRK (2 peptides), that indicated the sites of RAN translation initiation (Fig. 5c-e, Fig. S9a). These data, obtained for both LysC and trypsin digestion, suggest that, similar to the proteins synthesized by canonical translation, the initial amino acid incorporated into this RAN polyQ protein is methionine, which is removed by MetAPs followed by N-terminal acetylation of the second amino acid (Thr or Ser) by NatA complex. Interestingly, our results strongly suggest that the expanded CAG repeats in *ATXN3* can induce RAN translation initiation at neighboring non-cognate codons, i.e., CUU and ACU (Fig. 5c-e, Fig. S9a). Moreover, the heterogeneity of the RAN translation initiation events may not be limited to only these two favorable codons, as some indications of RAN translation initiation upstream and downstream of the CUU and ACU codons have also been indicated by LC-MS/MS analysis (Fig. S9a) and previous RAN translation analysis based on constructs with shorter 5' flanking sequences, e.g., 45/32-110CAG and 33/32-119CAG constructs (Fig. 4, Fig. S5).

In the case of RAN polyQ proteins isolated from the HEK293T cells transfected with the 33/32-119CAG construct, we failed to identify peptides showing RAN translation initiation site(s) upstream of the expanded CAG repeats, most likely because of the suboptimal peptide size (<5 amino acids) generated after LysC and trypsin digestion, which makes them difficult to identify in complex mixtures and through database searching (Fig. S10a). Similarly, in the case of all the RAN polyA proteins analyzed by LC-MS/MS, we were not able to identify, with high confidence, the location of RAN translation initiation (Fig. S8b, Fig. S9b, Fig. S10b). Based on our previously described experiments showing the extensive heterogeneity of RAN polyA protein sizes and because of the naturally occurring stop codon only eight amino acids upstream of the expanded repeats in *ATXN3*, we propose that the vast majority of the RAN polyA proteins were initiated within repeats at the GCA codons.

Cellular stress enhances SCA3 RAN translation of the CAG repeats

In response to diverse stress stimuli such as hypoxia, viral infections or the presence of misfolded proteins, eukaryotic cells activate a cytoprotective signaling pathway termed the integrated stress response (ISR) to restore cellular homeostasis. The core event of ISR is phosphorylation of the α subunit of eIF2, which leads to global protein synthesis attenuation accompanied by the induction of alternative mechanisms of translation initiation of selected mRNAs [61–64]. Recently, it has been

reported that RAN translation at CGG and GGGGCC repeats is enhanced by ISR activation [26, 27, 29, 36]. To examine whether this phenomenon can also occur in SCA3, cells transfected with selected RAN constructs (78/32-115CAG, 45/32-110CAG and 33/32-119CAG) were additionally exposed for 8 h to thapsigargin (TG), which causes endoplasmic reticulum (ER) stress, or to sodium arsenite (SA), which induces oxidative stress. As expected, both stressors led to increased eIF2 α phosphorylation and upregulation of transcription factor C/EBP homologous protein (CHOP), one of the effectors of ISR, in both untransfected and transfected cells (Fig. 6a and b; Fig. S11a-d). Next, we found that treatment with either of these stress inducers significantly increased RAN translation for the polyA proteins in the cells expressing all the studied constructs (Fig. 6d, e, g and h, Fig. S11f and h). We also observed enhanced RAN translation for the polyQ proteins in cells treated with TG (Fig. 6c and e, Fig. S11e). In the case of oxidative stress, elevated levels of RAN polyQ proteins were observed only in the cells expressing the 78/32-115CAG and 45/32-110CAG constructs (Fig. 6f and h, Fig. S11g). In contrast to the RAN polyA proteins, treatment with SA did not increase the RAN polyQ levels in the cells expressing the 33/32-119CAG construct (Figure 6f and h).

As many cellular stressors negatively interfere with cap-dependent translation, we also studied the effect of ER and oxidative stress on the translation in cells transfected with AUG-78/32-117CAG construct. It is worth noting that, in addition to AUG-initiated polyQ proteins and RAN polyA proteins, RAN polyQ proteins could also be produced in these cells. In the cells transfected with the AUG-78/32-117CAG plasmid, the levels of eIF2 α phosphorylation and CHOP were also markedly elevated under stress conditions (Fig. S11a-d). The administration of TG to cells expressing the AUG-78/32-117CAG construct caused a slight increase in polyQ protein levels (Fig. 6e, Fig. S11e), in contrast to the SA treatment, for which no such change was observed (Fig. 6h, Fig. S11g). Importantly, both treatments substantially elevated RAN translation for polyA proteins in the cells transfected with the AUG-78/32-117CAG plasmid (Fig. 6e and h; Fig. S11f and h). Thus, these data show that ISR can also enhance the production of RAN proteins at CAG repeats in SCA3.

DISCUSSION

The discovery of RAN translation has revolutionized not only the understanding of the mechanisms underlying repeat expansion disorders but has also indicated novel therapeutic targets to combat these devastating diseases. To date, this unconventional translation has been reported in eight distinct

diseases, including SCA8, DM1, c9ALS/FTD, FXTAS, HD, SCA31, DM2 and FECD, and a growing body of evidence clearly indicates that RAN translation products cause cellular toxicity [37–41]. Generally, RAN translation has been found to occur mainly across large repeat expansions in non-coding regions of the transcripts. Many reports have shown that these expanded repeats in RNA form a very stable secondary structure that is required for RAN translation [65–67]. It has been reported that mutant CAG repeats in polyQ diseases can also fold into stable hairpins, which might suggest that these repeats can also trigger RAN translation [66]. However, because of the location of these CAG repeats within open reading frames, studying RAN translation in polyQ diseases is technically challenging. To date, only Banez-Coronel and coworkers have provided strong evidence that this unusual translation can occur at CAG repeats in coding regions [15]. They showed the accumulation of four novel RAN proteins in HD human brain samples, which may be implicated in disease pathology.

Here, we report the first comprehensive study showing that RAN translation occurs at expanded CAG repeats in the *ATXN3* context and can contribute to the pathogenesis of SCA3. Furthermore, we provide, for the first time, some insights into the RAN translation mechanism in SCA3, showing that, in addition to the pathological length of expansion, the sequences surrounding the repeat region play important roles in SCA3 RAN translation initiation and efficiency. Moreover, SCA3 RAN translation is enhanced by the activation of ISR pathways.

We demonstrated the existence of two SCA3 RAN-translated polyQ and polyA proteins in cells transfected with constructs containing elongated CAG repeat regions and *ATXN3* sequence context without ATG codons. During the study of RAN translation in SCA8 and DM1, this unusual translation mode was also suggested for SCA3; however, a very short (20 bp) 5' flanking sequence of *ATXN3* was used [10]. In our study, we used longer sequences surrounding the expansion region (78 bp in the 5' direction of the repeats and 32 bp in the 3' direction of the repeats), as it has been demonstrated that both the 5' and 3' flanking sequences of *ATXN3* influence the CAG repeat structure, which is known to be important for RAN translation initiation. We previously showed that 3' terminal CAG repeats pair with an *ATXN3*-specific region named a pseudo-repeat sequence located in the 5' direction of the repeat tracts [68]. Another rationale for the length of flanking sequences was the occurrence of the last ATG codons upstream of the repeats and the presence of the first stop codon in the alanine reading frame downstream of the repeats. It is very possible that these selected *ATXN3* flanking regions represent the minimal SCA3 RAN-translational unit.

SCA3 RAN polyQ proteins, both in diffused form or as aggregates, exhibited cytoplasmic and nuclear localization, whereas diffused SCA3 RAN polyA proteins were found exclusively in the cytoplasm, which is consistent with previous data regarding the subcellular distribution of other proteins containing expanded glutamine or alanine tracts [10, 15, 69, 70]. It is worth mentioning that the expanded polyA tracts can serve as nuclear export signals, and the cytoplasmic accumulation of polyA-containing proteins is dependent on the length of the alanine tract [71]. Despite distinct cellular localization, SCA3 RAN polyQ and polyA proteins are frequently found in the same cell, which suggests not only the common mechanism for RAN translation in various frames in SCA3 but also the cumulative toxic effect of these two distinct proteins in cells. We were not able to detect RAN polyS proteins, which may be explained by not having the optimal sequence and/or structure context to enable efficient RAN translation from this reading frame.

To date, an enormous effort has been made to better recognize the role of RAN proteins in the pathogenesis of repeat expansion disorders. Using various cellular and animal models it has been shown that RAN translation affects proteasome activity and nucleocytoplasmic transport, activates oxidative and ER stress, impairs protein translation and stress granule dynamics and/or induces cell death [37–41]. Here, we demonstrate that SCA3 RAN polyQ and RAN polyA proteins, similar to the RAN proteins described for other repeat expansion diseases, can cause cellular toxicity. First, these RAN polyQ proteins form aggregates, which are common pathological features of most neurodegenerative diseases and are considered potentially harmful to cells [72–74]. Second, in the cells expressing SCA3 RAN proteins, apoptosis is induced. Finally, the presence of SCA3 RAN proteins is associated with changes in nuclear envelope morphology, as determined by nuclear pore complex staining. The last finding is particularly important because it shows, for the first time, that RAN translation of CAG repeats might affect nuclear membrane integrity, potentially leading to dysfunction of nucleocytoplasmic transport. It was previously shown using i.a. HD mouse models that mutant huntingtin disrupts the nuclear envelope architecture and nucleocytoplasmic transport [50, 51]. There are also several reports indicating that nucleocytoplasmic transport is also impaired by RAN proteins [51, 75–79]. Moreover, interaction of RAN polyglycine (polyG) proteins translated from CGG repeats in FXTAS with the nuclear lamina protein LAP2 β has been reported, which is associated with disorganization of the nuclear lamina structure [25]. Alterations in the nuclear lamina are also induced by dipeptide repeat proteins produced from GGGGCC repeats [75, 80]. The precise role of SCA3 RAN

proteins in disease pathogenesis requires further investigation; however, it is very likely that the toxicity of these proteins could be associated with nuclear dysfunction, such as an impaired nuclear envelope and aberrant nucleocytoplasmic transport, leading to cell death.

Consistent with previous work on other repeat expansions, RAN translation of CAG repeats in the *ATXN3* sequence context is also strongly influenced by repeat length [20–23]. We did not observe RAN products translated at a nonpathogenic number of CAG repeats, and the amount of both RAN polyQ and RAN polyA proteins substantially increased with longer CAG expansion tracts. Moreover, we revealed an important role of the sequences surrounding the repeat region on SCA3 RAN translation initiation and efficiency. We showed that truncating the 5' flanking sequence to 24 nt prevented RAN translation only in the glutamine frame, which suggested translation initiation upstream of the repeats, whereas the length of the downstream sequence affected RAN polyA proteins expression. Using mass spectrometry analysis, we confirmed that the synthesis of RAN polyQ proteins, in fact, is initiated at sequence preceding the CAG repeats. In the case of the RAN polyA proteins, we were not able to detect peptides indicating the initiation upstream of the repeats, which supports the hypothesis that translation in the alanine frame occurs within the repeats [10, 11]. Previous research on RAN translation in SCA8, FXTAS or c9ALS/FTD has provided evidence that the length and the content of the 5' flanking sequences influence the initiation of RAN translation in specific reading frames, whereas sequences located downstream of the repeats can affect RAN translation in SCA2 [10, 11, 24–26, 28–30, 81]. Insertion of a stop codon in the *ATXN8* sequence immediately before the CAG repeats impedes RAN translation in the glutamine frame but not in the alanine or serine frames [10]. Similarly, shortening the *FMR1* 5' UTR sequence located upstream of the CGG repeats or introducing stop codons at specific sites in this sequence inhibits the expression of RAN polyG but not RAN polyA proteins [11, 24, 25]. Further studies have shown that RAN polyG proteins are required for the initiation at an ACG near-cognate codon located 32 nt upstream of the CGG repeats [25]. Additionally, it has been revealed that RAN translation of GGGGCC repeats occurs at the near-cognate CUG codon located 24 nt before the repeats, and deletion of the sequence containing this codon abolishes translation [26, 28–30, 81]. The *ATXN3* 5' flanking sequence used in our research contains two near-cognate AAG codons in the glutamine frame, one located 45 nt and the second 9 nt in the 5' direction of the repeats. The presence of these codons could explain the appearance of two bands of RAN polyQ proteins, as well as the lack of expression of these proteins when the upstream sequence is shortened. However, the mutation of

one of these codons (45 nt upstream) and analysis of the RAN translation of the 33/32 construct have shown that this codon is not involved in the translation initiation for the RAN polyQ proteins. Moreover, our mass spectrometry analysis revealed that the SCA3 RAN translation in the glutamine frame is initiated at the neighboring non-cognate CUU and ACU codons. The observed results could be explained by the fact that the AAG codon is one of the least efficient near-cognate start codons; moreover, in the *ATXN3* sequence, it is in a poor Kozak context [53, 54]. This finding is in agreement with proteomic studies showing that ACG and CUG codons are indeed initiation sites in the RAN translation of CGG and GGGGCC repeats, respectively [25, 30]. Generally, these codons appear to be the most efficient near-cognate start codons, and they are embedded in a good Kozak consensus sequence in causative genes for FXTAS and c9ALS/FTD [25, 30, 53, 54]. Based on previous studies concerning structural requirements for translation initiation, we initially assumed, that a stable hairpin formed by expanded CAG repeats downstream from the AAG codons would increase the likelihood of initiation at these sites [82–85]. Why does RAN translation in the glutamine frame initiate from CUU and ACU codons in a weak Kozak motif? It is likely that the structural context of these non-cognate codons, together with the downstream hairpin structure formed by the CAG repeats, makes these sites more accessible to the ribosome for the initiation of translation compared to the sites of the near-cognate AAG codons. It has been previously shown that structural RNA elements in which the initiation AUG codon is embedded have a strong impact on translation i.e., placing an AUG codon within a highly stable hairpin motif substantially reduced translational efficiency [86–90]. Nevertheless, despite the initiation at CUU and ACU codons that encode leucine and threonine, respectively, our data indicate that methionine is likely to be the initial amino acid incorporated into the RAN polyQ proteins. However, due to the specific amino acids in the second position, this methionine is completely removed by MetAPs, and the next amino acid, threonine or serine, is acetylated by NAT complex [55–59]. Previous mass spectrometry analysis of proteins produced through RAN translation of CGG and GGGGCC repeats also revealed that, despite imperfect match methionine was the initial amino acid for polyG translated from the ACG codon and for glycine-alanine dipeptide repeat proteins translated from the CUG codon [25, 30]. In contrast to the results of our analysis, due to the presence of glutamic acid in the second position, methionine could not be excised and therefore was detected in the identified peptides [25, 30]. N-terminal methionine was also not observed in the peptides derived from SCA8 RAN polyA proteins, which was explained as a potential consequence of the removal by a specific peptidase

or an initiation mechanism without a requirement for methionine [10]. In conclusion, our findings suggest that RAN translation can also be initiated from non-cognate codons incorporating methionine as the first amino acid.

In the present study, the occurrence of RAN translation across a full-length *ATXN3* transcript was also examined. We used constructs expressing, in addition to different lengths of CAG repeats, two variants of *ATXN3* cDNA: one canonically translated because of the presence of an ATG start codon and the second one in which stop codons were introduced to abolish conventional *ATXN3* translation. Regardless of the constructs used, including those containing expanded CAG repeats, we were not able to observe RAN proteins. This could be due to the fact that RAN translation of full-length mutant *ATXN3* mRNA occurs but at undetectable level until the toxicity-inducing events favoring this unconventional mode of translation successively accumulate as a result of disease progression. We cannot also exclude the possibility that SCA3 RAN translation might occur across at not yet identified, shorter and more pathogenic *ATXN3* isoforms, such as the transcripts used in our study, resulting from aberrant splicing. To date, 56 alternative splicing variants of *ATXN3* have been described [91]. In our study, we used the transcript containing 11 exons that encode the predominant *ATXN3* isoform in human and murine brain tissue and an additional SNP (CGG instead GGG) immediately downstream of the CAG repeat, which is known to be associated with SCA3 [92–95]. As the presence of a CAG expansion sequence in the SCA3 YAC mouse model is associated with increased level of alternatively spliced *ATXN3* isoform lacking exon 11, it is possible that such transcript could be more susceptible to RAN translation [96]. It is worth noting that, dependent on expanded CAG repeats, deregulated splicing of *HTT* exon 1 has been shown to lead to the production of short polyadenylated mRNA encoding a highly toxic HTT fragment [97, 98]. However, it remains unexplained which transcript in HD patients is indeed RAN translated: the normally processed full-length *HTT*, the aberrantly spliced small mRNA or both. Further research on RAN translation in SCA3 is required, including an analysis of the presence of RAN proteins in patient brain tissues.

It is commonly known that, in response to stress conditions, such as ER stress from misfolded proteins, oxidative stress or viral infections, the ISR pathway is activated, leading to the phosphorylation of eIF2 α and causing global cap-dependent translation attenuation but upregulation of the translation of certain mRNAs in an unconventional manner [61–64]. Recent mechanistic studies on RAN translation clearly indicate that cellular stress is another trigger for this non-canonical mode of protein synthesis. It

has been reported that RAN translation of GGGGCC and CGG repeats is enhanced by diverse stress stimuli [26, 27, 29, 36]. Our results are consistent with these data and indicate that RAN translation of *ATXN3*-associated repeat expansion sequences can also be driven by ISR. We demonstrated enhanced production of both SCA3 RAN polyQ and RAN polyA proteins upon ER stress and oxidative stress, which was accompanied by increased levels of p-eIF2 α and CHOP. In contrast to previous studies, we did not report the inhibition of canonical translation under stress conditions. The differences among studies might be related to the control of the AUG-initiated translation. In the previous studies, vectors containing nanoluciferase or firefly luciferase sequences with AUG codons were used to study canonical translation, whereas in our research, we used the construct AUG-78/32-117CAG, which, in addition to AUG codon and *ATXN3*-specific sequences, also contained the expanded CAG repeat region [26, 27, 29]. As we demonstrated that RAN polyA proteins can be produced from this construct and that the amount of these proteins significantly increased after cellular stress induction, it is very likely that, under these conditions, RAN translation in the glutamine frame was also upregulated. Thus, we assume that the amount of polyQ proteins observed for the cells expressing the AUG-78/32-117CAG construct under ER and oxidative stress was the result of reduced canonical translation and increased RAN translation. Altogether, as age-dependent accumulation of misfolded mutant ataxin-3 can lead to activation of various cellular stress pathways, it is very likely that the amount of RAN proteins increase over time and, as a consequence, the pathogenesis of SCA3 is accelerated [1–4].

Overall, we provide new insights into the SCA3 RAN translation mechanism and pathogenesis, which may constitute possible novel targets of great therapeutic value in SCA3. In addition, our findings broaden our understanding of the requirements for RAN translation in general, particularly indicating the important role of the sequences surrounding an expansion region and the possibility of initiation from non-cognate codons.

DECLARATIONS

Funding

This work was supported by the National Science Centre, Poland [2015/19/D/NZ5/02183 to M.J.-C., 2012/06/A/NZ1/00094 to W.J.K., and 2015/19/B/NZ2/02453 to A.C.]; the Polish Ministry of Science and Higher Education (statutory funds for young investigators to M.J.-C.); and the Polish Ministry of Science and Higher Education under the KNOW program [01/KNOW2/2014].

Conflicts of interests

The authors have no conflict of interest to declare.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Code availability

Not applicable.

Authors' contributions

Authors contribution: M.J.-C. designed, coordinated and performed the research. A.C. designed and performed immunoprecipitation analysis and analyzed mass spectrometry data. A.A.K. and M.O.U.-T. assisted in some of the experiments. W.J.K. and A.F. supervised the study. M.J.-C. wrote the original draft with exception of mass spectrometry results which were written by A.C.. M.J.-C., A.C. and A.F. reviewed and edited the manuscript with editorial input from A.A.K. and M.O.U.-T. All authors read and approved the final manuscript.

Corresponding authors

Correspondence to Agnieszka Fiszer and Magdalena Jazurek-Ciesiolka

REFERENCES

1. Evers MM, Toonen LJA, van Roon-Mom WMC (2014) Ataxin-3 protein and RNA toxicity in spinocerebellar ataxia type 3: Current insights and emerging therapeutic strategies. *Mol Neurobiol* 49:1513–1531. <https://doi.org/10.1007/s12035-013-8596-2>

2. Da Silva JD, Teixeira-Castro A, Maciel P (2019) From Pathogenesis to Novel Therapeutics for Spinocerebellar Ataxia Type 3: Evading Potholes on the Way to Translation. *Neurotherapeutics* 16:1009–1031. <https://doi.org/10.1007/s13311-019-00798-1>
3. Matos CA, de Almeida LP, Nóbrega C (2019) Machado–Joseph disease/spinocerebellar ataxia type 3: lessons from disease pathogenesis and clues into therapy. *J Neurochem* 148:8–28. <https://doi.org/10.1111/jnc.14541>
4. McLoughlin HS, Moore LR, Paulson HL (2020) Pathogenesis of SCA3 and implications for other polyglutamine diseases. *Neurobiol Dis* 134:104635. <https://doi.org/10.1016/j.nbd.2019.104635>
5. Nalavade R, Griesche N, Ryan DP, et al (2013) Mechanisms of RNA-induced toxicity in CAG repeat disorders. *Cell Death Dis* 4:e752. <https://doi.org/10.1038/cddis.2013.276>
6. Chan HYE (2014) RNA-mediated pathogenic mechanisms in polyglutamine diseases and amyotrophic lateral sclerosis. *Front Cell Neurosci* 8:1–12. <https://doi.org/10.3389/fncel.2014.00431>
7. Mykowska A, Sobczak K, Wojciechowska M, et al (2011) CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Res* 39:8938–8951. <https://doi.org/10.1093/nar/gkr608>
8. Urbanek MO, Jazurek M, Switonski PM, et al (2016) Nuclear speckles are detention centers for transcripts containing expanded CAG repeats. *Biochim Biophys Acta* 1862:1513–1520. <https://doi.org/10.1016/j.bbadis.2016.05.015>
9. Tsoi H, Lau TCK, Tsang SY, et al (2012) CAG expansion induces nucleolar stress in polyglutamine diseases. *PNAS* 109:13428–13433. <https://doi.org/10.1073/pnas.1204089109>
10. Zu T, Gibbens B, Doty NS, et al (2011) Non-ATG-initiated translation directed by microsatellite expansions. *PNAS* 108:260–265. <https://doi.org/10.1073/pnas.1013343108>
11. Todd PK, Oh SY, Krans A, et al (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* 78:440–455. <https://doi.org/10.1016/j.neuron.2013.03.026>

12. Ash PEA, Bieniek KF, Gendron TF, et al (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 77:639–646. <https://doi.org/10.1016/j.neuron.2013.02.004>
13. Mori K, Weng SM, Arzberger T, et al (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTL/ALS. *Science* 339:1335–1338. <https://doi.org/10.1126/science.1232927>
14. Zu T, Liu Y, Banez-Coronel M, et al (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *PNAS* 110:E4968–E4977. <https://doi.org/10.1073/pnas.1315438110>
15. Bañez-Coronel M, Ayhan F, Tarabochia AD, et al (2015) RAN translation in Huntington Disease. *Neuron* 88:667–677. <https://doi.org/10.1016/j.neuron.2015.10.038>
16. Krans A, Kearse MG, Todd PK (2016) Repeat-associated non-AUG translation from antisense CGG repeats in Fragile X Tremor/Ataxia syndrome. *Ann Neurol* 80:871–881. <https://doi.org/10.1002/ana.24800>
17. Zu T, Cleary JD, Liu Y, et al (2017) RAN translation regulated by muscleblind proteins in myotonic dystrophy type 2. *Neuron* 95:1292–1305. <https://doi.org/10.1016/j.neuron.2017.08.039>
18. Ishiguro T, Sato N, Ueyama M, et al (2017) Regulatory role of RNA chaperone TDP-43 for RNA misfolding and repeat-associated translation in SCA31. *Neuron* 94:108–124. <https://doi.org/10.1016/j.neuron.2017.02.046>
19. Soragni E, Petrosyan L, Rinkoski TA, et al (2018) Repeat-associated non-ATG (RAN) translation in Fuchs' endothelial corneal dystrophy. *Investig Ophthalmol Vis Sci* 59:1888–1896. <https://doi.org/10.1167/iovs.17-23265>
20. Wojciechowska M, Olejniczak M, Galka-Marciniak P, et al (2014) RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. *Nucleic Acids Res* 42:11849–11864. <https://doi.org/10.1093/nar/gku794>
21. Green KM, Linsalata AE, Todd PK (2016) RAN translation-What makes it run? *Brain Res*

- 1647:30–42. <https://doi.org/10.1016/j.brainres.2016.04.003>
22. Zu T, Pattamatta A, Ranum LPW (2018) Repeat-associated non-ATG translation in neurological diseases. *Cold Spring Harb Perspect Biol* 10:a033019. <https://doi.org/10.1101/cshperspect.a033019>
23. Rodriguez CM, Todd PK (2019) New pathologic mechanisms in nucleotide repeat expansion disorders. *Neurobiol Dis* 130:104515. <https://doi.org/10.1016/j.nbd.2019.104515>
24. Kearse MG, Green KM, Krans A, et al (2016) CGG repeat-associated non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. *Mol Cell* 62:314–322. <https://doi.org/10.1016/j.molcel.2016.02.034>
25. Sellier C, Buijsen RAM, He F, et al (2017) Translation of expanded CGG repeats into FMRpolyG is pathogenic and may contribute to fragile X tremor ataxia syndrome. *Neuron* 93:331–347. <https://doi.org/10.1016/j.neuron.2016.12.016>
26. Green KM, Glineburg MR, Kearse MG, et al (2017) RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat Commun* 8:2005. <https://doi.org/10.1038/s41467-017-02200-0>
27. Cheng W, Wang S, Mestre AA, et al (2018) C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2 α phosphorylation. *Nat Commun* 9:51. <https://doi.org/10.1038/s41467-017-02495-z>
28. Tabet R, Schaeffer L, Freyermuth F, et al (2018) CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat Commun* 9:152. <https://doi.org/10.1038/s41467-017-02643-5>
29. Sonobe Y, Ghadge G, Masaki K, et al (2018) Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiol Dis* 116:155–165. <https://doi.org/10.1016/j.nbd.2018.05.009>
30. Boivin M, Pfister V, Gaucherot A, et al (2020) Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders. *EMBO J* 39:e100574. <https://doi.org/10.15252/emboj.2018100574>

31. Ayhan F, Perez BA, Shorrock HK, et al (2018) SCA8 RAN polySer protein preferentially accumulates in white matter regions and is regulated by eIF3F. *EMBO J* 37:e99023. <https://doi.org/10.15252/emboj.201899023>
32. Goodman LD, Prudencio M, Srinivasan AR, et al (2019) eIF4B and eIF4H mediate GR production from expanded G4C2 in a Drosophila model for C9orf72-associated ALS. *Acta Neuropathol Commun* 7:62. <https://doi.org/10.1186/s40478-019-0711-9>
33. Linsalata AE, He F, Malik AM, et al (2019) DDX3X and specific initiation factors modulate FMR1 repeat-associated non-AUG-initiated translation. *EMBO Rep* 20:e47498. <https://doi.org/10.15252/embr.201847498>
34. Cheng W, Wang S, Zhang Z, et al (2019) CRISPR-Cas9 screens identify the RNA helicase DDX3X as a repressor of C9ORF72 (GGGGCC)_n repeat-associated non-AUG translation. *Neuron* 104:885-898.e8. <https://doi.org/10.1016/j.neuron.2019.09.003>
35. Yamada SB, Gendron TF, Niccoli T, et al (2019) RPS25 is required for efficient RAN translation of C9orf72 and other neurodegenerative disease-associated nucleotide repeats. *Nat Neurosci* 22:1383–1388. <https://doi.org/10.1038/s41593-019-0455-7>
36. Westergard T, McAvoy K, Russell K, et al (2019) Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol Med* 11:e9423. <https://doi.org/10.15252/emmm.201809423>
37. Glineburg MR, Todd PK, Charlet-Berguerand N, Sellier C (2018) Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in Fragile X Tremor Ataxia Syndrome. *Brain Res* 1693:43–54. <https://doi.org/10.1016/j.brainres.2018.02.006>
38. Banez-Coronel M, Ranum LPW (2019) Repeat-associated non-AUG (RAN) translation: insights from pathology. *Lab Investig* 99:929–942. <https://doi.org/10.1038/s41374-019-0241-x>
39. Nguyen L, Cleary JD, Ranum LPW (2019) Repeat-associated non-ATG Translation: Molecular mechanisms and contribution to neurological disease. *Annu Rev Neurosci* 42:227–247. <https://doi.org/10.1146/annurev-neuro-070918-050405>
40. Freibaum BD, Taylor JP (2017) The role of dipeptide repeats in C9ORF72-related ALS-FTD.

- Front Mol Neurosci 10:35. <https://doi.org/10.3389/fnmol.2017.00035>
41. Jiang J, Ravits J (2019) Pathogenic mechanisms and therapy development for C9orf72 amyotrophic lateral sclerosis/frontotemporal dementia. *Neurotherapeutics* 16:1115–1132. <https://doi.org/10.1007/s13311-019-00797-2>
 42. Scoles DR, Ho MHT, Dansithong W, et al (2015) Repeat Associated non-AUG translation (RAN Translation) dependent on sequence downstream of the ATXN2 CAG repeat. *PLoS One* 10:e0128769. <https://doi.org/10.1371/journal.pone.0128769>
 43. Switonski PM, Fiszler A, Kazmierska K, et al (2011) Mouse ataxin-3 functional knock-out model. *Neuromolecular Med* 13:54–65. <https://doi.org/10.1007/s12017-010-8137-3>
 44. Figura G, Koscianska E, Krzyzosiak W (2015) In vitro expansion of CAG, CAA, and mixed CAG/CAA repeats. *Int J Mol Sci* 16:18741–18751. <https://doi.org/10.3390/ijms160818741>
 45. Malinowska A, Kistowski M, Bakun M, et al (2012) Diffprot - software for non-parametric statistical analysis of differential proteomics data. *J Proteomics* 75:4062–4073. <https://doi.org/10.1016/j.jprot.2012.05.030>
 46. Nojima J, Oma Y, Futai E, et al (2009) Biochemical analysis of oligomerization of expanded polyalanine repeat proteins. *J Neurosci Res* 87:2290–2296. <https://doi.org/10.1002/jnr.22052>
 47. Polling S, Mok YF, Ramdzan YM, et al (2014) Misfolded polyglutamine, polyalanine, and superoxide dismutase 1 aggregate via distinct pathways in the cell. *J Biol Chem* 289:6669–6680. <https://doi.org/10.1074/jbc.M113.520189>
 48. Polling S, Ormsby AR, Wood RJ, et al (2015) Polyalanine expansions drive a shift into α -helical clusters without amyloid-fibril formation. *Nat Struct Mol Biol* 22:1008–1015. <https://doi.org/10.1038/nsmb.3127>
 49. Chapple JP, Bros-Facer V, Butler R, Gallo JM (2008) Focal distortion of the nuclear envelope by huntingtin aggregates revealed by lamin immunostaining. *Neurosci Lett* 447:172–174. <https://doi.org/10.1016/j.neulet.2008.09.075>
 50. Gasset-Rosa F, Chillon-Marinhas C, Goginashvili A, et al (2017) Polyglutamine-expanded huntingtin exacerbates age-related disruption of nuclear integrity and nucleocytoplasmic

- transport. *Neuron* 94:48–57. <https://doi.org/10.1016/j.neuron.2017.03.027>
51. Grima JC, Daigle JG, Arbez N, et al (2017) Mutant huntingtin disrupts the nuclear pore complex. *Neuron* 94:93–107. <https://doi.org/10.1016/j.neuron.2017.03.023>
52. Shahin V (2017) Strategic disruption of nuclear pores structure, integrity and barrier for nuclear apoptosis. *Semin Cell Dev Biol* 68:85–90. <https://doi.org/10.1016/j.semcdb.2017.07.002>
53. Peabody DS (1989) Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem* 264:5031–5035
54. Kearse MG, Wilusz JE (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.* 31:1717–1731
55. Giglione C, Boularot A, Meinell T (2004) Protein N-terminal methionine excision. *Cell Mol Life Sci* 61:1455–1474. <https://doi.org/10.1007/s00018-004-3466-8>
56. Varland S, Osberg C, Arnesen T (2015) N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics* 15:2385–2401. <https://doi.org/10.1002/pmic.201400619>
57. Polevoda B, Norbeck J, Takakura H, et al (1999) Identification and specificities of N-terminal acetyltransferases from *Saccharomyces cerevisiae*. *EMBO J* 18:6155–6168. <https://doi.org/10.1093/emboj/18.21.6155>
58. Vetting MW, de Carvalho LPS, Yu M, et al (2005) Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys* 433:212–226. <https://doi.org/10.1016/j.abb.2004.09.003>
59. Arnesen T, Van Damme P, Polevoda B, et al (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *PNAS* 106:8157–8162. <https://doi.org/10.1073/pnas.0901931106>
60. Van Damme P, Arnesen T, Gevaert K (2011) Protein alpha-N-acetylation studied by N-terminomics. *FEBS J* 278:3822–3834. <https://doi.org/10.1111/j.1742-4658.2011.08230.x>
61. Pakos-Zebrucka K, Koryga I, Mnich K, et al (2016) The integrated stress response. *EMBO Rep* 17:1374–1395. <https://doi.org/10.15252/embr.201642195>

62. Young SK, Wek RC (2016) Upstream open reading frames differentially regulate genespecific translation in the integrated stress response. *J Biol Chem* 291:16927–16935.
<https://doi.org/10.1074/jbc.R116.733899>
63. Vasudevan D, Ryoo HD (2017) Two distinct nodes of translational inhibition in the integrated stress response. *BMB Rep* 50:539–545. <https://doi.org/10.5483/BMBRep.2017.50.11.157>
64. Bond S, Lopez-Lloreda C, Gannon PJ, et al (2020) The integrated stress response and phosphorylated eukaryotic initiation factor 2 α in neurodegeneration. *J Neuropathol Exp Neurol* 79:123–143. <https://doi.org/10.1093/jnen/nlz129>
65. Bernat V, Disney MD (2015) RNA structures as mediators of neurological diseases and as drug targets. *Neuron* 87:28–46. <https://doi.org/10.1016/j.neuron.2015.06.012>
66. Ciesiolka A, Jazurek M, Drazkowska K, Krzyzosiak WJ (2017) Structural characteristics of simple RNA repeats associated with disease and their deleterious protein interactions. *Front Cell Neurosci* 11:97. <https://doi.org/10.3389/fncel.2017.00097>
67. Hale MA, Johnson NE, Berglund JA (2019) Repeat-associated RNA structure and aberrant splicing. *Biochim Biophys Acta - Gene Regul Mech* 1862:194405.
<https://doi.org/10.1016/j.bbagrm.2019.07.006>
68. Michlewski G, Krzyzosiak WJ (2004) Molecular architecture of CAG repeats in human disease related transcripts. *J Mol Biol* 340:665–679. <https://doi.org/10.1016/j.jmb.2004.05.021>
69. Oma Y, Kino Y, Sasagawa N, Ishiura S (2004) Intracellular localization of homopolymeric amino acid-containing proteins expressed in mammalian cells. *J Biol Chem* 279:21217–21222.
<https://doi.org/10.1074/jbc.M309887200>
70. Pelassa I, Corà D, Cesano F, et al (2014) Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Hum Mol Genet* 23:3402–3420. <https://doi.org/10.1093/hmg/ddu049>
71. Li L, Ng NKL, Koon AC, Chan HYE (2017) Expanded polyalanine tracts function as nuclear export signals and promote protein mislocalization via eEF1A1 factor. *J Biol Chem* 292:5784–5800. <https://doi.org/10.1074/jbc.M116.763599>

72. Sweeney P, Park H, Baumann M, et al (2017) Protein misfolding in neurodegenerative diseases: implications and strategies. *Transl Neurodegener* 6:6.
<https://doi.org/10.1186/s40035-017-0077-5>
73. Chung CG, Lee H, Lee SB (2018) Mechanisms of protein toxicity in neurodegenerative diseases. *Cell Mol Life Sci* 75:3159–3180. <https://doi.org/10.1007/s00018-018-2854-4>
74. Soto C, Pritzkow S (2018) Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nat Neurosci* 21:1332–1340. <https://doi.org/10.1038/s41593-018-0235-9>
75. Freibaum BD, Lu Y, Lopez-Gonzalez R, et al (2015) GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* 525:129–133.
<https://doi.org/10.1038/nature14974>
76. Jovičič A, Mertens J, Boeynaems S, et al (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* 18:1226–1229.
<https://doi.org/10.1038/nn.4085>
77. Boeynaems S, Bogaert E, Michiels E, et al (2016) Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Sci Rep* 6:20877.
<https://doi.org/10.1038/srep20877>
78. Zhang YJ, Gendron TF, Grima JC, et al (2016) C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat Neurosci* 19:668–677.
<https://doi.org/10.1038/nn.4272>
79. Shi KY, Mori E, Nizami ZF, et al (2017) Toxic PR n poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. *PNAS* 114:E1111–E1117.
<https://doi.org/10.1073/pnas.1620293114>
80. Zhang YJ, Guo L, Gonzales PK, et al (2019) Heterochromatin anomalies and double-stranded RNA accumulation underlie C9orf72 poly(PR) toxicity. *Science* 363:eaav2606.
<https://doi.org/10.1126/science.aav2606>
81. Almeida S, Krishnan G, Rushe M, et al (2019) Production of poly(GA) in C9ORF72 patient

- motor neurons derived from induced pluripotent stem cells. *Acta Neuropathol* 138:1099–1101.
<https://doi.org/10.1007/s00401-019-02083-z>
82. Kozak M (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *PNAS* 87:8301–8305. <https://doi.org/10.1073/pnas.87.21.8301>
 83. Kochetov A V., Palyanov A, Titov II, et al (2007) AUG_hairpin: prediction of a downstream secondary structure influencing the recognition of a translation start site. *BMC Bioinformatics* 8:318. <https://doi.org/10.1186/1471-2105-8-318>
 84. Lee S, Liu B, Lee S, et al (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *PNAS* 109:E2424-2432.
<https://doi.org/10.1073/pnas.1207846109>
 85. Liang H, Chen X, Yin Q, et al (2017) PTEN β is an alternatively translated isoform of PTEN that regulates rDNA transcription. *Nat Commun* 8:14771. <https://doi.org/10.1038/ncomms14771>
 86. Kozak M (1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *PNAS* 83:2850–2854. <https://doi.org/10.1073/pnas.83.9.2850>
 87. de Smit MH, van Duin J (1990) Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. *PNAS* 87:7668–7672.
<https://doi.org/10.1073/pnas.87.19.7668>
 88. Vega Laso MR, Zhu D, Sagliocco F, et al (1993) Inhibition of translational initiation in the yeast *Saccharomyces cerevisiae* as a function of the stability and position of hairpin structures in the mRNA leader. *J Biol Chem* 268:6453–6462
 89. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324:255–258. <https://doi.org/10.1126/science.1170160>
 90. Zydowicz-Machtel P, Swiatkowska A, Popenda L, et al (2018) Variants of the 5'-terminal region of p53 mRNA influence the ribosomal scanning and translation efficiency. *Sci Rep* 8:1533.
<https://doi.org/10.1038/s41598-018-20010-2>
 91. Bettencourt C, Santos C, Montiel R, et al (2010) Increased transcript diversity: novel splicing variants of Machado-Joseph Disease gene (ATXN3). *Neurogenetics* 11:193–202.

<https://doi.org/10.1007/s10048-009-0216-y>

92. Harris GM, Dodelzon K, Gong L, et al (2010) Splice isoforms of the polyglutamine disease protein ataxin-3 exhibit similar enzymatic yet different aggregation properties. *PLoS One* 5:e13695. <https://doi.org/10.1371/journal.pone.0013695>
93. Igarashi S, Takiyama Y, Cancel G, et al (1996) Intergenerational instability of the CAG repeat of the gene for Machado-Joseph disease (MJD1) is affected by the genotype of the normal chromosome: implications for the molecular mechanisms of the instability of the CAG repeat. *Hum Mol Genet* 5:923–932. <https://doi.org/10.1093/hmg/5.7.923>
94. Matsumura R, Takayanagi T, Murata K, et al (1996) Relationship of (CAG)_nC configuration to repeat instability of the Machado-Joseph disease gene. *Hum Genet* 98:643–645. <https://doi.org/10.1007/s004390050276>
95. Maciel P, Caspar C, Guimarães L, et al (1999) Study of three intragenic polymorphisms in the Machado-Joseph disease gene (MJD1) in relation to genetic instability of the (CAG)_n tract. *Eur J Hum Genet* 7:147–156. <https://doi.org/10.1038/sj.ejhg.5200264>
96. Ramani B, Harris GM, Huang R, et al (2015) A knockin mouse model of spinocerebellar ataxia type 3 exhibits prominent aggregate pathology and aberrant splicing of the disease gene transcript. *Hum Mol Genet* 24:1211–1224. <https://doi.org/10.1093/hmg/ddu532>
97. Sathasivam K, Neueder A, Gipson TA, et al (2013) Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *PNAS* 110:2366–2370. <https://doi.org/10.1073/pnas.1221891110>
98. Neueder A, Landles C, Ghosh R, et al (2017) The pathogenic exon 1 HTT protein is produced by incomplete splicing in Huntington's disease patients. *Sci Rep* 7:1307. <https://doi.org/10.1038/s41598-017-01510-z>

FIGURE LEGENDS

Fig. 1 RAN translation of the CAG repeats in the *ATXN3* context. **a** Diagram of the constructs developed to monitor SCA3 RAN translation (upper panel) in reference to canonical translation (lower panel). Both constructs contain 78 bp and 32 bp *ATXN3* flanking sequences (FS) upstream and downstream of

repeat region, respectively, and the indicated CAG repeat numbers. Immediately downstream of the *ATXN3*-specific sequence, Myc-, HA- and His-tags were placed in the alanine, glutamine and serine reading frames, respectively. For all the analyses presented in this figure, lysates were obtained from HEK293T cells 48 h after transfection with the indicated constructs. **b** Expression of the AUG-initiated and RAN polyQ proteins was analyzed by western blotting using an anti-HA antibody. RAN translation products are indicated by blue arrows. Aggregated AUG-initiated polyQ proteins are indicated by a red asterisk. **c** Expression of RAN polyA proteins was analyzed by western blotting using an anti-Myc antibody. Specific products are indicated by black arrows. **d** Expression of RAN polyS proteins analyzed by western blotting using an anti-His antibody. No specific signals were detected. For all the western blot analyses, the GAPDH level was used as a loading control

Fig. 2 SCA3 RAN polyQ proteins form aggregates. **a** Representative images of immunofluorescence staining for AUG-initiated and RAN polyQ proteins (green signal, anti-HA antibody) and RAN polyA proteins (red signal, anti-Myc antibody) in the HEK293T cells 48 h after transfection with the indicated constructs. DAPI (blue signal) was used to stain nuclei. Scale bar=10 μ m. **b** Quantitation results of the cells containing polyQ aggregates; 100% represents cells expressing AUG-initiated or RAN polyQ proteins. The graph bars represent the mean value $\pm$ SEM from 4 biological replicates. One-way ANOVA with Bonferroni's multiple comparisons, * $p < 0.05$, ** $p < 0.0001$

Fig. 3 SCA3 RAN polyA proteins affect the integrity of the nuclear membrane. Representative images of immunofluorescence staining for RAN polyA proteins (red signal, anti-Myc antibody), and the nuclear pore complex (NPC, green signal) in the HEK293T cells 48 h after transfection with the indicated constructs. DAPI (blue signal) was used to stain nuclei. Scale bar=10 μ m

Fig. 4 SCA3 RAN translation in the glutamine frame greatly depends on the 5' flanking sequence. **a** *ATXN3* 5' and 3' flanking sequences of the 78/32-115CAG, 45/32-110CAG, 45/17-117CAG, 33/32-119CAG, 33/17-118CAG, 24/32-116CAG, 24/17-119CAG, 18/32-119CAG and 18/17-116CAG constructs. The blue arrows indicate the length of the flanking sequence upstream and downstream of the repeats used in particular RAN translation constructs. For all the analyses presented in this figure, lysates were obtained from HEK293T cells 48 h after their transfection with the indicated constructs. **b** Expression of the RAN polyQ proteins was analyzed by western blotting using an anti-HA antibody. Aggregated RAN polyQ proteins are indicated by the red asterisk. **c** Expression of the RAN polyA

proteins was analyzed by western blotting using an anti-Myc antibody. **d** Relative expression of the RAN polyA proteins from the indicated constructs normalized to that of the 45/32-110CAG, 33/32-119CAG, 24/32-116CAG or 18/32-119CAG constructs. Specific RAN polyA products used in the quantitative analysis are indicated by black arrows. The graph bars represent the mean value $\pm$ SEM from 4 biological replicates. Two-tailed *t* test, **p*<0.005, ***p*<0.0005. For all the western blot analyses, GAPDH was used as a loading control. **e** Representative images of immunofluorescence staining for RAN polyQ proteins (green signal, anti-HA antibody) and RAN polyA proteins (red signal, anti-Myc antibody) in the HEK293T cells expressing 33/32-119CAG construct. DAPI (blue signal) was used to stain nuclei. Scale bar=20 μ m

Fig. 5 SCA3 RAN translation of the expanded CAG repeats is initiated at upstream non-cognate codons. Schematic representation of *ATXN3* ORF-derived constructs: **a** AUG-78/32-117CAG and **c** 78/32-115CAG is shown with the indicated expanded CAG tracts (red, bold) and 5' and 3' flanking sequences (black, bold). HA-tag sequence is also indicated (green, bold). The identified canonical translation-dependent AUG start codon (**a**) and non-cognate RAN translation-dependent codons (**c**) are highlighted (light blue, bold). Amino acid sequences of AUG-initiated and RAN polyQ proteins are indicated in navy blue. **b** Representative LC-MS/MS spectra of the N-terminal part of the HA-immunoprecipitated polyQ proteins (^{AC}SGTNLTSEELRK) from cells transfected with AUG-78/32-117CAG construct. **d**, **e** Representative LC-MS/MS spectra of the N-terminal part of the HA-immunoprecipitated RAN polyQ proteins (^{AC}TSEELRK; ^{AC}SEELRK) from cells transfected with 78/32-115CAG construct, indicating the RAN translation initiation sites

Fig. 6 SCA3 RAN translation of the CAG repeats is upregulated upon cellular stress. For all the analyses presented in this figure, lysates were obtained from HEK293T cells 20 h after their transfection with the indicated constructs followed by 8 h of treatment with 1 μ M thapsigargin (TG) or 50 μ M sodium arsenite (SA). **a**, **b** Western blot analysis of phospho-eIF2 α at the Ser51 site, eIF2 α and CHOP in the HEK293T cells that were either transfected or not with the indicated constructs and treated or not with TG or SA. **c**, **d** Expression of the RAN polyQ and RAN polyA proteins after treatment or not with TG was analyzed by western blotting using anti-HA and anti-Myc antibodies, respectively. **e** Relative expression of the RAN polyQ and RAN polyA proteins and AUG-initiated polyQ proteins upon ER stress induction by TG and normalized to that of the untreated cells. In Fig. S11e and f, western blot analysis of the AUG-initiated polyQ, RAN polyQ and RAN polyA proteins in the HEK293T cells transfected with the AUG-

117CAG and 45/32-110CAG constructs, and treated or not with TG is shown. The graph bars represent the mean value $\pm$ SEM from 5 biological replicates. Two-tailed *t* test, **p*<0.05, ***p*<0.0005. **f, g** Expression of the RAN polyQ and RAN polyA proteins after treatment or not with SA was analyzed by western blotting using anti-HA and anti-Myc antibodies, respectively. **h** Relative expression of the RAN polyQ and RAN polyA proteins and AUG-initiated polyQ proteins upon oxidative stress induction by SA and normalized to that of the untreated cells. In Fig. S11g and h, western blot analysis of the AUG-initiated polyQ, RAN polyQ and RAN polyA proteins in the HEK293T cells transfected with the AUG-117CAG and 45/32-110CAG constructs, and treated or not with SA. The graph bars represent the mean value $\pm$ SEM from 5 biological replicates. Two-tailed *t* test, **p*<0.05, ***p*<0.001. Specific RAN polyA products used in the quantitative analysis are indicated by black arrows. For all the western blot analyses, GAPDH was used as a loading control

Fig. 1

Fig. 2

Fig. 3

Fig. 4**a** **b** **c** **d** **e**

Fig. 5

Fig. 6