Short internal open reading frames regulate the translation of N-terminally truncated proteoforms

Raphael Fettig[1,4], Zita Gonda[1,4], Niklas Walter[1,4], Paul Sallmann[1], Christiane Thanisch[1], Markus Winter[1], Susanne Bauer[1], Lei Zhang[2], Greta Linden[2], Margarethe Litfin[1], Christelle Etard[1], Olivier Armant[3], Olalla Vázquez[2], Olivier Kassel[1,5,*]

Affiliations

[1] Karlsruhe Institute of Technology (KIT), Institute for Biological and Chemical Systems - Biological Information Processing (IBCS-BIP), Karlsruhe, Germany.

[2] Philipps-Universität Marburg, Marburg, Germany.

[3] Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSE-ENV/SERPEN/LECO, Cadarache, France.

[4] These authors contributed equally

[5] Lead contact

* Correspondence: olivier.kassel@kit.edu

Summary

We have identified an mRNA cis-regulatory element which regulates the translation of N-terminally truncated proteoforms. nTRIP6 represents a short nuclear proteoform of the cytoplasmic LIM domain protein TRIP6. We have previously reported that nTRIP6 regulates the dynamics of skeletal muscle progenitor differentiation. Here we show that nTRIP6 is generated by translation initiation at an internal AUG and that its translation is repressed by a short open reading frame (sORF) immediately upstream of the nTRIP6 AUG. In an *in vitro* model of myogenic differentiation, the translation of nTRIP6 is transiently de-repressed in a mechanistic Target of Rapamycin Complex 1-dependent manner. Mining of a published translation initiation sites dataset revealed the existence of other internal sORFs which repress the translation of N-terminally truncated proteoforms. Based on their similarity with repressing upstream sORFs in mRNA 5' leader sequences, we have named these elements internal upstream ORFs (iuORFs).

Keywords: translation; short ORFs; proteoforms; myogenesis; differentiation; mTORC1

Introduction

The number of proteins that can be expressed in a cell far exceeds the number of genes. Alternative splicing and alternative translation of mRNAs drastically increase the coding capacity of the genome by generating multiple proteoforms. How the expression of these various proteoforms is regulated in a physiological context is still poorly understood, in particular in the case of alternative translation. In some cases, the proteoforms regulate the function of the "canonical" proteins[1–3], as for example the case of truncated transcription factors (TFs) which act as dominant-negative inhibitors. However, there are also examples of proteoforms which exert a function distinct from that of their canonical counterparts. This occurs in particular when the canonical protein and its variant are located in different subcellular compartments[1–3]. The protein TRIP6[4,5] represents one such example. TRIP6 belongs to the ZYXIN family of LIM domain-containing proteins that are localized in the cytosol and enriched at sites of cell adhesion where they regulate adhesion and migration. These proteins are also present in the nucleus where they exert transcriptional co-regulator functions for various TFs[6]. In the case of TRIP6, this nuclear function is exerted by a shorter, exclusively nuclear proteoform, which we have termed nTRIP6[7–10]. We have recently reported a role for nTRIP6 in the regulation of myogenesis[11]. This process relies on resident adult stem cells, the so-called satellite cells[12]. Upon muscle damage, these cells are activated and give rise to proliferating progenitors, the myoblasts. These then differentiate into committed precursors, the myocytes, which finally fuse together to form multinucleated myofibres. We have shown in an *in vitro* model of myogenesis that nTRIP6 prevents premature differentiation of myoblasts into myocytes. This repression is required for the subsequent late differentiation and fusion of myocytes. Furthermore, the expression of nTRIP6 is transiently upregulated at the transition between proliferation and differentiation, while that of TRIP6 is not[11].

Here, we report that nTRIP6 is expressed by alternative translation initiation at an internal AUG and that a short open reading frame (ORF) within the *Trip6* coding sequence represses nTRIP6 translation. Furthermore, the mechanistic target of rapamycin complex 1 (mTORC1) transiently de-represses nTRIP6 translation during early myogenesis. Based on this prototypical example, we identified other mRNAs in which short internal ORFs regulate the translation of N-terminally truncated proteoforms. Thus, our work reveals novel translation cis-regulatory elements that we have termed internal upstream ORFs (iuORFs).

Results

nTRIP6 is generated by alternative translation at an internal AUG

A nuclear export signal (NES) within the N-terminal pre-LIM region of TRIP6 is responsible for its cytosolic localization[4,5]. We therefore predicted that the nuclear proteoform nTRIP6 lacks a functional NES. However, it should still harbor the C-terminal LIM domains as well as the interaction domains within the pre-LIM region which are required for its transcriptional co-regulator function[7,10,13]. Transfection of NIH-3T3 fibroblasts with a C-terminally tagged *Trip6* construct gave rise to two proteins with sizes corresponding to those of TRIP6 and nTRIP6 (Figure 1A,B). Deletion of the annotated translation initiation codon (AUG1) abolished the expression of the long proteoform but strongly increased the expression of the short proteoform (Figure 1A,B). This result strongly suggests that nTRIP6 arises from alternative translation initiation after leaky scanning, the process by which the scanning ribosome skips the suboptimal first initiation codon and initiates at a downstream site[14]. We identified a putative initiation codon (AUG2), in frame with AUG1 and located in the middle of the NES-encoding sequence. Mutation of AUG2 into a non-initiating codon abrogated the expression of the short proteoform (Figure 1A,B). To confirm that nTRIP6 is translated at AUG2, we designed a peptide nucleic acid (PNA) to block translation initiation[15,16] at this site (Figure S1). In an *in vitro* transcription/translation assay, the AUG2-targeting PNA inhibited the translation of the short proteoform, as compared to a mispaired control PNA (Misp), without affecting that of the long proteoform (Figure 1C). Deletion of AUG1 abolished the translation of the long proteoform and increased the translation of the short proteoform, which was inhibited by the AUG2 PNA (Figure 1C). Thus, nTRIP6 arises from translation initiation at AUG2 after leaky scanning.

Our previous observation that the levels of nTRIP6 transiently increased during C2C12 myoblast differentiation whereas those of TRIP6 did not vary[11] suggests an increased translation at AUG2. To address this question, we used the AUG2 and the control PNAs fused to a cell-penetrating

peptide[15,16] in a C2C12 differentiation assay. In the control PNA treated cells, nTRIP6 expression transiently increased during the proliferation phase while TRIP6 levels did not vary. The PNA targeting AUG2 abolished the increase in nTRIP6 expression without affecting the levels of TRIP6 (Figure 1D,E). Thus, the transient increase in nTRIP6 expression occurs via increased translation initiation at AUG2. We have previously shown that nTRIP6 represses premature myoblast differentiation, allowing proper myocyte differentiation and fusion at later stages[11]. Accordingly, treatment of C2C12 myoblasts with the AUG2 PNA accelerated the expression of *Myog* mRNA, used as an indicator of early myocytic differentiation[17] and delayed the expression of *Tnni2*, a late differentiation gene[18], as compared to untreated cells or cells treated with the control PNA (Figure 1F,G). Similarly, during the early myocytic differentiation phase (day 1 after the induction of differentiation), the AUG2 PNA reduced the number of cells expressing MYH3 (embryonic myosin heavy chain), another index of late differentiation (Fig. 1H,I). Finally, cell fusion, which started at day 2 and strongly increased at day 3 in untreated cells and in cells treated with the control PNA, was significantly inhibited in cells treated with the AUG2 PNA (Figure 1H,J). Thus, the function of nTRIP6 in the temporal regulation of myoblast differentiation and fusion relies on a transient increase in translation at AUG2.

The mTOR pathway transiently stimulates nTRIP6 translation during myogenesis

In order to identify signaling pathways involved in the upregulation of nTRIP6 translation, we screened a library of 184 unique Medaka kinases[19,20] for those altering the nTRIP6/TRIP6 ratio. This ratio was normally distributed (Figure S2A,B) and analysis of the Z scores (Figure S2C; Table S1) revealed 8 kinases which significantly increased it (Table 1), including notably mTOR. Interestingly, AKT1, CHUK and MAP3K7, other hits from the screen, are direct or indirect activators of mTORC1[21]. We thus tested the effect of mTORC1 on nTRIP6 expression. In proliferating C2C12 myoblasts, treatment with IGF-1, a potent inducer of mTORC1[21], increased the expression of nTRIP6 in a dose-dependent manner without affecting that of TRIP6 (Figure

2A,B). Furthermore, a constitutively active mutant of RHEB (RHEB-Q64L)[22], an upstream activator of mTORC1, increased the expression of nTRIP6 from the tagged construct without affecting the expression of TRIP6 (Figure 2C,D). Thus, nTRIP6 translation is stimulated by the mTORC1 pathway. We next investigated whether mTORC1 was responsible for the up-regulation of nTRIP6 translation during C2C12 cell differentiation. mTORC1 activity was elevated in the proliferation phase, reaching a maximum at day -1 (relative to the induction of differentiation), and then decreased prior to differentiation (Figure 2E,F). Treatment with the mTORC1 inhibitor Everolimus abolished the transient increase in nTRIP6 (Figure 2G,H). Thus, the transient increase in mTORC1 activity promotes the increase in nTRIP6 translation during early myogenesis.

A short internal upstream open reading frame (iuORF) represses nTRIP6 translation

The stimulatory effect of mTORC1 on nTRIP6 translation is reminiscent of the regulation of translation by short upstream ORFs (uORFs) in the 5' regulatory regions of mRNAs which typically repress translation initiation at the downstream main ORF translation initiation site (TIS)[23,24]. Indeed, apart from its role in stimulating translation initiation of a subset of mRNAs[25], mTORC1 stimulates re-initiation downstream of uORFs[26–28]. Interestingly, the expression of ATF-4, a prototypical protein whose translation is repressed by a uORF[23,24] was also transiently increased during myoblast differentiation and this increase was also inhibited by Everolimus (Figure S3). We identified a conserved short ORF in the *Trip6* mRNA coding sequence (Figure 3A; Figure S4) located immediately upstream of AUG2. Mutation of the short ORF initiation codon into a non-initiating codon in the tagged TRIP6 construct led to an increased expression of nTRIP6 (Figure 3A,B). Furthermore, expression of nTRIP6 from the AUG1-deleted construct was further increased upon mutation of the short ORF AUG (Figure 3A,B), confirming the repressive role of the short ORF for initiation at AUG2 after leaky scanning through AUG1. Thus, this short internal ORF behaves like "classical" repressive uORFs in 5' leader sequences. We therefore designated this regulatory element an internal uORF (iuORF). To investigate whether

other mRNAs also harbor iuORFs, we curated a published ribosome profiling dataset that identified over 13,000 TISs in mouse embryonic stem cells[29]. Based on stringent filtration criteria (see STRAR Methods), among the 4,994 genes analyzed we identified 20 mRNAs exhibiting a short internal ORF in another frame than the canonical ORF and in close proximity to a TIS for an N-terminally truncated proteoform. Among these short internal ORFs 5 were upstream, with the stop codon immediately upstream of the truncated proteoform TIS (Figure 3C), and 15 were overlapping, with the stop codon downstream of the truncated proteoform TIS (Figure 3D). In the case of uORFs in 5' regulatory regions, the sequence context of the initiation codon, the length of the short ORF and its distance to the downstream TIS are the most relevant features to regulate initiation at the next downstream TIS[23,24]. In the identified short internal ORFs these features were highly conserved in the human and rat mRNAs (Table S2). We validated this finding by studying the contribution of the identified TISs in *Kctd20* mRNA, which bears an upstream iuORF and in *Tfdp1* mRNA which harbors an overlapping iuORF (Figure 3E-H). Transfection of NIH-3T3 cells with a C-terminally tagged KCTD20 construct gave rise to a protein with a size corresponding to the usage of the annotated first AUG, which was indeed not expressed when this AUG was mutated into a non-initiating codon. A slightly smaller protein, presumably translated from a downstream non-canonical TIS was also detected, which was still present when AUG1 was mutated. The wild-type construct also gave rise to a short proteoform with a size corresponding to the predicted size of the postulated N-terminally truncated proteoform. Its expression was strongly increased when the first AUG was mutated, and was abolished when the AUG for the postulated N-terminally truncated proteoform (AUG2) was mutated into a non-initiating codon (Figure 3E,F). These results confirm that this short proteoform is translated from AUG2 after leaky scanning at AUG1. Furthermore, mutation of the identified upstream iuORF initiation codon into a non-initiating codon led to an increased expression of the short proteoform (Figure 3E,F). Similarly, a C-terminally tagged TFDP1 construct gave rise to a protein with a size corresponding to the usage of the annotated first AUG, which was not expressed when this AUG

was mutated into a non-initiating codon. A shorter proteoform with a size corresponding to that of the postulated N-terminally truncated proteoform was expressed at very low levels from the wild-type construct. Its expression was slightly increased from the construct with a mutated AUG1, and strongly increased when the identified overlapping iuORF initiation codon was mutated into a non-initiating codon (Figure 3G,H). Furthermore, similarly to what we observed with nTRIP6, activation of mTORC1 by transfection of RHEB-Q64L increased the expression of the short KCTD20 proteoform from the tagged construct without affecting the expression of canonical long proteoform (Figure 3I,J). However, neither the short nor the canonical long proteoforms of TFDP1 were regulated by RHEB-Q64L (Figure 3K,L).

Discussion

Using *Trip6* mRNA as a model, we have identified a novel cis-regulatory element which regulates the translation of N-terminally truncated proteoforms.

We show that nTRIP6 is translated from an internal AUG located in the middle of the NES-encoding sequence. Given that the NES is the main determinant of the cytosolic localization of TRIP6[4,5], the nuclear localization of nTRIP6[7,10] is a logical consequence of the truncated NES. Our results show that nTRIP6 translation occurs after leaky ribosome scanning at the first AUG responsible for TRIP6 translation. Leaky scanning occurs when the first initiation codon is in a suboptimal context [14]. Indeed, the *Trip6* AUG1 is immediately followed by a conserved T and not by the optimal G. Still, TRIP6 is expressed at very high levels compared to nTRIP6 in all cells tested[7–10]. This difference in the relative levels of the two proteoforms is compatible with their functions. Indeed, TRIP6 is a regulator of the cytoskeleton whereas nTRIP6 is a transcriptional regulator. These are generally expressed at significant lower levels than cytoskeletal regulatory proteins[30].

Ribosome profiling experiments have revealed that many N-terminally truncated proteoforms are generated by internal translation initiation[29,31]. Here, we show that the translation of such proteoforms can be regulated. Indeed, nTRIP6 translation is repressed by the short ORF that we have identified in the coding sequence. Given its similarities with the repressive uORFs described in the 5' regulatory regions of many mRNAs[23,24], we have termed this novel regulatory element an iuORF, for internal uORF. The first hint that this iuORF might be functional came from the observation that the sequence context of the iuORF AUG, the length of the iuORF and its distance to the nTRIP6 AUG, which are most relevant for uORFs to repress initiation at downstream AUGs[23,24], were highly conserved. We confirmed the functionality of the iuORF by showing that mutating the iuAUG into a non-initiating codon increased nTRIP6 expression. Ribosome profiling experiments have also revealed the existence of short ORFs in the coding sequences of mRNAs[29,32]. However, their function remained elusive. Our analysis of these short ORFs in a published ribosome profiling dataset[29] led to the identification of other iuORFs which regulate the translation of N-terminally truncated proteoforms. In this analysis of 4,994 mRNAs, we identified only 20 such probable iuORFs which present, as in the case of *Trip6* mRNA iuORF, a high degree of conservation of the features the most relevant for a regulatory role. Using *Kctd20* and *Tfdp1* mRNAs as examples of mRNAs bearing an upstream and an overlapping iuORF, respectively, we confirmed that iuORFs do repress the translation of N-terminally truncated proteoforms.

Finally, our work demonstrates that the translation of N-terminally truncated proteoforms can be regulated independently of the translation of the canonical ORF. Specifically, mTORC1 increased nTRIP6 translation without affecting TRIP6 levels and increased the expression of the short KCTD20 proteoform, which is also repressed by an upstream iuORF. mTORC1 has been shown to inhibit post-termination ribosome recycling after the translation of uORFs, and thereby to promote re-initiation at downstream start codons[27,33,34].Thus, mTORC1 most likely favors re-initiation at the nTRIP6 and at the short KCTD20 proteoform AUGs after translation of the

iuORF. Interestingly, the expression of the short TFDP1 proteoform was not increased by mTORC1. The most likely explanation for this difference is that the iuORF in *Tfdp1* mRNA overlaps with the TIS for the short proteoform. Therefore, re-initiation downstream of the iuORF cannot occur at this TIS. Thus, given the similarities that we observed between upstream iuORFs and uORFs, it seems very likely that mTORC1 stimulates re-initiation downstream of upstream iuORFs to increase the translation of the N-terminally truncated proteoforms. Furthermore, our results show that this mechanism operates in a physiological context. The transient increase in nTRIP6 expression during myoblast differentiation is preceded by a transient increase in mTORC1 activity and blocked by an mTORC1 inhibitor. Thus, during early myogenesis, the increase in nTRIP6 expression occurs via an mTORC1-dependent stimulation of translation initiation at AUG2, downstream of the repressing iuORF. Our observation that the expression of ATF4, a prototypical example of a protein whose translation is repressed by a uORF[23,24], is also transiently up-regulated in an mTORC1-dependent manner in the same time window shows that this mechanism, common to uORF and iuORF-regulated mRNAs operates generally during myogenesis.

In conclusion, we have characterized iuORFs as novel mRNA cis-regulatory elements which regulate the translation of N-terminally truncated proteoforms.

Acknowledgments:

The authors are grateful to Jochen Wittbrodt (Centre for Organismal Studies, Heidelberg, Germany) and Gary Davidson (Institute for Biological and Chemical Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany) for providing the Medaka kinase library. We thank Nicholas S. Foulkes and Andrew C.B. Cato for stimulating discussions and critically reading the manuscript. This work was supported in part by the German Science Foundation (grants 315384510 to O.K. and 425970020 to O.K. and O.V.).

Author contributions:

Conceptualization, O.K. and O.V.; Methodology, R.F., Z.G., C.T., M.W., M.L., O.A., O.V. and O.K.; Investigation, R.F, Z.G., N.W., P.S., C.T., M.W., S.B., M.L. and C.E.; Writing – Original Draft, O.K.; Writing – Review & Editing, C.E., O.A., O.V. and O.K.; Funding Acquisition, O.V. and O.K.; Resources, L.Z., G.L., M.L. and C.E.; Supervision, O.V. and O.K.

Declaration of interests: the authors declare no competing interests.

Figure legends

Figure 1: nTRIP6 is generated by alternative translation at an internal AUG. A) Schematic representation of the constructs used. The green box represents the Nuclear Export Signal encoding sequence. B) Lysates of NIH-3T3 fibroblasts transfected with either an empty vector (V) or the indicated HA-tagged TRIP6 constructs, together with eGFP as a transfection control were subjected to Western blotting using antibodies against the HA tag and eGFP. Results are representative of three independent experiments. C) *In vitro* transcription and translation of V5-tagged Trip6 CDS constructs with (wt) or without the first AUG (ΔAUG1), as well as of eGFP as in internal translation control, in the presence of a peptide nucleic acid (PNA) targeting the second AUG (AUG2) or a mispaired control PNA (Misp), and analysis by Western blotting using

anti-V5 and anti-eGFP antibodies. Results are representative of three independent experiments. D,E) C2C12 myoblasts were treated with either the cell-penetrating AUG2 or Misp PNAs, harvested at the indicated time point of a differentiation experiment and subjected to Western blotting using antibodies against TRIP6 / nTRIP6 and GR as a loading control. Representative blots are shown (D). The expression of TRIP6 and nTRIP6 relative to the loading control is presented as mean ± SD of three independent experiments (E). F-J) C2C12 cells treated with either the AUG2 or the Misp PNAs were subjected to a differentiation experiment. F,G) The relative levels of the *Myog* (F) and *Tnni2* (G) mRNAs were determined by reverse transcription and real-time PCR. Results are plotted relative to the expression of the *Rplp0* gene (mean ± SD of three independent experiments). H-J) Cells were fixed at the indicated day and subjected to immunofluorescence analysis using an antibody against MYH3. Nuclei were counterstained using DAPI. H) Representative images are presented (scale bar 200 µm). The percentage of MYH3 expressing mononuclear cells (I) and the fusion index (percentage of nuclei within fused myotubes) (J) are presented as mean ± SD of three independent experiments.

See also Figure S1.

Figure 2. mTORC1 mediates the increase in nTRIP6 translation during myogenesis. A,B) Lysates of C2C12 cells treated with IGF-I at the indicated concentration for 6 h were analyzed by Western blotting using antibodies against TRIP6 / nTRIP6, phosphorylated p70S6K (P-p70S6K) and GR as a loading control. Membranes were stripped and re-probed with a phosphorylation-independent p70S6K antibody. Representative Western blots are shown (A). The expression of TRIP6 and nTRIP6 relative to the loading control is presented as mean ± SD of three independent experiments (B). C,D) Lysates of C2C12 cells transfected with the HA-tagged *Trip6* construct together with either an empty vector (V) or a V5-tagged constitutively active mutant of RHEB (RHEB-Q64L) at increasing doses were subjected to Western blotting using antibodies against the HA tag, the V5 tag and GR as a loading control. (C) Representative Western blots are shown. (D) The expression of nTRIP6 relative to that of TRIP6 is presented as mean ± SD of three

independent experiments (α: P=0.032). E,F) Lysates of C2C12 myoblasts harvested at the indicated day of a differentiation experiment were subjected to Western blotting using an anti-phosphorylated p70S6 kinase antibody (P-p70S6K). Membranes were stripped and re-probed with a phosphorylation-independent p70S6K antibody. Representative blots are shown (E). The amount of P-p70S6K normalized to the total amount of p70S6K is presented as mean ± SD of three independent experiments (F). G,H) Lysates of C2C12 myoblasts treated as indicated with either solvent alone or Everolimus (Everol.; 100nM) were subjected to Western blotting using antibodies against TRIP6 / nTRIP6, P-p70S6K and GR as a loading control. Membranes were stripped and re-probed with a phosphorylation-independent p70S6K antibody. Representative blots are shown (G). The expression of TRIP6 and nTRIP6 relative to the loading control is presented as mean ± SD of three independent experiments (H). In each graph individual values are depicted by symbols, each representing an independent experiment.

See also Figure S2, S3, Table S1.

Figure 3. Short internal ORFs repress the translation on N-terminally truncated proteoforms. A) Schematic representation of the TRIP6 constructs used. The internal upstream open reading frame (iuORF) is indicated as a grey box. B) Lysates of NIH-3T3 fibroblasts transfected with either an empty vector (V) or the indicated constructs, together with eGFP as a transfection (TF) control were analyzed by Western blotting using anti-HA and anti-eGFP antibodies. Results are representative of three independent experiments. C,D) Schematic representation of other mRNAs harboring upstream (C) or overlapping (D) iuORFs. cTIS: canonical translation initiation site; iuTIS and iuSTOP: TIS and stop codon for the iuORF; nTIS: TIS for the N-terminally truncated proteoform ORF; cSTOP: canonical stop codon. The numbers indicate the positions on the mRNA. E-H) NIH-3T3 fibroblasts were transfected with either an empty vector (V) or the indicated C-terminally tagged *Kctd20* (E,F) or *Tfdp1* (G,H) CDS construct together with mCherry as a transfection (TF) control. Cell lysates were analyzed by Western blotting using anti-HA and anti-mCherry antibodies. Results are representative of three

independent experiments. I-L) Lysates of C2C12 cells transfected with either *Kctd20-HA* (I,J) or *Tfdp1-HA* (K,L) together with either an empty vector (V) or the V5-tagged constitutively active mutant RHEB-Q64L at increasing doses were subjected to Western blotting using antibodies against the HA tag, the V5 tag and GR as a loading control. (I,K) Representative Western blots are shown. The expression of the short proteoform relative to that of the long proteoform of KCTD20 (J) or TFDP1 (L) is presented as mean ± SD of three independent experiments; individual values are depicted by symbols, each representing an independent experiment.

See also Figure S4, Table S2.

Table 1. Kinases which increase the expression of nTRIP6. See also Table S2.

Kinase	nTRIP6/TRIP6	Z-score[a]	*P* value
Mark1	0.0964	3.016	2.56×10^{-3}
Map3k7	0.0896	2.578	9.94×10^{-3}
Eif2ak1	0.0880	2.471	1.35×10^{-2}
Cdc7	0.0875	2.443	1.71×10^{-2}
Chuk	0.0866	2.385	1.71×10^{-2}
mTor	0.0847	2.258	2.39×10^{-2}
Akt1	0.0832	2.163	3.05×10^{-2}
Pkn2	0.0804	1.983	4.74×10^{-2}
Vector[b]	0.0371±0.015		

[a] Z-scores and derived *P* values were determined for the entire kinase set (See STAR Methods).

[b] The ratio in cells transfected with the empty vector is presented a mean ± SD of 16 determinations.

STAR Methods

Resource availability

Lead contact:

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Olivier Kassel (Olivier.kassel@kit.edu).

Materials availability

Plasmids generated in this study will be made available on request after completion of a Material Transfer Agreement.

Experimental model and subject details

C2C12 myoblasts, NIH-3T3 fibroblasts and HEK-293 cells (all from ATCC, LGC Standards GmbH, Wesel, Germany) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS). Cell lines were authenticated by morphological examination and were routinely checked for mycoplasma contamination.

Method details

Plasmid constructs

In pcDNA3.1-TRIP6-HA the mouse *Trip6* CDS with a C-terminal HA epitope was cloned by PCR between the NheI and EcoRI sites of pcDNA3.1 (Invitrogen, Karlsruhe, Germany). The KpnI / XhoI fragment of pcDNA3.1-TRIP6-HA, lacking the first ATG, was cloned into pcDNA3.1 to obtain pcDNA3.1-TRIP6ΔAUG1-HA. pcDNA3.1-TRIP6-mAUG2-HA and pcDNA3.1-TRIP6-iuACG-HA were obtained by mutating the nTRIP6 ATG (ATG2) and the iuORF ATG (iuATG) into ACG, respectively, using the QuikChange site-directed mutagenesis kit (Agilent, Waldbronn, Germany). The KpnI / XhoI fragment of pcDNA3.1-TRIP6-iuACG-HA was cloned into pcDNA3.1 to obtain pcDNA3.1-TRIP6ΔAUG1-iuACG-HA. The *Trip6* CDS constructs C-terminally tagged with the V5 epitope pcDNA3.1-TRIP6-V5 and pcDNA3.1-TRIP6ΔAUG1-V5 were cloned using a similar strategy. The constitutively active RHEB mutant RHEB-Q64L [22] N-terminally tagged with the V5 epitope was synthesized as a DNA string (GeneArt Strings DNA Fragments, Thermo Fisher Scientific, Schwerte, Germany) and cloned into pcDNA3.1 to obtain pcDNA3.1-V5-RHEB-Q64L. C-terminally HA-tagged KCTD20 and TFDP1 constructs were cloned as synthesized DNA strings into pcDNA3.1. The QuikChange

site-directed mutagenesis kit was used to mutate the canonical initiation codon (AUG1), the initiation codon for the N-terminally truncated proteoform (AUG2) and the initiation codon for the short internal ORF (iuAUG) into non-initiating codons (ACG) to obtain pcDNA3.1-KCTD20-mAUG1-HA, pcDNA3.1-KCTD20-mAUG2-HA, pcDNA3.1-KCTD20-miuAUG-HA, pcDNA3.1-TFDP1-mAUG1-HA, pcDNA3.1-TFDP1-mAUG2-HA and pcDNA3.1-TFDP1-miuAUG-HA. pcDNA3.1-eGFP and pcDNA3.1-mCherry were obtained by cloning the corresponding string DNA fragments (ThermoFisher Scientific) into pcDNA3.1. All constructs were verified by sequencing.

Cell transfection and cellular assays

We used a standardized protocol for the differentiation of C2C12 myoblasts[11]. Cells were seeded at a density of $5x10^3$ cells / cm^2 in growth medium (GM, 10% FCS-containing DMEM) at day -3, relative to the induction of differentiation at day 0. When cells reached at day 0, differentiation was induced by changing the medium to differentiation medium (DM, 2% horse serum-containing DMEM), which was then replaced every second day. Where indicated, cells were treated with the mTORC1 inhibitor Everolimus (Sigma-Aldrich) at a concentration of 100 nM. NIH-3T3 fibroblasts and C2C12 myoblasts were transfected using *Trans*IT-X2 (Mirus Bio LLC) and HEK-293 cells with ScreenFect-A (ScreenFect, Eggenstein-Leopoldshafen, Germany). A PNA was designed to target the mouse *Trip6* mRNA sequence (NM_011639.3) from positions 481 to 498 which encompass AUG2: NH_2-GTCCAGATCAGCCAACAT-R_8-$CONH_2$. A mispaired PNA (Misp) was used as a control: NH_2-<u>C</u>TCCA<u>C</u>ATCA<u>C</u>CCAAGAT-R_8-$CONH_2$; the mispaired bases are underlined. The cell-penetrating moiety of the PNAs consists of an octoarginine peptide (R8). The PNAs were synthesized manually by standard Fmoc-based PNA solid-phase protocols as described [35] (TentaGel® S RAM resin; 10 µmol scale; after the 12th PNA-monomer coupling, the used equivalents increased from 4 to 8 for 1h). Peptide-PNA conjugates were purified by preparative reverse-phase (RP) HPLC and identified by high-resolution mass spectrometry (Figure S1). To increase the penetration, the PNAs were applied to C2C12 cells at a concentration

of 10 µM in DMEM with reduced FCS (2.5%) for 2h, after which FCS concentration was increased to 10%.

For the kinase over-expression screen, a library of 184 unique Medaka kinases cloned into pCMV-Sport6.1 [19,20] was kindly provided by Jochen Wittbrodt (Centre for Organismal Studies, Heidelberg, Germany) and Gary Davidson (Institute for Biological and Chemical Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany). HEK-293 cells were co-transfected in 96-well plates with the V5-tagged TRIP6 construct and either an empty vector as a control or the library using ScreenFect-A. Cells were lysed 24h later, and the relative expression of TRIP6 and nTRIP6 was assessed by Western Blot analysis (see below).

RNA isolation and quantitative Real-time PCR (qRT-PCR)

Total RNA was extracted using PeqGOLD TriFast™ (Peqlab Biotechnologie, Erlangen, Germany) and reverse-transcribed into cDNA. *Myog* (myogenin) and *Tnni2* mRNAs, as well as the transcript of the large ribosomal subunit P0 gene (*Rplp0*) used for normalization, were quantified by real-time PCR using the ABI Prism Sequence Detection System 7000 (Applied Biosystems, Foster City, CA). The primers (ThermoFisher Scientific) are described in the Key resources table.

Western Blotting

Western blotting analyses were performed using the following antibodies: a custom-made rabbit anti-TRIP6 monoclonal antibody [10]; anti-V5 (R960-25, Thermo Fisher Scientific); anti-HA (clone 3F10, Roche Applied Science, Mannheim, Germany); anti-GFP (clone D5.1, Cell Signaling Technology); anti-mCherry (Ab167453, Abcam); anti-ATF4 (11815, Cell Signaling Technology); anti-phospho (Thr389)-p70 S6 Kinase (clone 108D2, Cell Signaling Technology). To normalize the phospho (Thr389)-p70 S6 Kinase signals membranes were stripped and reprobed with a

phosphorylation-independent anti-p70 S6 Kinase antibody (Cell Signaling Technology). The expression of the "classical" housekeeping genes β-actin and GAPDH is regulated during myogenesis[36–38]. Therefore, we selected the glucocorticoid receptor (GR) as a loading control, given that it is stably expressed during C2C12 differentiation[39]. To this end we used an anti-GR antibody (sc-393232, Santa Cruz, Heidelberg, Germany). Signals were detected by enhanced chemoluminescence using the ChemiDoc Touch Imaging System (BioRad laboratories, Munich, Germany). Signal quantification was performed within the linear range of detection using the Image Lab software (Bio-Rad laboratories). Linear brightness and contrast adjustments were made for illustration purposes only after the analysis had been made.

In vitro transcription and translation

The plasmids pcDNA3.1-TRIP6-V5 and pcDNA3.1-TRIP6ΔAUG1-V5, as well as pcDNA3.1-eGFP as a control were used as templates for *in vitro* transcription and translation using the T7 TNT® Coupled Reticulocyte Lysate System (Promega, Mannheim, Germany) according to the manufacturer's instructions, except that the reactions were performed for 90 min at 30°C in the presence of 5 µM PNA targeting the 2nd AUG of *Trip6* mRNA or its mispaired control. The translated products were analyzed by Western Blotting.

In silico identification of iuORFs

In order to identify other mRNAs harboring short internal ORFs that putatively regulate the translation of N-terminally truncated proteoforms, we curated a published ribosome profiling dataset [29]. Out of the 4,994 mRNAs analyzed, we first selected those in which the canonical TIS was identified and that harbored at least one internal (3' to the canonical TIS) out of frame, short (≤ 100 codons) ORF and at least one in frame internal TIS for a truncated proteoform. The resulting mRNAs were then filtered to keep only those in which the TIS of the short ORF was

located 5' to the truncated proteoform TIS, and in which the stop codon of the short ORF was located either 3' to the truncated proteoform TIS (overlapping short ORF) or at a maximum distance of 30 nucleotide 5' to the truncated proteoform TIS (upstream short ORF). When several short ORFs fulfilled these criteria within the same mRNA, only the one with the highest initiation score [29] was kept. We then performed a conservation analysis of the features most relevant for the translation of N-terminally truncated proteoforms and for short ORFs to regulate initiation at downstream AUGs [23,24], i.e. the sequence contexts of the canonical, short ORF and truncated proteoform initiation codons relative to the Kozak sequence [14], the distance from the canonical TIS to the short ORF TIS, the distance from the short ORF stop codon to the truncated proteoform TIS and the length of the short ORF. Human and rat orthologues to mouse genes were selected using the closest homologue in the phylogenetic tree using clustalw. Conservation of the start codons and length of the ORF were obtained by screening the multiple alignments with the EMBOSS [41] functions extractalign and transeq.

Immunofluorescence, microscopy and image analysis

Immunofluorescence analysis was performed on C2C12 cells grown and differentiated on glass coverslips coated with collagen Type I (Sigma-Aldrich), fixed for 10 min in 10% formalin, permeabilized for 10 min in 0.5% Triton X-100 in PBS and blocked for 1h in 5% BSA in PBS. The primary antibody was a mouse anti-MYH3 (F1.652-b, Developmental Studies Hybridoma Bank, deposited by Blau H.M.) and the secondary antibody was an Alexa Fluor 488-conjugated anti-mouse antibody (Invitrogen). Nuclei were counter-stained with DAPI (Thermo Fisher Scientific), Cells were imaged by confocal microscopy on a Zeiss LSM 800 (Zeiss, Jena, Germany). Cells images were acquired in tiling mode using a 10x/0.3 Plan-Neofluar objective resulting in 3 X 2 mm^2 images. Images were analyzed using Fiji [42]. The number of MYH3 positive mononuclear cells was determined by a combination of automated segmentation and manual counting. The fusion index was calculated as the percentage of nuclei within fused

myotubes. Linear brightness and contrast adjustments were made for illustration purposes, but only after the analysis had been made.

Statistical analysis

Statistical analyses were performed using R. Where indicated, significant differences were assessed by two-sided t-test analysis, with values of $P < 0.05$ sufficient to reject the null hypothesis. A Bonferroni correction was applied when multiple comparisons were performed. For the kinase overexpression screen, the normal distribution of the nTRIP6/TRIP6 ratio was first confirmed by a Shapiro-Wilk normality test ($W = 0.9803$, $P = 1.07 \times 10^{-2}$) as well as by the Kernel density and Q-Q plots (Figure S2). This allowed us to calculate Z scores and to derive 2-sided P values in order to determine the "hits", kinases which significantly altered the nTRIP6/TRIP6 ratio.

Data and code availability

All data supporting the findings of this study are available within the paper and are available from the corresponding author upon request.

This paper does not report original code.

Supplemental information

Document S1. Figures S1–S4

Table S1. Effect of overexpressed kinases on the relative expression of nTRIP6 and TRIP6. Related to Table 1 and Figure 2

Table S2. Conservation of the short internal ORFs and translation initiation sites for N-terminally truncated proteoforms. Related to Figure 3.

References

1. Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. Gene *514*, 1–30. 10.1016/j.gene.2012.07.083.

2. Orre, L.M., Vesterlund, M., Pan, Y., Arslan, T., Zhu, Y., Woodbridge, A.F., Frings, O., Fredlund, E., and Lehtiö, J. (2019). SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization. Mol Cell *73*, 166-182.e7. 10.1016/j.molcel.2018.11.035.

3. Bogaert, A., Fernandez, E., and Gevaert, K. (2020). N-Terminal Proteoforms in Human Disease. Trends Biochem Sci *45*, 308–320. 10.1016/j.tibs.2019.12.009.

4. Lin, V.T.G., and Lin, F.-T. (2011). TRIP6: An adaptor protein that regulates cell motility, antiapoptotic signaling and transcriptional activity. Cell Signal *23*, 1691–1697. 10.1016/j.cellsig.2011.06.004.

5. Willier, S., Butt, E., Richter, G.H.S., Burdach, S., and Grunewald, T.G.P. (2011). Defining the role of TRIP6 in cell physiology and cancer. Biol Cell *103*, 573–591. 10.1042/bc20110077.

6. Hervy, M., Hoffman, L., and Beckerle, M.C. (2006). From the membrane to the nucleus and back again: bifunctional focal adhesion proteins. Curr Opin Cell Biol *18*, 524–532. 10.1016/j.ceb.2006.08.006.

7. Kassel, O., Schneider, S., Heilbock, C., Litfin, M., Göttlicher, M., and Herrlich, P. (2004). A nuclear isoform of the focal adhesion LIM-domain protein Trip6 integrates activating and repressing signals at AP-1- and NF- B-regulated promoters. Gene Dev *18*, 2518–2528. 10.1101/gad.322404.

8. Diefenbacher, M., Sekula, S., Heilbock, C., Maier, J.V., Litfin, M., Dam, H. van, Castellazzi, M., Herrlich, P., and Kassel, O. (2008). Restriction to Fos Family Members of Trip6-Dependent Coactivation and Glucocorticoid Receptor-Dependent Trans-Repression of Activator Protein-1. Mol Endocrinol *22*, 1767–1780. 10.1210/me.2007-0574.

9. Diefenbacher, M.E., Litfin, M., Herrlich, P., and Kassel, O. (2010). The nuclear isoform of the LIM domain protein Trip6 integrates activating and repressing signals at the promoter-bound glucocorticoid receptor. Mol Cell Endocrinol *320*, 58–66. 10.1016/j.mce.2010.02.010.

10. Kemler, D., Dahley, O., Roßwag, S., Litfin, M., and Kassel, O. (2016). The LIM domain protein nTRIP6 acts as a co-repressor for the transcription factor MEF2C in myoblasts. Sci Rep *6*, 27746. 10.1038/srep27746.

11. Abbariki, T.N., Gonda, Z., Kemler, D., Urbanek, P., Wagner, T., Litfin, M., Wang, Z.-Q., Herrlich, P., and Kassel, O. (2021). The LIM domain protein nTRIP6 modulates the dynamics of myogenic differentiation. Sci Rep *11*, 12904. 10.1038/s41598-021-92331-8.

12. Schmidt, M., Schüler, S.C., Hüttner, S.S., Eyss, B. von, and Maltzahn, J. von (2019). Adult stem cells at work: regenerating skeletal muscle. Cell Mol Life Sci *76*, 2559–2570. 10.1007/s00018-019-03093-6.

13. Diefenbacher, M.E., Reich, D., Dahley, O., Kemler, D., Litfin, M., Herrlich, P., and Kassel, O. (2014). The LIM Domain Protein nTRIP6 Recruits the Mediator Complex to AP-1-Regulated Promoters. Plos One *9*, e97549. 10.1371/journal.pone.0097549.

14. Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. Gene *299*, 1–34. 10.1016/s0378-1119(02)01056-9.

15. Marin, V.L., Roy, S., and Armitage, B.A. (2004). Recent advances in the development of peptide nucleic acid as a gene-targeted drug. Expert Opin Biol Th *4*, 337–348. 10.1517/14712598.4.3.337.

16. Gupta, A., Mishra, A., and Puri, N. (2017). Peptide nucleic acids: Advanced tools for biomedical applications. J Biotechnol *259*, 148–159. 10.1016/j.jbiotec.2017.07.026.

17. Edmondson, D.G., and Olson, E.N. (1989). A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Gene Dev *3*, 628–640. 10.1101/gad.3.5.628.

18. Lin, Z., Lu, M.-H., Schultheiss, T., Choi, J., Holtzer, S., Dilullo, C., Fischman, D.A., and Holtzer, H. (1994). Sequential appearance of muscle-specific proteins in myoblasts as a function of time after cell division: Evidence for a conserved myoblast differentiation program in skeletal muscle. Cell Motil Cytoskel *29*, 1–19. 10.1002/cm.970290102.

19. Chen, Q., Su, Y., Wesslowski, J., Hagemann, A.I., Ramialison, M., Wittbrodt, J., Scholpp, S., and Davidson, G. (2014). Tyrosine phosphorylation of LRP6 by Src and Fer inhibits Wnt/β-catenin signalling. EMBO Rep *15*, 1254–1267. 10.15252/embr.201439644.

20. Souren, M., Martinez-Morales, J., Makri, P., Wittbrodt, B., and Wittbrodt, J. (2009). A global survey identifies novel upstream components of the Ath5 neurogenic network. Genome Biol *10*, R92. 10.1186/gb-2009-10-9-r92.

21. Saxton, R.A., and Sabatini, D.M. (2017). mTOR Signaling in Growth, Metabolism, and Disease. Cell *168*, 960–976. 10.1016/j.cell.2017.02.004.

22. Jiang, H., and Vogt, P.K. (2008). Constitutively active Rheb induces oncogenic transformation. Oncogene *27*, 5729–5740. 10.1038/onc.2008.180.

23. Chen, H.-H., and Tarn, W.-Y. (2019). uORF-mediated translational control: recently elucidated mechanisms and implications in cancer. RNA Biol *16*, 1–12. 10.1080/15476286.2019.1632634.

24. Zhang, H., Wang, Y., and Lu, J. (2019). Function and Evolution of Upstream ORFs in Eukaryotes. Trends Biochem Sci *44*, 782–794. 10.1016/j.tibs.2019.03.002.

25. Masvidal, L., Hulea, L., Furic, L., Topisirovic, I., and Larsson, O. (2017). mTOR-sensitive translation: Cleared fog reveals more trees. RNA Biol *14*, 1299–1305. 10.1080/15476286.2017.1290041.

26. Calkhoven, C.F., Müller, C., and Leutz, A. (2000). Translational control of C/EBPalpha and C/EBPbeta isoform expression. Gene Dev *14*, 1920–1932.

27. Zidek, L.M., Ackermann, T., Hartleben, G., Eichwald, S., Kortman, G., Kiehntopf, M., Leutz, A., Sonenberg, N., Wang, Z.-Q., Maltzahn, J. von, et al. (2015). Deficiency in mTORC1-controlled C/EBPβ-mRNA translation improves metabolic health in mice. EMBO Rep *16*, 1022–1036. 10.15252/embr.201439837.

28. Chen, Y.-J., Tan, B.C.-M., Cheng, Y.-Y., Chen, J.-S., and Lee, S.-C. (2010). Differential regulation of CHOP translation by phosphorylated eIF4E under stress conditions. Nucleic Acids Res *38*, 764–777. 10.1093/nar/gkp1034.

29. Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell *147*, 789–802. 10.1016/j.cell.2011.10.002.

30. Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011). The quantitative proteome of a human cell line. Mol Syst Biol *7*, 549. 10.1038/msb.2011.82.

31. Damme, P.V., Gawron, D., Criekinge, W.V., and Menschaert, G. (2014). N-terminal Proteomics and Ribosome Profiling Provide a Comprehensive View of the Alternative Translation Initiation Landscape in Mice and Men. Mol Cell Proteomics *13*, 1245–1261. 10.1074/mcp.m113.036442.

32. Bazzini, A.A., Johnstone, T.G., Christiano, R., Mackowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., et al. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J *33*, 981–993. 10.1002/embj.201488411.

33. Schepetilnikov, M., Dimitrova, M., Mancera-Martínez, E., Geldreich, A., Keller, M., and Ryabova, L.A. (2013). TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3h. EMBO J *32*, 1087–1102. 10.1038/emboj.2013.61.

34. Gunišová, S., Hronová, V., Mohammad, M.P., Hinnebusch, A.G., and Valášek, L.S. (2018). Please do not recycle! Translation reinitiation in microbes and higher eukaryotes. FEMS Microbiol Rev *42*, 165–192. 10.1093/femsre/fux059.

35. Vázquez, O., and Seitz, O. (2014). Cytotoxic peptide–PNA conjugates obtained by RNA-programmed peptidyl transfer with turnover. Chem Sci *5*, 2850–2854. 10.1039/c4sc00299g.

36. Hildyard, J.C.W., and Wells, D.J. (2014). Identification and Validation of Quantitative PCR Reference Genes Suitable for Normalizing Expression in Normal and Dystrophic Cell Culture Models of Myogenesis. PLoS Curr. *6*, ecurrents.md.faafdde4bea8df4aa7d06cd5553119a6. 10.1371/currents.md.faafdde4bea8df4aa7d06cd5553119a6.

37. Otey, C.A., Kalnoski, M.H., and Bulinski, J.C. (1988). Immunolocalization of muscle and nonmuscle isoforms of actin in myogenic cells and adult skeletal muscle. Cell Motil. Cytoskelet. *9*, 337–348. 10.1002/cm.970090406.

38. Cox, R.D., Garner, I., and Buckingham, M.E. (1990). Transcriptional regulation of actin and myosin genes during differentiation of a mouse muscle cell line. Differentiation *43*, 183–191. 10.1111/j.1432-0436.1990.tb00445.x.

39. Sun, L., Trausch-Azar, J.S., Muglia, L.J., and Schwartz, A.L. (2008). Glucocorticoids differentially regulate degradation of MyoD and Id1 by N-terminal ubiquitination to promote muscle protein catabolism. Proc Natl Acad Sci *105*, 3339–3344. 10.1073/pnas.0800165105.

40. Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc *7*, 1534–1550. 10.1038/nprot.2012.086.

41. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet *16*, 276–277. 10.1016/s0168-9525(00)02024-2.

42. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676–682. 10.1038/nmeth.2019.

Figure 1: nTRIP6 is generated by alternative translation at an internal AUG

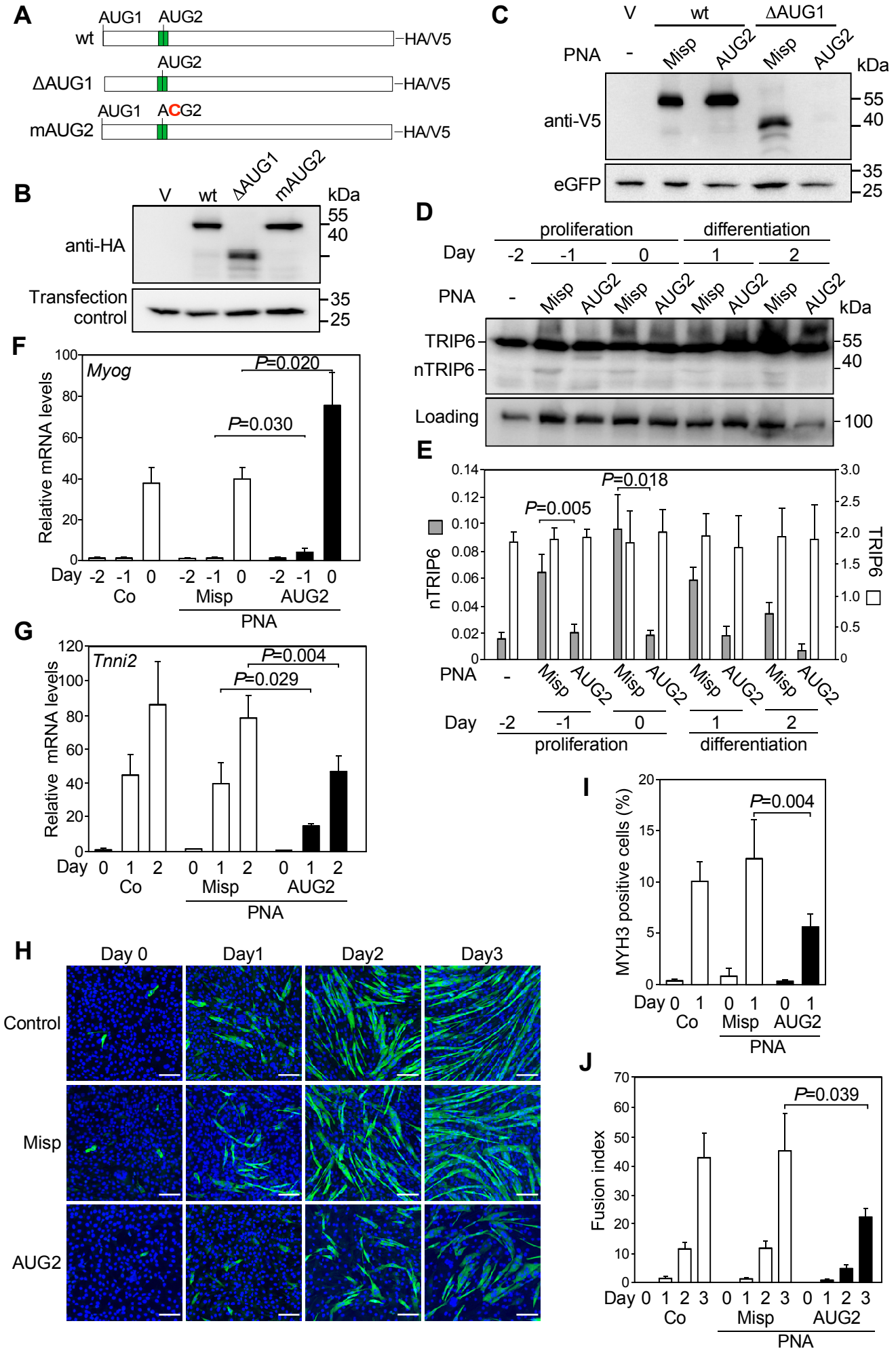

Figure 2: mTORC1 mediates the increase in nTRIP6 translation during myogenesis

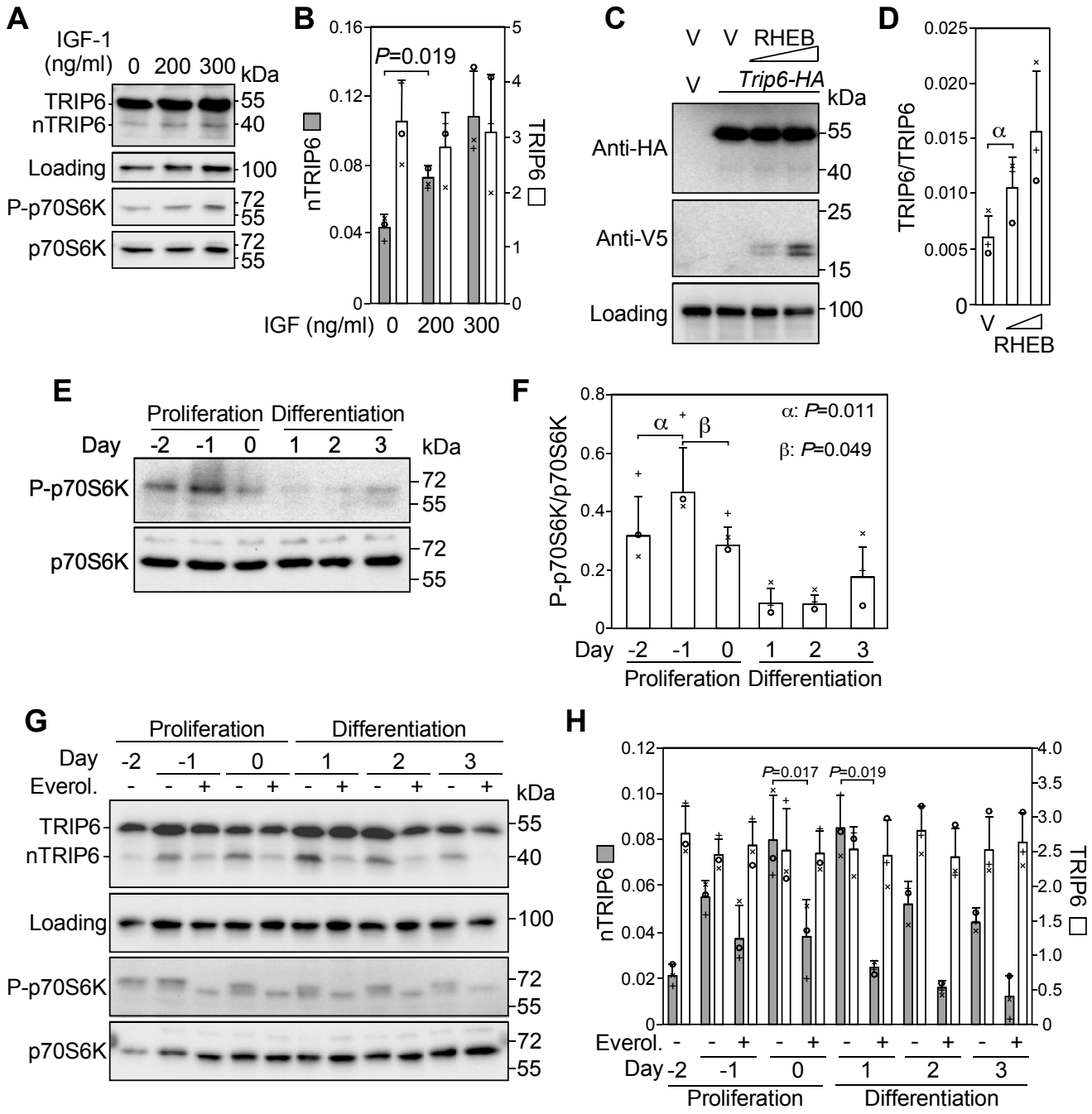
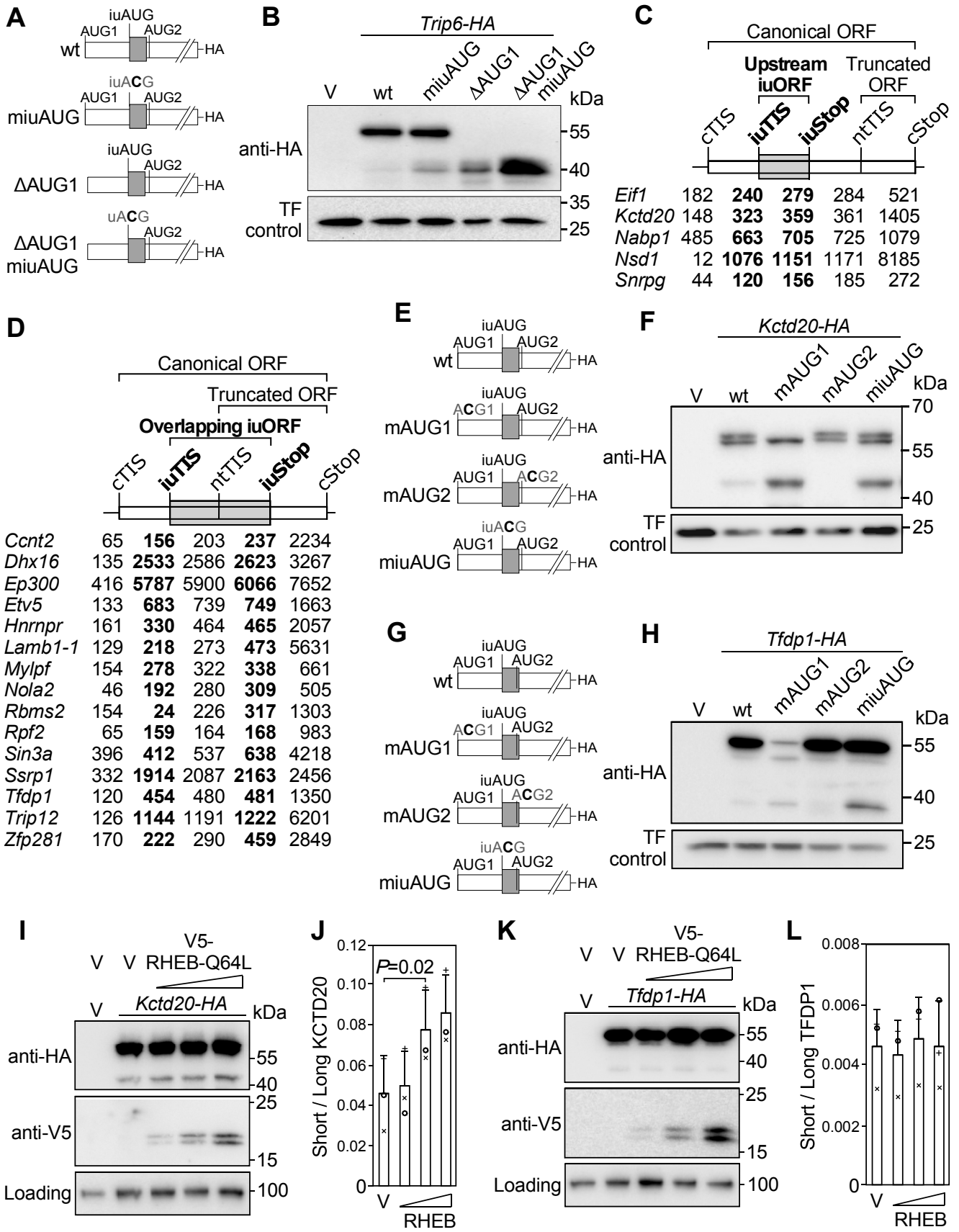
Figure 3. Short internal ORFs repress the translation on N-terminally truncated proteoforms