

REVIEW

Open Access

Non-AUG translation initiation in mammals



Dmitry E. Andreev^{1,2*}, Gary Loughran³, Alla D. Fedorova³, Maria S. Mikhaylova², Ivan N. Shatsky² and Pavel V. Baranov^{3*} 

*Correspondence:
cycloheximide@yandex.ru;
p.baranov@ucc.ie

¹ Shemyakin-Ovchinnikov
Institute of Bioorganic
Chemistry, RAS,
Moscow 117997, Russia

³ School of Biochemistry
and Cell Biology, University
College Cork, Cork T12 XF62,
Ireland

Full list of author information
is available at the end of the
article

Abstract

Recent proteogenomic studies revealed extensive translation outside of annotated protein coding regions, such as non-coding RNAs and untranslated regions of mRNAs. This non-canonical translation is largely due to start codon plurality within the same RNA. This plurality is often due to the failure of some scanning ribosomes to recognize potential start codons leading to initiation downstream—a process termed leaky scanning. Codons other than AUG (non-AUG) are particularly leaky due to their inefficiency. Here we discuss our current understanding of non-AUG initiation. We argue for a near-ubiquitous role of non-AUG initiation in shaping the dynamic composition of mammalian proteomes.

Introduction

The initiation of translation of most eukaryotic mRNAs occurs via the so-called scanning mechanism which was championed by Marylin Kozak [1, 2]. The start codon, usually AUG, is recognized by the preinitiation complex (PIC) consisting of a 40S ribosomal subunit loaded with eukaryotic initiation factors (eIFs) and Met-tRNA_i. The PIC first enters at the 5' end of mRNA via recognition of m⁷G cap and then scans the 5' “untranslated” region (leader) in the 5'-to-3' direction. Recognition of a suitable start codon is followed by dissociation of initiation factors and 60S subunit joining (reviewed in [3–5]).

At a mechanistic level, start codon recognition depends on codon-anticodon interaction between AUG and Met-tRNA_i. During the elongation phase, all codons are decoded in the A-site; however, the start codon is recognized in the ribosome P-site. Base pairing in the P-site is not monitored with the same strictness as in the A-site where the Watson-Crick geometry of the base pairs involving the first two subcodon positions is strictly enforced by the ribosome [6, 7].

Presumably, the reduced control of the base-pairing geometry in the P-site enables Met-tRNA_i to recognize codons other than AUG as starts, albeit less efficiently. In addition to the identity of the potential start codon, the likelihood of translation initiation is influenced by the surrounding nucleotide context which has been extensively investigated. Kozak reported that the 6 nt preceding the initiation codon and the



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

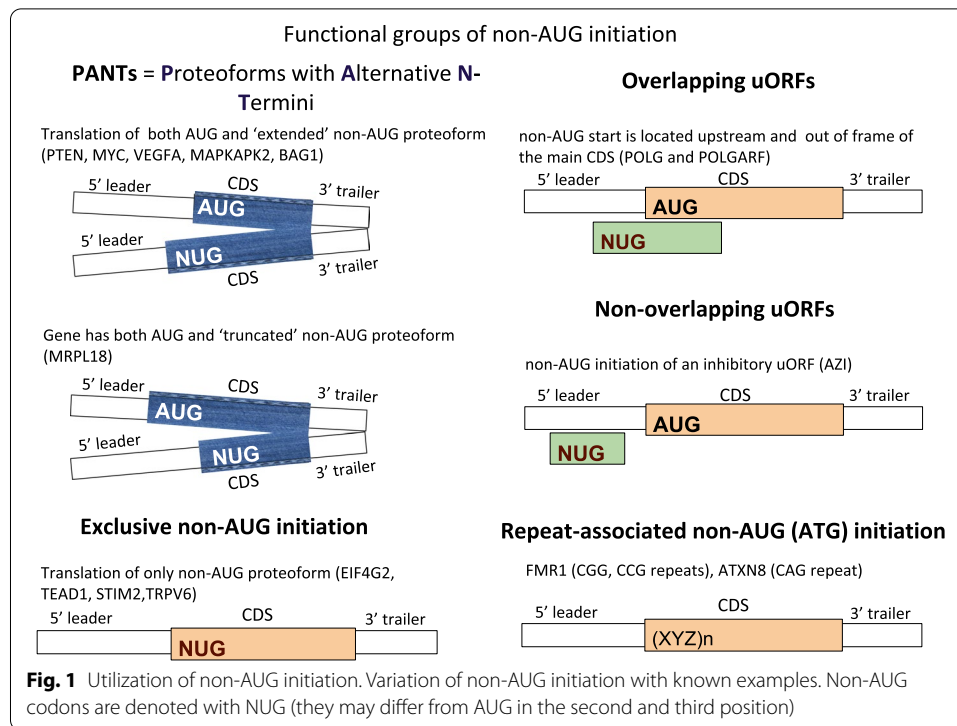
nucleotide immediately downstream influence translation initiation efficiency, with GCCRCCAUGG (R = A or G) being the consensus sequence in vertebrate mRNAs [8–10]. Specifically, positions –3 and +4 (where +1 refers to A in AUG) were found to be the most important for efficient AUG recognition, with A/G in –3 and G in +4 being optimal. All possible context efficiencies for AUG initiation from –6 to +5 have been recently characterized quantitatively and this analysis finds the strongest context to be RYMRMVAUGGC (Y = C or U; M = A or C; V = G, C or A) [11].

Not all mRNAs possess optimal context AUGs at the start of their coding sequence (CDS). Noderer et al. [11] applied a quantitative estimation of all protein coding AUG initiation efficiencies across the human transcriptome and found that although the distribution is shifted toward more efficient translation initiation starts (TISs), thousands of mRNAs possess non-optimal and “weak” context starts. Some of the weakest context initiation codons are highly conserved [12–14]. The difference in initiation efficiencies between optimal and weak naturally occurring AUG contexts is estimated to be up to 12-fold [11]. Suboptimal AUG TISs facilitate leaky scanning and it has been shown that in such cases the next downstream start tends to be located in the same reading frame, thus allowing translation of additional “truncated” proteoforms [15].

More than half of all human mRNAs have at least one AUG codon upstream (uAUG) of their annotated TIS (we searched current transcript annotations and found that 58% in the current versions of RefSeq (annotation release 109.20210226) and 56% in Gencode (version 37)). Their potential use as TISs could result in translation of so-called upstream Open Reading Frames (uORFs). uORF translation usually results in the synthesis of short polypeptides, some of which have been shown to be functional, e.g., in *ASNSD1* [16], *MIEF1* [17–20], *MKKS* [21], and *SLC35A4* [17]. However, it is believed that most translated uORFs only have a mild inhibitory effect on downstream translation [22]. The average inhibitory effect of uORFs is mild either because most uORF starts are leaky (non-AUG or AUG in a weak context), or also because ribosomes terminating after translation of short ORFs are often capable of reinitiating [23].

Nonetheless, translation initiation on most mRNAs likely occurs at more than one AUG. Furthermore, modified ribosome profiling techniques that enrich ribosome footprints at TISs revealed that non-AUG TISs are even more abundant than AUG TISs [24–26], suggesting that TIS plurality is pervasive in mammalian mRNAs. Furthermore, the stringency of start codon selection is plastic and can be regulated by the cellular levels of specific factors [12–14, 27] leading to altered ratios of AUG vs non-AUG initiation which subsequently varies across different (patho)physiological conditions. Despite this growing appreciation of the role of non-AUG initiation in shaping mammalian proteomes, its exploration is still in its infancy. In this review, we provide a glimpse into recent findings and argue for the need for increased awareness of this phenomenon in genome annotations and gene expression studies.

While non-AUG initiation is known to take place in prokaryotes, as well as in lower eukaryotes such as yeast and in plants, this is outside the scope of this review and is discussed elsewhere [28–30]. We mention relevant studies in other eukaryotes only when they provide important mechanistic insights in a historical context.



Large proteins whose synthesis initiates at non-AUG start codons

Depending on its location within a particular mRNA, alternative TISs yield either translation of an alternative ORF or, if it is located in the same reading frame as the main AUG-initiated ORF, it can yield N-terminally extended or truncated proteoforms (Proteoforms with Alternative N Termini, PANTs) (Fig. 1).

One example of non-AUG initiation on a mammalian mRNA was demonstrated for the oncogenic transcription factor *MYC* (a.k.a. c-Myc). Hann et al. [31] detected two c-Myc proteoforms that originate from a single mRNA with a single AUG-initiated ORF. It was subsequently shown that N-terminally extended Myc (p67, myc1) starts from a CUG codon located 15 triplets upstream of the main ORF AUG. The N-terminally extended proteoform of c-Myc differentially regulates transcription through a non-canonical DNA-binding site and thus is functionally distinct from the AUG-initiated proteoform, the latter appears to be more favorable for a phenotype supportive of cancer progression than the CUG-initiated proteoform [32, 33]. Intriguingly, among the few known instances of non-AUG initiated proteoforms, several are expressed from genes implicated in cancer. These include *VEGFA* encoding vascular endothelial growth factors whose extended proteoform is secreted [34], *BAG1* encoding Bcl-2-associated athanogene-1 [35–40], and *FGF2* whose CUG-initiated proteoforms are targeted to the nucleus [41–43].

Ivanov et al. [44] reasoned that the functionality of non-AUG extended proteoforms should constrain their evolution. Searches for sequences upstream of annotated AUG TISs that are evolutionary conserved identified 42 novel non-AUG initiated proteoforms including a CUG-initiated proteoform of the *PTEN* gene whose product is extended by 173 residues. *PTEN* is a tumor suppressor and an antagonist of the phosphoinositide-3

kinase (PI3K) pathway. It was later reported that the extended proteoform is secreted from cells and can enter other cells to alter PI3K signaling [45]. A subsequent study that focused on identifying the exact TIS revealed several non-AUG TISs in addition to CUG with an AUU being the most efficient [46]. This study also cast doubt on the secretory potential of any N-terminally extended PTEN proteoforms. While such phylogenetic approaches can successfully identify functionally important extensions, several known instances of reported non-AUG TISs are poorly conserved. However, poor conservation does not preclude their phenotypic relevance. For example, contrary to expectations, one of the most extensively studied examples of non-AUG initiation, the above mentioned CUG codon in *MYC*, is absent in several mammals (e.g., substituted to CUA codon in cow, sheep, and some other ungulates).

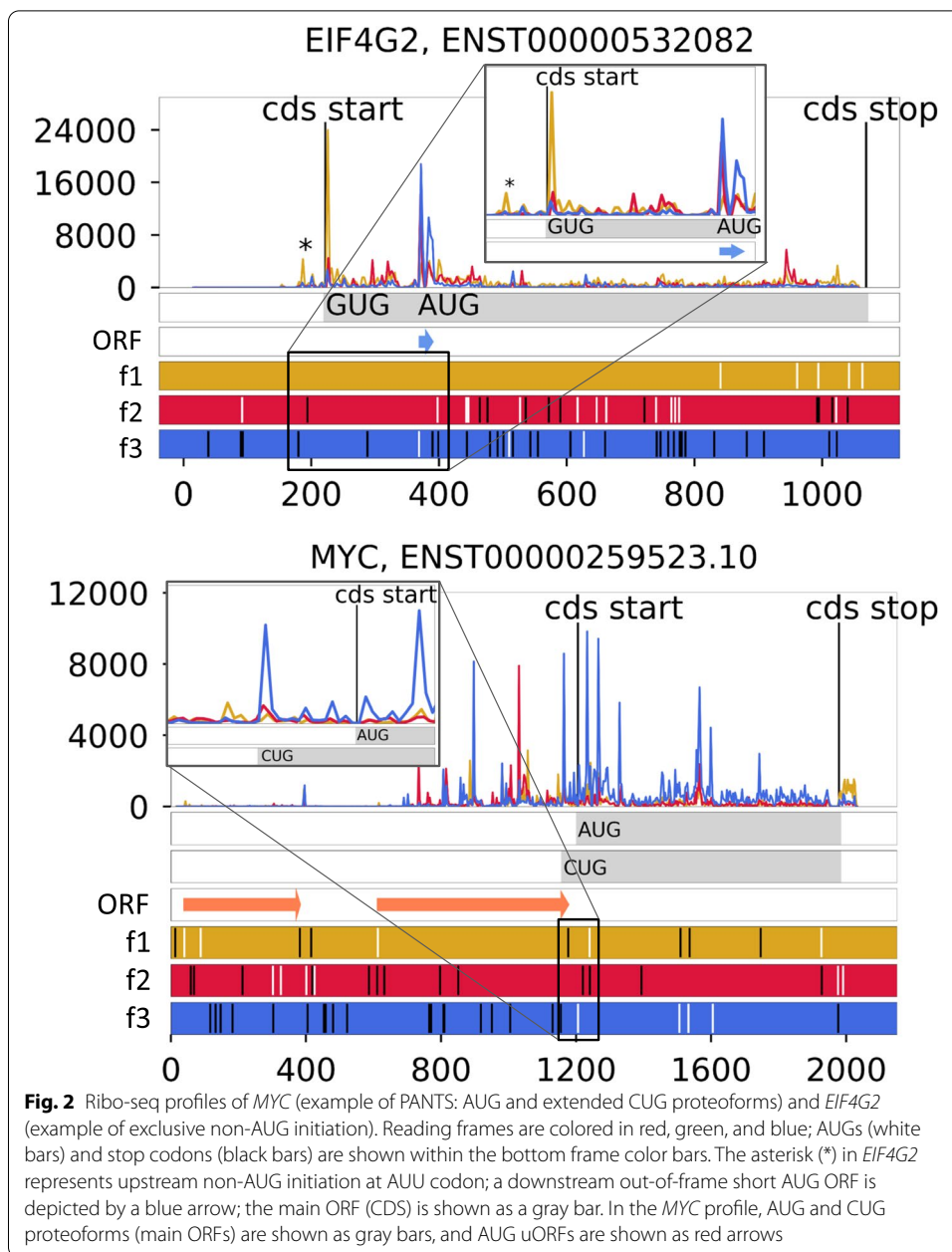
In the above examples, non-AUG TISs are located upstream of the main AUG start. Detectable non-AUG TISs are more frequent upstream of AUG TISs as expected according to the scanning model of translation initiation as fewer PICs would be available for non-AUG initiation downstream of efficient AUG initiators [47]. Yet, despite this, N-terminally truncated non-AUG initiated proteoforms are known (Fig. 1). A remarkable example was reported for *MRPL18* encoding mitochondrial ribosomal protein L18 [48]. Here initiation at a CUG codon occurs downstream of the main AUG codon under heat shock stress conditions. The truncated ribosomal L18 protein lacks a mitochondrial localization sequence and is instead incorporated into cytoplasmic ribosomes. These “hybrid” MRPL18 containing ribosomes have been proposed to be involved in the translation of stress induced mRNAs.

Interestingly, there are some mRNAs where the predominant initiation codon is non-AUG (Fig. 1). The most well-known example is in the *EIF4G2* [49] gene which codes for translation initiation factor eIF4G2 a.k.a. DAP5. Translation of *EIF4G2* mRNA is thought to occur exclusively at a GUG codon, although examination of publicly available ribosome profiling data in Trips-Viz [50] suggests the existence of an additional minor non-AUG TIS shortly upstream (Fig. 2).

Non-AUG initiation is a simple mechanism enabling multiple TISs on the same mRNA, but it does not lead to similar proteoforms with heterogeneous N-termini when it occurs in a reading frame different from the CDS (Fig. 1). An example of a protein encoded by a non-AUG initiated ORF that overlaps the main CDS was recently discovered in *POLG* mRNA [51, 52]. *POLG* encodes the catalytic subunit of mitochondrial DNA polymerase. Initiation at a CUG codon located upstream of the *POLG* CDS yields a long protein (260 amino acids in humans) termed POLGARF. While the functional role of POLGARF is still not clear, there is evidence that POLGARF may participate in extracellular signaling. Ectopically expressed POLGARF can be cleaved and secreted upon serum stimulation [52]. The likely importance of POLGARF is supported by phylogenetic conservation of its overlapping reading frame in placental mammals [51, 52]. Since the *POLGARF* coding ORF overlaps the *POLG* coding ORF by 725 nucleotides, there are implications for the interpretation of *POLG* synonymous mutations.

Non-AUG initiation of short ORFs (sORFs) translation

While it is likely that more novel proteins and proteoforms of known proteins synthesized as a result of non-AUG initiation are yet to be discovered, it is apparent that the



vast majority of non-AUG initiation leads to the translation of sORFs. In fact, ribosome profiling of ribosomes enriched at start codons suggests that translation of most sORFs initiates at non-AUG codons [24, 25] and this was confirmed with proteomic studies [53–55]. Because of leaky scanning, productive non-AUG codons are far more likely to occur upstream of AUG codons [47]. Subsequently most non-AUG sORFs are expected to be located at the beginning of RNA transcripts, though translation of non-AUG initiated ORFs downstream of the CDS TIS have also been reported [56]. What are the functional and phenotypic consequences of translation of these sORFs? It is certainly possible, and even likely, that this abundant but inefficient translation is driven by the high evolutionary cost of optimizing the translational apparatus for accurate translation

initiation. However, even if most non-AUG initiation is non-adaptive [57], it is clear that translation of at least some non-AUG sORFs is beneficial as evident from deep evolutionary conservation of some of them. An example is the high frequency of conserved non-AUG initiated uORFs in genes involved in the regulation of polyamine levels [58]. The regulatory roles of many uORFs are well established, with many enabling gene-specific regulation in response to particular conditions, e.g., polyamine [59] or magnesium [60] concentrations as well as during stress [17, 61]. Non-AUG initiated uORFs provide yet another regulatory layer to uORFs because of the apparently dynamic and regulatable stringency of translation initiation as we will see further below. During Oxygen and Glucose Deprivation (OGD), many non-AUG uORFs whose translation was reciprocal to that of its corresponding CDS were found. This coincided with altered translation of mRNAs involved in the regulation of translation initiation stringency [62] supporting their autoregulation discussed in more detail further below.

Even if the products of non-AUG sORFs or their translation have no specific functions, their translation is not inconsequential for the organism's phenotype. In addition to imposing an energetic burden, since protein synthesis is the most energetically demanding cellular process [63], the products of sORF translation could become a source for antigen presentation and thus contribute to the immune response [64, 65]. Of note, in addition to being recognized by standard Met-tRNA_i, a Leu-tRNA was identified that can also initiate (inefficiently) at CUG codons [66]. Given the low level of evolutionary selection acting on most non-AUG starts, the repertoire of translated sORFs may vary among individuals that may contribute to personalized immunopeptidomes. Apart from its likely role in shaping the immunopeptidome, promiscuous translation of non-AUG initiation has been implicated in cancer development linking it to cell survival and proliferation [67, 68]. The pathogenicity of non-AUG initiation at sORFs has received particular attention in relation to neurodegenerative diseases associated with repeat expansions, termed RAN (repeat-associated non-ATG) translation. Long repetitive sequences have been shown to stimulate translation initiation upstream; more details on the molecular processes and associated disease etiology can be found in several reviews dedicated to this topic [69–71]. However, it is important to note that during RAN translation, recognition of non-AUG codons seems to occur via the standard cap-dependent mechanism and involves canonical initiation factors and regulators [72–74].

Nucleotide contexts that influence non-AUG initiation

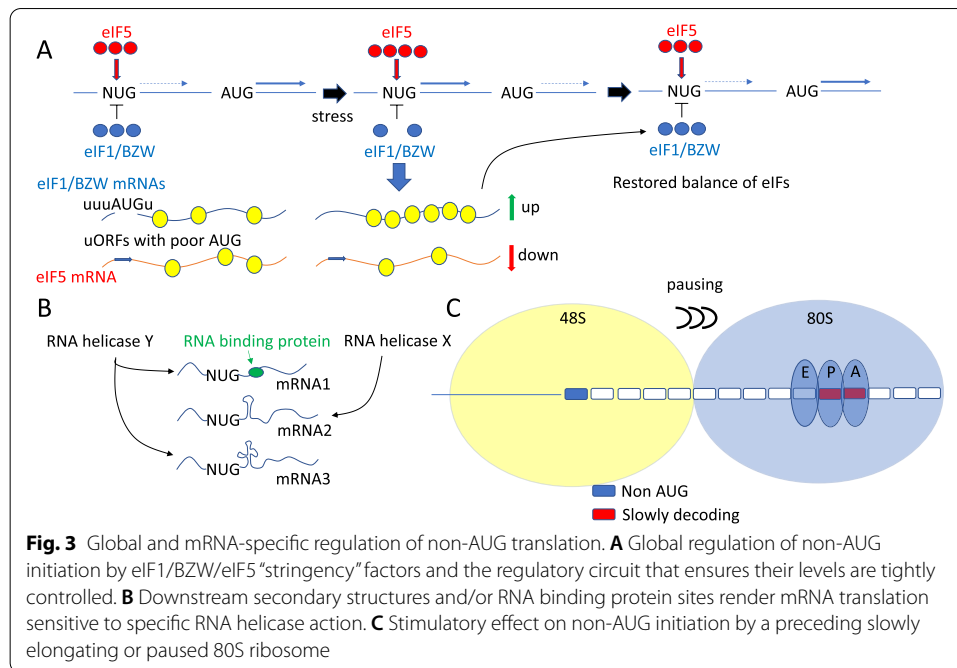
Perhaps the most important feature, other than mRNA position, that affects non-AUG initiation efficiency is the codon identity. Various approaches have been implemented to investigate initiation efficiencies of near cognate codons including reporter constructs [75] or ribosome profiling of cells treated with antibiotics that enrich for ribosomes at initiation sites [24, 25]. All of these studies reveal that the most efficient non-AUG codon is CUG, which is typically at least 2–3 times more efficient than any other non-AUG codon. It is not clear whether the CUG codon is the most efficient simply due to the minimal topological constraints of its codon-anticodon duplex [76] or for different reasons—we expect that this question will be addressed soon with structural studies of near cognate translation initiation complexes.

As AUG start codon recognition critically depends on its nucleotide context, it is not surprising that context has an even higher influence on non-AUG initiation, as additional forces that stabilize the preinitiation complex facilitate recognition of an inefficient start codon. This implies that non-AUG initiation could be sensitive to nucleotide positions which have little significant influence on initiation at AUG codons. Indeed, early studies based on reporter constructs indicated that nucleotides at +5 and +6 are important for the efficiency of non-AUG initiation (both studies concluding that the optimal nucleotides at positions +5 and +6 are A and U, respectively) [77, 78].

The development of high throughput approaches for massively parallel variant analysis of initiation efficiency, FACS-seq, allowed estimation of all individual context variants in cells. In these studies, libraries of GFP reporters in which several positions around the start codon were randomized were transfected into cells. The cells were sorted and binned by relative expression levels of the reporters before amplification and sequencing. This approach was first implemented for AUG contexts [11], and later extended to non-AUG codons [75]. FACS-seq studies provided new important insights into the determinants of near cognate start codon efficiencies. It directly demonstrated that the optimal context for initiation at non-AUG and AUG codons is not exactly the same. Certain nucleotide contexts enable non-AUG initiation with efficiencies comparable to AUG initiation. These findings challenged the hitherto prevailing view that the efficiency of non-AUG initiation is generally no more than 5–10% of that of an AUG. Another non-trivial observation was the variation of optimal contexts among different non-AUG codons. Certain contexts allow generally less efficient non-AUG codons to outperform the most efficient non-AUG codons (e.g., AUU vs CUG).

Although the above study was limited to the context within –4 to +4 positions and cannot be used to verify the claim about +5 and +6 positions, certain general conclusions can be made regarding the context of non-AUG initiation. Firstly, non-AUG initiation is critically dependent on the +4 nucleotide being guanine. While +4G is also optimal for AUG initiation, adenine in this position is usually tolerated without strong decreases in translation efficiency. In contrast, a +4G/A substitution has a dramatic effect on non-AUG initiation which can decrease by up to 10 times for CUG. Secondly, the –4 position also has a strong influence on near cognate initiation. Nucleotide changes at the –4 position caused the efficiency of non-AUG TISs to vary more than 2-folds, but the efficiency of AUG TISs varies only by ~10%. In particular, C (or also A, depending on the assay) in the –4 position caused a ~70% increase in the expression from all non-AUG codons compared to G in the same position. Altogether, with the early findings highlighting the influence of the +5 and +6 positions (which were not addressed in the FACS seq study), it is possible to predict that the most efficient non-AUG initiator (codon and context) as (C/A)(A/G)(C/A)(A/C/G)CUGGAU. Analysis of phylogenetic conservation of nucleotide contexts for *EIF4G2*, *R3CCH1*, and *POLGARG* non-AUG contexts (which were shown to achieve 20–40% of initiation relative to AUG) supports the importance of –4, –3, and +4 positions as C, A, and G respectively [52].

Notably, many known non-AUG starts implicated in the synthesis of functional proteoforms do not bear optimal near cognate contexts and hence can be considered as sub-optimal non-AUG initiators. For instance, the human *MYC* context (GACGCUGG) and



BAG1 context (GGGCCUGG) do not have $-4C$, and the *VEGFA* context (CGCGCUG A) lacks $+4G$. On the other hand, the non-optimality of non-AUG initiation is expected from the simple fact that in many cases initiation codons other than CUG are utilized.

Factors and intracellular conditions involved in global regulation of non-AUG initiation

In this and the following two sections, we will discuss 3 different ways of non-AUG initiation regulation illustrated in Fig. 3. The first example describes transcriptome-wide regulation by the modulation of the activity/availability of certain initiation factors involved in the stringency of start codon selection (Fig. 3A). The second example outlines changes in the composition of scanning complexes that affect their ability to unwind downstream RNA secondary structures known to stimulate initiation (Fig. 3B). This may selectively affect some initiation sites more than others depending on local RNA structures. The last example requires paused 80S ribosomes positioned downstream (Fig. 3C); again, such regulation is transcript specific.

The idea that some components of the translation apparatus can modulate start codon recognition arose from yeast genetic experiments when the Donahue group performed a genetic screening for *sui/SUI* alleles (suppressors of initiator codon mutants), taking advantage of these mutants’ ability to confer translation of a mutant *his4* allele despite the absence of an AUG start codon. These *sui/SUI* genes encode initiation factors eIF1, all three subunits of eIF2, and eIF5 [79–81].

eIF1 is a critical factor involved in the stringency of start codon selection. Initial studies from Pestova’s group demonstrated that in a mammalian reconstituted translation initiation system, omission of eIF1 prevents formation of the 48S complex on AUG codons and instead leads to an accumulation of 48S complexes upstream of the AUG [82]. Omission of eIF1 from this *in vitro* translation system leads to 48S complex formation at UUG and GUG codons located closer to the 5’ end of β -globin mRNA than the

AUG initiator [83]. The rate of dissociation of eIF1 from 48S seems to be a crucial determinant of start codon recognition [84]. Subsequently, it was shown that accurate translation initiation entails control of binding and release of eIF1 by other factors such as eIF3 [85, 86] and eIF5 [87, 88].

eIF5 also plays an important role in scanning and start codon selection [89, 90]. During the scanning process, eIF5 stimulates hydrolysis of GTP in the ternary complex (TC). However, the release of the phosphate (P_i) from eIF2-GDP- P_i is prevented until the start codon is recognized. Upon start codon recognition, the N-terminal domain of eIF5 occupies the eIF1 binding site on the ribosome and prevents eIF1 rebinding thus stabilizing the codon-anticodon interaction and promoting initiation [91].

The *BZW2* and *BZW1* genes encode translation regulatory proteins termed eIF5 mimic proteins 5MP1 and 5MP2 respectively. Both proteins are homologous to eIF5 at its C-terminal domain but lack its N terminal GTPase activating GAP domain. Like eIF5, 5MP1/2 can bind eIF2, but due to the absence of GAP activity they act as dominant negative translation repressors [13, 92].

Both eIF1, eIF5, and 5MP1 and 2, when overexpressed in mammalian cells, affect initiation at suboptimal codons, including non-AUG codons, at least with reporter constructs. Specifically, eIF1, 5MP1, and 5MP2 decrease initiation on suboptimal codons, while eIF5 stimulates it [14, 27]. These factors are likely very prominent in the regulation of non-AUG initiation at the genome wide level. As mentioned earlier, their own synthesis is autoregulated *via* a network of evolutionary conserved negative feedback loops. The initiation contexts of the *BZW1/2* and *EIF1* AUG codons are very weak, and high levels of 5MP1/2 or eIF1 (increases stringency) suppressing translation of their own mRNAs [12, 13]. The 5' leader of the eIF5 mRNA contains several inhibitory uORFs with evolutionary conserved poor initiation contexts [27]. Consequently, high levels of eIF5 (reduces stringency) also suppresses translation of its own mRNA by increasing translation of its inhibitory uORFs. This regulatory network (Fig. 3A) is interconnected, and these factors can also regulate each other's translation, e.g., increased eIF1 levels enhance eIF5 translation and vice versa. This conserved genetically encoded circuit likely serves to ensure stringent start codon selection appropriate for specific conditions.

In agreement with this, depletion of eIF1 in cells showed a somewhat modest effect on global translation [93]. Indeed, upon eIF1 deprivation, only 245 out of 10,000 actively translated transcripts showed changes in translation. As expected, under these conditions eIF1 depletion resulted in a concomitant decrease in eIF5 translation, thus restoring the balance of "stringency" factors and limiting the global effect on translation from suboptimal start site initiations [93]. As mentioned earlier, the cell response to metabolic stress such as OGD illustrates this regulatory system in action [62]. OGD leads to an increase of translation in 5' leaders which includes increased initiation at some non-AUG codons, which may indicate a relaxed stringency of translation initiation. Interestingly, after 60 min of OGD, *EIF5* is translationally repressed and this is accompanied by a marked increase in *EIF1* mRNA translation. This change is expected to alter the balance between eIF1 and eIF5 in order to make initiation more stringent in response to the stress. Notably, eIF1B, the paralog of eIF1 which shares 92% sequence identity, may also be implicated in start codon selection and the regulation of start codon selection stringency [93].

Helicase apparatus and its interplay with non-AUG initiation

In her pioneering work, Kozak proposed that RNA secondary structures located downstream of sub-optimal starts can enhance initiation [94]. The stimulatory effect of stem-loop structures is position dependent with an optimal location 14–16 nt downstream of a start codon which roughly corresponds to the region occupied by the ribosome based on RNase protection [95–97]. The most likely explanation is that a scanning ribosome complex pauses while unwinding the secondary structure, which positions the ribosome P-site close to the start codon thus increasing the dwell time at this codon and subsequently increasing the probability of start codon selection. In line with this observation, it has been recently shown that translation initiation of a CUG-initiated protein POLGARF is stimulated 3-fold by a 3' RNA stem loop. In combination with an efficient nucleotide context, it makes this CUG codon one of the most efficient non-AUG start sites with efficiencies up to 60–70% of AUG [52].

Whereas, in general, downstream structures can enhance non-AUG initiation, they vary and may respond differently to the compositions of scanning complexes. While it seems that the core helicase, eIF4A, is critical for scanning [98, 99], additional helicases and accessory factors are involved in the unwinding of specific structures upon scanning [100]. The unwinding of such “difficult” structures may increase dwell time of scanning ribosomes in the absence of corresponding factors and hence increase initiation at upstream non-AUG codons. This raises the intriguing idea that specific secondary structures downstream of suboptimal start codons can render non-AUG initiation sensitive to a specific RNA helicase activity. For instance, Murat and colleagues [101] investigated the role of two accessory RNA helicases, DHX36 and DHX9, in human cells using ribosome profiling. They demonstrated that depletion of the rG4-quadruplex unwinding helicases DHX36 and DHX9 promotes translation of rG4-associated uORFs while reducing the translation of their corresponding CDS. Mechanistically, these helicases assist in the unwinding of G quadruplexes (and perhaps other structures) which otherwise serve as roadblocks for scanning ribosomes, and upon helicase depletion such structures can promote initiation at suboptimal codons. A striking example of such a mechanism was demonstrated for *DDX23* mRNA translation, which is dependent on DHX9 helicase. A specific RNA motif located in the *DDX23* 5' leader, which binds DHX9, is located upstream of an AUG-initiated short uORF. siRNA-mediated depletion of DHX9 cause increased translation of this uORF which is accompanied by a decrease in its CDS translation [101].

It is reasonable to propose that similar mechanisms can also operate for non-AUG codons. Indeed, direct evidence of such a mechanism has come from experiments with yeast. Guenther and colleagues [102] investigated the role of DEAD-box RNA helicase Ded1p (DDX3 is its mammalian ortholog) in translation. Whereas Ded1p regulates global translation, some mRNAs are more sensitive to Ded1p depletion. It was further demonstrated that Ded1p depletion leads to increases in initiation at non-AUG codons located immediately upstream of RNA secondary structures.

In principle, similar initiation-inducing roadblocks can be created by RNA binding proteins. One prominent example is the regulation of *msl-2* mRNA translation by Sex lethal (SXL) protein, which is critical for dosage compensation in *Drosophila*. Medenbach and colleagues [103] demonstrated that SXL binding downstream of a short uORF

imposes a strong negative effect on main ORF translation, by increasing initiation at the uORF and augmenting its impediment to downstream translation. It is reasonable to expect that this mechanism will also operate for non-AUG codons and that specific RNA helicases may be required to displace a particular RNA binding protein during scanning. Thus, nucleotide sequence motifs located immediately downstream of non-AUG codons may confer specificity of mRNA translation to various accessory helicases and/or RNA binding proteins, and this allows protein-specific translation control of non-AUG initiation for selected mRNAs (Fig. 2B).

Interplay between elongation and initiation rates can be important for non-AUG initiation

Like RNA structures and RNA binding proteins, a paused elongating 80S ribosome could also provide a stimulatory effect on initiation upstream. Such a possibility was initially proposed to occur during translation of barley yellow dwarf virus RNA [104] and is likely responsible for an observed increase in non-AUG initiation transcriptome-wide in response to mild cycloheximide treatment, which slows down, but does not completely inhibit elongation [105]. This was shown not only using reporters, but also on endogenous mRNAs, e.g., initiation at GUG in *EIF4G2* mRNA. Whereas this effect was observed under artificial conditions of antibiotic treatment, it is possible that similar effects can be realized under physiological conditions when the elongation rate is significantly decreased. Global translation elongation rates are controlled *via* eEF2K-mediated phosphorylation of eEF2. eEF2K is a calmodulin-dependent kinase which phosphorylates eEF2 on T56 and prevents its binding to the ribosome [106–110]. Many signaling pathways control eEF2 activity indirectly by regulating the activity of eEF2K. For example, SAPK4/p38delta [111], p90(RSK1), and p70 S6 kinase [112] phosphorylate eEF2K and inhibit its activity, while AMP-activated protein kinase (AMPK) activates eEF2K [113, 114]. It is therefore reasonable to propose that eEF2K-mediated eEF2 phosphorylation, which occurs not only upon various stress conditions but also during mitosis [115], can globally enhance non-AUG initiation in a way that is reminiscent of low dose cycloheximide treatment. However, as stress conditions may also result in decreased initiation, this can counteract ribosome queue formation caused by slow elongation.

This mechanism implies that slowly decoded sequences downstream of a suboptimal codon could selectively enhance non-AUG initiation. Indeed, such an example was reported for the *AZINI* mRNA which codes for antizyme inhibitor, a key regulator of polyamine biosynthesis. Elevated polyamine levels repress translation of the *AZINI* main ORF, and this is dependent on the presence of an AUU-initiated inhibitory uORF [58]. A subsequent study revealed the mechanism of uORF mediated translation control—a PPW motif encoded at the C-terminus of the uORF promotes ribosome stalling when polyamine levels are high, and this stalling enhances initiation at the uORF AUU codon ~30 codons upstream [116], thus confirming Dinesh-Kumar's [104] hypothesis and providing a possible mechanistic explanation for the later observation of an increase in global non-AUG initiation in response to elongation inhibitors [105].

Site-specific ribosomal pauses can control the use of non-optimal start codons at a genome-wide level. eIF5A is a highly conserved translation factor which functions during both translation elongation and termination. During translation elongation, it facilitates peptide bond formation between amino acids where the peptidyl transfer reaction

is thought to be inefficient [117–119]. Manjunath and colleagues [120] showed that depletion of eIF5A enhances upstream translation within some 5' leaders across yeast and human transcriptomes. Among those mRNAs that are regulated upon eIF5A depletion is *MYC* mRNA, where there is increased synthesis of an N-terminally extended proteoform. Importantly, the eIF5A-dependent pause site should be located close enough to the suboptimal start codon to promote its utilization upon eIF5A depletion, which is consistent with the queuing model. A similar role for eIF5A was also shown for the regulation of the AUU-initiated AZIN1 uORF [116]. The effect of eIF5A on non-AUG initiation was also documented in yeast and this type of regulation has been found during meiosis [121].

It is expected that any site-specific pause of elongating ribosomes may have similar effects on non-optimal start codons. Such ribosome pauses can be induced by rare codons, unavailability of cognate charged tRNAs or nascent chain-mediated ribosome slow down. Therefore, the amino acid sequence downstream of non-AUG or sub-optimal start codons may be a critical determinant of mRNA specific initiation (Fig. 3C).

Conclusion

It is becoming increasingly clear that initiation at non-AUG codons is widely utilized in mammalian cells and significantly contributes to “hidden proteome” diversity. Such non-canonical start codons have a wide range of initiation efficiencies, from zero activity to efficiencies comparable with that of AUG codons. A specific nucleotide context is required for efficient non-AUG initiation which can be further enhanced by downstream stimulators (such as an RNA secondary structure), and, in some rare cases, such codons are recognized almost as well as AUG codons by the translation machinery, as the initiation factors that usually “sense” suboptimal codons fail to affect initiation on such “super-optimal” non-AUG starts [52]. Nevertheless, even the most efficiently initiated non-AUG codons appear to be “leaky” (Fig. 2, translation of downstream AUG initiated ORFs is evident).

Recent studies demonstrated that non-AUG initiation can be exceptionally sensitive to conditions that pause scanning ribosome progression. This includes downstream secondary structures which are likely difficult to unwind by conventional RNA helicase eIF4A activity, and slowly decoded triplets which may induce a queue of elongating 80S ribosomes. In this regard, non-AUG codons are more dependent on downstream sequences than AUG codons, where good nucleotide context is usually enough to provide efficient initiation. Downstream sequences can provide selective translation control of mRNA under various conditions. In our opinion, it is likely that specific non-AUG codons in 5' leaders are among the most widespread and important genetically encoded regulatory elements in the mammalian genome. Identification of active non-AUG initiation start sites is essential for genome annotation and for the discovery of functional micropeptides and proteoforms [122].

The implication of non-AUG initiation in various human diseases, such as cancer [67, 68] and neurological disorders [69, 123], prompted many to suggest it as a therapeutic target. While affecting non-AUG initiation is unlikely to be specific due to the ubiquitous nature of this phenomenon, the sensitivity of particular non-AUG initiators to specific factors and conditions suggests that such an approach could hold some promise.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-022-02674-2>.

Additional file 1. Peer review history.

Acknowledgements

We would like to acknowledge financial support from Russian Science Foundation to D.E.A. [№20-14-00121], from SFI-HRB-Wellcome Trust Biomedical Research Partnership (Investigator in Science award) to P.V.B. [210692/Z/18/] and from Science Foundation Ireland (SFI Centre for Research Training in Genomics Data Science) to A.D.F. [18/CRT/6214].

Review history

The review history is available as Additional file 1.

Peer review information

Tim Sands was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

D.E.A. and P.V.B. conceptualized the work and drafted the review. D.E.A. and A.D.F. designed the figures. All authors reviewed and approved the literature, discussed, and edited the manuscript.

Availability of data and materials

No supporting data or materials are provided for this manuscript.

Declarations

Competing interests

G.L. and P.V.B. are co-founders of RiboMaps Ltd. Detection of translation initiation sites with ribosome profiling is one of the services provided by RiboMaps. Therefore, these authors may indirectly benefit from the publication of this review as it would increase general awareness of translation initiation at non-AUG codons among the readers and potential clients.

Author details

¹Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow 117997, Russia. ²Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119992, Russia. ³School of Biochemistry and Cell Biology, University College Cork, Cork T12 XF62, Ireland.

Received: 14 August 2021 Accepted: 14 April 2022

Published online: 09 May 2022

References

1. Kozak M. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell*. 1978;15:1109–23.
2. Kozak M. Evaluation of the “scanning model” for initiation of protein synthesis in eucaryotes. *Cell*. 1980;22:7–8.
3. Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*. 2010;11:113–27.
4. Hinnebusch AG. The scanning mechanism of eukaryotic translation initiation. *Annu Rev Biochem*. 2014;83:779–812.
5. Kozak M. Initiation of translation in prokaryotes and eukaryotes. *Gene*. 1999;234:187–208.
6. Ogle JM, Brodersen DE, Clemons WM Jr, Tarry MJ, Carter AP, Ramakrishnan V. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science*. 2001;292:897–902.
7. Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G. A new understanding of the decoding principle on the ribosome. *Nature*. 2012;484:256–9.
8. Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*. 1986;44:283–92.
9. Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*. 1987;15:8125–48.
10. Kozak M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J Mol Biol*. 1987;196:947–50.
11. Noderer WL, Flockhart RJ, Bhaduri A, Diaz de Arce AJ, Zhang J, Khavari PA, et al. Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Mol Syst Biol*. 2014;10:748.
12. Ivanov IP, Loughran G, Sachs MS, Atkins JF. Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). *Proc Natl Acad Sci U S A*. 2010;107:18056–60.
13. Loughran G, Firth AE, Atkins JF, Ivanov IP. Translational autoregulation of BZW1 and BZW2 expression by modulating the stringency of start codon selection. *PLoS One*. 2018;13:e0192648.
14. Tang L, Morris J, Wan J, Moore C, Fujita Y, Gillaspie S, et al. Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. *Nucleic Acids Res*. 2017;45:11941–53.

15. Benitez-Cantos MS, Yordanova MM, O'Connor PBF, Zhdanov AV, Kovalchuk SI, Papkovsky DB, et al. Translation initiation downstream from annotated start codons in human mRNAs coevolves with the Kozak context. *Genome Res.* 2020;30:974–84.
16. Cloutier P, Poitras C, Faubert D, Bouchard A, Blanchette M, Gauthier MS, et al. Upstream ORF-encoded ASDURF is a novel prefoldin-like subunit of the PAQosome. *J Proteome Res.* 2020;19:18–27.
17. Andreev DE, O'Connor PB, Fahey C, Kenny EM, Terenin IM, Dmitriev SE, et al. Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife.* 2015;4:e03971.
18. Brown A, Rathore S, Kimanius D, Aibara S, Bai XC, Rorbach J, et al. Structures of the human mitochondrial ribosome in native states of assembly. *Nat Struct Mol Biol.* 2017;24:866–9.
19. Delcourt V, Brunelle M, Roy AV, Jacques JF, Salzert M, Fournier I, et al. The protein coded by a short open reading frame, not by the annotated coding sequence, is the main gene product of the dual-coding gene MIEF1. *Mol Cell Proteomics.* 2018;17:2402–11.
20. Rathore A, Chu Q, Tan D, Martinez TF, Donaldson CJ, Diedrich JK, et al. MIEF1 microprotein regulates mitochondrial translation. *Biochemistry.* 2018;57:5564–75.
21. Akimoto C, Sakashita E, Kasashima K, Kuroiwa K, Tominaga K, Hamamoto T, et al. Translational repression of the McKusick-Kaufman syndrome transcript by unique upstream open reading frames encoding mitochondrial proteins with alternative polyadenylation sites. *Biochim Biophys Acta.* 2013;1830:2728–38.
22. Chew GL, Pauli A, Schier AF. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat Commun.* 2016;7:11663.
23. Gunisova S, Hronova V, Mohammad MP, Hinnebusch AG, Valasek LS. Please do not recycle! Translation reinitiation in microbes and higher eukaryotes. *FEMS Microbiol Rev.* 2018;42:165–92.
24. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell.* 2011;147:789–802.
25. Lee S, Liu B, Lee S, Huang SX, Shen B, Qian SB. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A.* 2012;109:E2424–32.
26. Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, et al. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* 2012;22:2208–18.
27. Loughran G, Sachs MS, Atkins JF, Ivanov IP. Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. *Nucleic Acids Res.* 2012;40:2898–906.
28. Asano K. Why is start codon selection so precise in eukaryotes? *Translation (Austin).* 2014;2:e28387.
29. Hinnebusch AG, Lorsch JR. The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harb Perspect Biol.* 2012;4(10):a011544.
30. Spealman P, Naik AW, May GE, Kuersten S, Freeberg L, Murphy RF, et al. Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. *Genome Res.* 2018;28:214–22.
31. Hann SR, King MW, Bentley DL, Anderson CW, Eisenman RN. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell.* 1988;52:185–95.
32. Hann SR, Dixit M, Sears RC, Sealy L. The alternatively initiated c-Myc proteins differentially regulate transcription through a noncanonical DNA-binding site. *Genes Dev.* 1994;8:2441–52.
33. Sato K, Masuda T, Hu Q, Tobo T, Gillaspie S, Niida A, et al. Novel oncogene 5MP1 reprograms c-Myc translation initiation to drive malignant phenotypes in colorectal cancer. *EBioMedicine.* 2019;44:387–402.
34. Tee MK, Jaffe RB. A precursor form of vascular endothelial growth factor arises by initiation from an upstream in-frame CUG codon. *Biochem J.* 2001;359:219–26.
35. Yang X, Chernenko G, Hao Y, Ding Z, Pater MM, Pater A, et al. Human BAG-1/RAP46 protein is generated as four isoforms by alternative translation initiation and overexpressed in cancer cells. *Oncogene.* 1998;17:981–9.
36. Takayama S, Krajewski S, Krajewska M, Kitada S, Zapata JM, Kochel K, et al. Expression and location of Hsp70/Hsc-binding anti-apoptotic protein BAG-1 and its variants in normal tissues and tumor cell lines. *Cancer Res.* 1998;58:3116–31.
37. Froesch BA, Takayama S, Reed JC. BAG-1L protein enhances androgen receptor function. *J Biol Chem.* 1998;273:11660–6.
38. Shatkina L, Mink S, Rogatsch H, Klocker H, Langer G, Nestl A, et al. The cochaperone Bag-1L enhances androgen receptor action via interaction with the NH2-terminal region of the receptor. *Mol Cell Biol.* 2003;23:7189–97.
39. Cutress RI, Townsend PA, Sharp A, Maison A, Wood L, Lee R, et al. The nuclear BAG-1 isoform, BAG-1L, enhances oestrogen-dependent transcription. *Oncogene.* 2003;22:4973–82.
40. Cato L, Neeb A, Sharp A, Buzon V, Ficarro SB, Yang L, et al. Development of Bag-1L as a therapeutic target in androgen receptor-dependent prostate cancer. *Elife.* 2017;6:e27159.
41. Bugler B, Amalric F, Prats H. Alternative initiation of translation determines cytoplasmic or nuclear localization of basic fibroblast growth factor. *Mol Cell Biol.* 1991;11:573–7.
42. Arnaud E, Touriol C, Boutonnet C, Gensac MC, Vagner S, Prats H, et al. A new 34-kilodalton isoform of human fibroblast growth factor 2 is cap dependently synthesized by using a non-AUG start codon and behaves as a survival factor. *Mol Cell Biol.* 1999;19:505–14.
43. Couderc B, Prats H, Bayard F, Amalric F. Potential oncogenic effects of basic fibroblast growth factor requires cooperation between CUG and AUG-initiated forms. *Cell Regul.* 1991;2:709–18.
44. Ivanov IP, Firth AE, Michel AM, Atkins JF, Baranov PV. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.* 2011;39:4220–34.
45. Hopkins BD, Fine B, Steinbach N, Dendy M, Rapp Z, Shaw J, et al. A secreted PTEN phosphatase that enters cells to alter signaling and survival. *Science.* 2013;341:399–402.
46. Tzani I, Ivanov IP, Andreev DE, Dmitriev RI, Dean KA, Baranov PV, et al. Systematic analysis of the PTEN 5' leader identifies a major AUU initiated proteoform. *Open Biol.* 2016;6(5):150203.
47. Michel AM, Andreev DE, Baranov PV. Computational approach for calculating the probability of eukaryotic translation initiation from ribo-seq data that takes into account leaky scanning. *BMC Bioinformatics.* 2014;15:380.

48. Zhang X, Gao X, Coots RA, Conn CS, Liu B, Qian SB. Translational control of the cytosolic stress response by mitochondrial ribosomal protein L18. *Nat Struct Mol Biol.* 2015;22:404–10.
49. Imataka H, Olsen HS, Sonenberg N. A new translational regulator with homology to eukaryotic translation initiation factor 4G. *EMBO J.* 1997;16:817–25.
50. Kiniry SJ, Judge CE, Michel AM, Baranov PV. Trips-Viz: an environment for the analysis of public and user-generated ribosome profiling data. *Nucleic Acids Res.* 2021;49(W1):W662–W670.
51. Khan YA, Jungreis I, Wright JC, Mudge JM, Choudhary JS, Firth AE, et al. Evidence for a novel overlapping coding sequence in POLG initiated at a CUG start codon. *BMC Genet.* 2020;21:25.
52. Loughran G, Zhdanov AV, Mikhaylova MS, Rozov FN, Datskevich PN, Kovalchuk SI, et al. Unusually efficient CUG initiation of an overlapping reading frame in POLG mRNA yields novel protein POLGARF. *Proc Natl Acad Sci U S A.* 2020;117:24936–46.
53. Ma J, Ward CC, Jungreis I, Slavoff SA, Schwaid AG, Neveu J, et al. Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *J Proteome Res.* 2014;13:1757–65.
54. Zhang Q, Wu E, Tang Y, Zhang L, Wang J, Hao Y, et al. Deeply mining a universe of peptides encoded by long noncoding RNAs. *Mol Cell Proteomics.* 2021;20:100109.
55. Brunet MA, Lucier JF, Levesque M, Leblanc S, Jacques JF, Al-Saedi HRH, et al. OpenProt 2021: deeper functional annotation of the coding potential of eukaryotic genomes. *Nucleic Acids Res.* 2021;49:D380–8.
56. Wu Q, Wright M, Gogol MM, Bradford WD, Zhang N, Bazzini AA. Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J.* 2020;39:e104763.
57. Xu C, Zhang J. Mammalian alternative translation initiation is mostly nonadaptive. *Mol Biol Evol.* 2020;37:2015–28.
58. Ivanov IP, Loughran G, Atkins JF. uORFs with unusual translational start codons autoregulate expression of eukaryotic ornithine decarboxylase homologs. *Proc Natl Acad Sci U S A.* 2008;105:10079–84.
59. Hill JR, Morris DR. Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA. Dependence on translation and coding capacity of the cis-acting upstream open reading frame. *J Biol Chem.* 1993;268:726–31.
60. Hardy S, Kostantin E, Wang SJ, Hristova T, Galicia-Vazquez G, Baranov PV, et al. Magnesium-sensitive upstream ORF controls PRL phosphatase expression to mediate energy metabolism. *Proc Natl Acad Sci U S A.* 2019;116:2925–34.
61. Starck SR, Tsai JC, Chen K, Shodiya M, Wang L, Yahiro K, et al. Translation from the 5' untranslated region shapes the integrated stress response. *Science.* 2016;351:aad3867.
62. Andreev DE, O'Connor PB, Zhdanov AV, Dmitriev RI, Shatsky IN, Papkovsky DB, et al. Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol.* 2015;16:90.
63. Lynch M, Marinov GK. The bioenergetic costs of a gene. *Proc Natl Acad Sci U S A.* 2015;112:15690–5.
64. Bullock TN, Eisenlohr LC. Ribosomal scanning past the primary initiation codon as a mechanism for expression of CTL epitopes encoded in alternative reading frames. *J Exp Med.* 1996;184:1319–29.
65. Wei J, Kishnot RJ, Angel M, Conn CS, Dalla-Venezia N, Marcel V, et al. Ribosomal proteins regulate MHC class I peptide generation for immunosurveillance. *Mol Cell.* 2019;73(1162–1173):e1165.
66. Starck SR, Jiang V, Pavon-Eternod M, Prasad S, McCarthy B, Pan T, et al. Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science.* 2012;336:1719–23.
67. Prensner JR, Enache OM, Luria V, Klauser KR, Dempster JM, et al. Noncanonical open reading frames encode functional proteins essential for cancer cell survival. *Nat Biotechnol.* 2021;39:697–704.
68. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, et al. Translation from unconventional 5' start sites drives tumour initiation. *Nature.* 2017;541:494–9.
69. Nguyen L, Cleary JD, Ranum LPW. Repeat-associated non-ATG translation: molecular mechanisms and contribution to neurological disease. *Annu Rev Neurosci.* 2019;42:227–47.
70. Todd TW, Petrucelli L. Insights into the pathogenic mechanisms of Chromosome 9 open reading frame 72 (C9orf72) repeat expansions. *J Neurochem.* 2016;138(Suppl 1):145–62.
71. Rodriguez CM, Todd PK. New pathologic mechanisms in nucleotide repeat expansion disorders. *Neurobiol Dis.* 2019;130:104515.
72. Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, et al. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron.* 2013;78:440–55.
73. Kearse MG, Green KM, Krans A, Rodriguez CM, Linsalata AE, Goldstrohm AC, et al. CGG repeat-associated non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. *Mol Cell.* 2016;62:314–22.
74. Singh CR, Glineburg MR, Moore C, Tani N, Jaiswal R, Zou Y, et al. Human oncoprotein 5MP suppresses general and repeat-associated non-AUG translation via eIF3 by a common mechanism. *Cell Rep.* 2021;36:109376.
75. Diaz de Arce AJ, Noderer WL, Wang CL. Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. *Nucleic Acids Res.* 2018;46:985–94.
76. Kameda T, Asano K, Togashi Y. Free energy landscape of RNA binding dynamics in start codon recognition by eukaryotic ribosomal pre-initiation complex. *PLoS Comput Biol.* 2021;17:e1009068.
77. Boeck R, Kolakofsky D. Positions +5 and +6 can be major determinants of the efficiency of non-AUG initiation codons for protein synthesis. *EMBO J.* 1994;13:3608–17.
78. Grunert S, Jackson RJ. The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons. *EMBO J.* 1994;13:3618–30.
79. Donahue TF, Cigan AM, Pabich EK, Valavicius BC. Mutations at a Zn(II) finger motif in the yeast eIF-2 beta gene alter ribosomal start-site selection during the scanning process. *Cell.* 1988;54:621–32.
80. Castilho-Valavicius B, Yoon H, Donahue TF. Genetic characterization of the *Saccharomyces cerevisiae* translational initiation suppressors sui1, sui2 and SUI3 and their effects on HIS4 expression. *Genetics.* 1990;124:483–95.
81. Huang HK, Yoon H, Hannig EM, Donahue TF. GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev.* 1997;11:2396–413.
82. Pestova TV, Borukhov SI, Hellen CU. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature.* 1998;394:854–9.

83. Andreev D, Hauryliuk V, Terenin I, Dmitriev S, Ehrenberg M, Shatsky I. The bacterial toxin RelE induces specific mRNA cleavage in the A site of the eukaryote ribosome. *RNA*. 2008;14:233–9.
84. Cheung YN, Maag D, Mitchell SF, Fekete CA, Algire MA, Takacs JE, et al. Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. *Genes Dev*. 2007;21:1217–30.
85. Obayashi E, Luna RE, Nagata T, Martin-Marcos P, Hiraishi H, Singh CR, et al. Molecular landscape of the ribosome pre-initiation complex during mRNA scanning: structural role for eIF3c and its control by eIF5. *Cell Rep*. 2017;18:2651–63.
86. Valasek L, Nielsen KH, Zhang F, Fekete CA, Hinnebusch AG. Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection. *Mol Cell Biol*. 2004;24:9437–55.
87. Yamamoto Y, Singh CR, Marintchev A, Hall NS, Hannig EM, Wagner G, et al. The eukaryotic initiation factor (eIF) 5 HEAT domain mediates multifactor assembly and scanning with distinct interfaces to eIF1, eIF2, eIF3, and eIF4G. *Proc Natl Acad Sci U S A*. 2005;102:16164–9.
88. Luna RE, Arthanari H, Hiraishi H, Nanda J, Martin-Marcos P, Markus MA, et al. The C-terminal domain of eukaryotic initiation factor 5 promotes start codon recognition by its dynamic interplay with eIF1 and eIF2beta. *Cell Rep*. 2012;1:689–702.
89. Singh CR, Curtis C, Yamamoto Y, Hall NS, Kruse DS, He H, et al. Eukaryotic translation initiation factor 5 is critical for integrity of the scanning preinitiation complex and accurate control of GCN4 translation. *Mol Cell Biol*. 2005;25:5480–91.
90. Asano K, Clayton J, Shalev A, Hinnebusch AG. A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate in vivo. *Genes Dev*. 2000;14:2534–46.
91. Llacer JL, Hussain T, Saini AK, Nanda JS, Kaur S, Gordiyenko Y, et al. Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *Elife*. 2018;7:e39273.
92. Singh CR, Watanabe R, Zhou D, Jennings MD, Fukao A, Lee B, et al. Mechanisms of translational regulation by a human eIF5-mimic protein. *Nucleic Acids Res*. 2011;39:8314–28.
93. Fijalkowska D, Verbruggen S, Ndah E, Jonckheere V, Menschaert G, Van Damme P. eIF1 modulates the recognition of suboptimal translation initiation sites and steers gene expression via uORFs. *Nucleic Acids Res*. 2017;45:7997–8013.
94. Kozak M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci U S A*. 1990;87:8301–5.
95. Kozak M, Shatkin AJ. Sequences and properties of two ribosome binding sites from the small size class of reovirus messenger RNA. *J Biol Chem*. 1977;252:6895–908.
96. Steitz JA. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature*. 1969;224:957–64.
97. Wolin SL, Walter P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J*. 1988;7:3559–69.
98. Ray BK, Lawson TG, Kramer JC, Cladaras MH, Grifo JA, Abramson RD, et al. ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J Biol Chem*. 1985;260:7651–8.
99. Grifo JA, Tahara SM, Leis JP, Morgan MA, Shatkin AJ, Merrick WC. Characterization of eukaryotic initiation factor 4A, a protein involved in ATP-dependent binding of globin mRNA. *J Biol Chem*. 1982;257:5246–52.
100. Parsyan A, Svitkin Y, Shahbazian D, Gkogkas C, Lasko P, Merrick WC, et al. mRNA helicases: the tacticians of translational control. *Nat Rev Mol Cell Biol*. 2011;12:235–45.
101. Murat P, Marsico G, Herdy B, Ghanbarian AT, Portella G, Balasubramanian S. RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol*. 2018;19:229.
102. Guenther UP, Weinberg DE, Zubradt MM, Tedeschi FA, Stawicki BN, Zagore LL, et al. The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. *Nature*. 2018;559:130–4.
103. Medenbach J, Seiler M, Hentze MW. Translational control via protein-regulated upstream open reading frames. *Cell*. 2011;145:902–13.
104. Dinesh-Kumar SP, Miller WA. Control of start codon choice on a plant viral RNA encoding overlapping genes. *Plant Cell*. 1993;5:679–92.
105. Kears MG, Goldman DH, Choi J, Nwaezeapu C, Liang D, Green KM, et al. Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. *Genes Dev*. 2019;33:871–85.
106. Price NT, Redpath NT, Severinov KV, Campbell DG, Russell JM, Proud CG. Identification of the phosphorylation sites in elongation factor-2 from rabbit reticulocytes. *FEBS Lett*. 1991;282:253–8.
107. Ovchinnikov LP, Motuz LP, Natapov PG, Averbuch LJ, Wettenhall RE, Szyszka R, et al. Three phosphorylation sites in elongation factor 2. *FEBS Lett*. 1990;275:209–12.
108. Nairn AC, Palfrey HC. Identification of the major Mr 100,000 substrate for calmodulin-dependent protein kinase III in mammalian cells as elongation factor-2. *J Biol Chem*. 1987;262:17299–303.
109. Ryazanov AG. Ca²⁺/calmodulin-dependent phosphorylation of elongation factor 2. *FEBS Lett*. 1987;214:331–4.
110. Ryazanov AG, Davydova EK. Mechanism of elongation factor 2 (EF-2) inactivation upon phosphorylation. Phosphorylated EF-2 is unable to catalyze translocation. *FEBS Lett*. 1989;251:187–90.
111. Knebel A, Morrice N, Cohen P. A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38delta. *EMBO J*. 2001;20:4360–9.
112. Wang X, Li W, Williams M, Terada N, Alessi DR, Proud CG. Regulation of elongation factor 2 kinase by p90(RSK1) and p70 S6 kinase. *EMBO J*. 2001;20:4370–9.
113. Browne GJ, Finn SG, Proud CG. Stimulation of the AMP-activated protein kinase leads to activation of eukaryotic elongation factor 2 kinase and to its phosphorylation at a novel site, serine 398. *J Biol Chem*. 2004;279:12220–31.
114. Horman S, Browne G, Krause U, Patel J, Vertommen D, Bertrand L, et al. Activation of AMP-activated protein kinase leads to the phosphorylation of elongation factor 2 and an inhibition of protein synthesis. *Curr Biol*. 2002;12:1419–23.

115. Celis JE, Madsen P, Ryazanov AG. Increased phosphorylation of elongation factor 2 during mitosis in transformed human amnion cells correlates with a decreased rate of protein synthesis. *Proc Natl Acad Sci U S A*. 1990;87:4231–5.
116. Ivanov IP, Shin BS, Loughran G, Tzani I, Young-Baird SK, Cao C, et al. Polyamine control of translation elongation regulates start site selection on antizyme inhibitor mRNA via ribosome queuing. *Mol Cell*. 2018;70(254-264):e256.
117. Pelechano V, Alepuz P. eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Res*. 2017;45:7326–38.
118. Saini P, Eyler DE, Green R, Dever TE. Hypusine-containing protein eIF5A promotes translation elongation. *Nature*. 2009;459:118–21.
119. Schuller AP, Wu CC, Dever TE, Buskirk AR, Green R. eIF5A functions globally in translation elongation and termination. *Mol Cell*. 2017;66(194-205):e195.
120. Manjunath H, Zhang H, Rehfeld F, Han J, Chang TC, Mendell JT. Suppression of ribosomal pausing by eIF5A is necessary to maintain the fidelity of start codon selection. *Cell Rep*. 2019;29(3134-3146):e3136.
121. Eisenberg AR, Higdon AL, Hollerer I, Fields AP, Jungreis I, Diamond PD, et al. Translation initiation site profiling reveals widespread synthesis of non-AUG-initiated protein isoforms in yeast. *Cell Syst*. 2020;11(145-160):e145.
122. Chen J, Brunner AD, Cogan JZ, Nunez JK, Fields AP, Adamson B, et al. Pervasive functional translation of non-canonical human open reading frames. *Science*. 2020;367:1140–6.
123. Zu T, Pattamatta A, Ranum LPW. Repeat-associated non-ATG translation in neurological diseases. *Cold Spring Harb Perspect Biol*. 2018;10(12):a033019.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

