Trends in Cell Biology

CellPress

Special Issue: Ribosomal Dynamics and Cellular Homeostasis

Review

Mechanisms suppressing noncoding translation

Jordan S. Kesner[1,2] and Xuebing Wu[1,2,*]

The majority of the DNA sequence in our genome is noncoding and not intended for synthesizing proteins. Nonetheless, genome-wide mapping of ribosome footprints has revealed widespread translation in annotated noncoding sequences, including long noncoding RNAs (lncRNAs), untranslated regions (UTRs), and introns of mRNAs. How cells suppress the translation of potentially toxic proteins from various noncoding sequences remains poorly understood. This review summarizes mechanisms for the mitigation of noncoding translation, including the BCL2-associated athanogene 6 (BAG6)-mediated proteasomal degradation pathway, which has emerged as a unifying mechanism to suppress the translation of diverse noncoding sequences in metazoan cells.

Highlights

Translation is widespread in annotated noncoding sequences, including untranslated regions (UTRs), introns, and long noncoding RNAs (lncRNAs), especially in contexts such as cancer, aging, and neurodegeneration.

Unbiased genetic and biochemical screens both identified the BCL2-associated athanogene 6 (BAG6) pathway for mediation of the proteasomal degradation of diverse noncoding translation products.

BAG6 recognizes a hydrophobic C-terminal tail, a common feature of proteins translated from all types of noncoding sequences. This results from the U-rich nature of the noncoding genome and the strong bias of U-rich codons for hydrophobic amino acids in the genetic code.

Pervasive noncoding translation

The simplicity of the central dogma, which states that genetic information flows from DNA to RNA to protein, conceals the complexity that most of the DNA sequence in our genome, and even most of the RNA sequence in our transcriptome, does not encode functional proteins [1]. Noncoding sequences, including those in intergenic regions, introns, and the UTRs of protein-coding genes, constitute over 98% of our genome. Over the past two decades, extensive transcriptomic surveys have revealed pervasive transcription of the noncoding genome, generating tens of thousands of **lncRNAs** (see Glossary) [1,2]. These lncRNAs are mRNA-like transcripts that lack long **open reading frames (ORFs)**. Additionally, the extensive use of alternative promoters, splice sites, and **intronic polyadenylation** [poly(A)] sites in coding genes produces thousands of aberrant mRNA isoforms [3–5]. These aberrant mRNAs contain **noncanonical ORFs** that include annotated intronic or UTR sequences. Unexpectedly, unbiased mapping of ribosome footprints using **ribosome profiling** has revealed pervasive translation in annotated noncoding sequences, including lncRNAs, introns, and the UTRs of mRNAs (hereinafter referred to as **noncoding translation**, Figure 1) in the cells of various eukaryotic species (human, mouse, zebrafish, fruit fly, *Arabidopsis*, and yeast) [6–16]. For instance, a meta-analysis of over 600 ribosome profiling datasets in humans identified 58 383 translated noncanonical ORFs, far exceeding the 33 251 canonical ORFs identified as being translated [15].

The prevalence of noncoding translation in mammalian cells underscores the need for a deeper understanding of its consequences and mitigation mechanisms. The translation of noncoding sequences is expected to predominantly produce nonfunctional proteins that are likely to misfold and/or aggregate, potentially causing cytotoxicity. Furthermore, the increased activity of noncoding translation that has been observed in aging [9,17,18], neurodegeneration [17,19], and cancer [12,20,21] similarly highlights the importance of understanding how cells suppress noncoding translation. The existence of mechanisms mitigating noncoding translation was revealed in a seminal study from the Andrew Fire laboratory showing considerable protein loss as a result of stop codon **readthrough** and subsequent translation of 3′ UTRs in both worms and human cells [22]. However, the molecular details of the mitigation mechanisms, including the signals

[1]Department of Medicine, Columbia University Irving Medical Center, New York, NY 10032, USA
[2]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032, USA

*Correspondence:
xuebing.wu@columbia.edu (X. Wu).

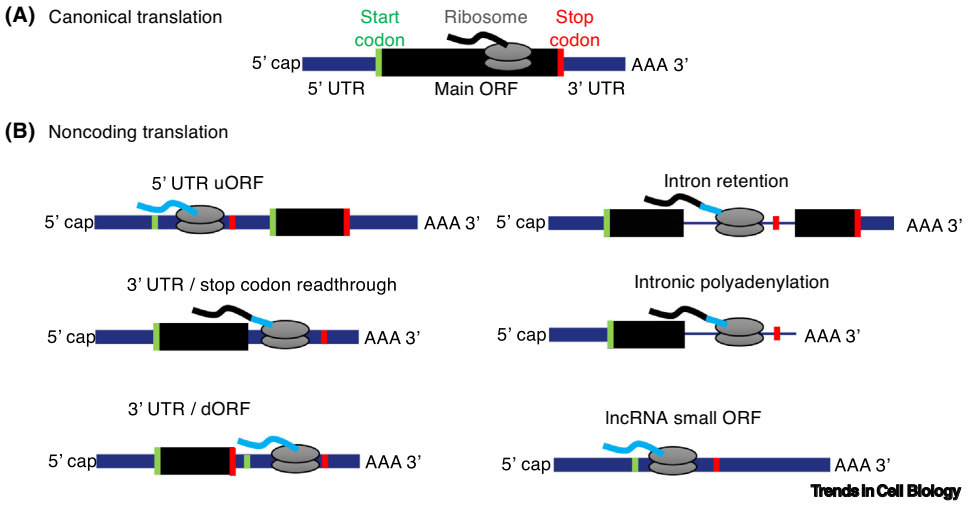

Figure 1. Noncoding translation. (A) In canonical translation, functional proteins are encoded by an open reading frame (ORF) (thick black horizontal bars) between a start codon (vertical green bars) and a stop codon (vertical red bars). Also shown are the ribosome and associated nascent polypeptide. Canonical untranslated regions (UTRs) are shown in dark blue. (B) Noncoding translation, or translation in annotated noncoding sequences. 5′ UTR: many 5′ UTRs contain one or more upstream ORFs (uORFs) that will be translated into small peptides (light-blue line associated with the ribosome). 3′ UTR: readthrough of the stop codon, caused by stop codon mutations or ribosome decoding errors, extends the main protein with a C-terminal tail (light blue). Ribosomes can also initiate translation in 3′ UTRs containing a downstream ORF (dORF). Intron: aberrant mRNA isoforms can be generated by either intron retention or intronic polyadenylation. In both cases, the original protein is truncated and extended at the C-terminal end with a peptide encoded by the 5′ end of the intron (light blue). Long noncoding RNA (lncRNA): Many lncRNAs contain a small ORF and are often translated into micropeptides (light blue).

that trigger surveillance and the protein factors that detect these signals, have remained elusive until recently [23,24].

Preventing noncoding translation

Translation is one of the most energy-demanding processes in cells [25]. It is therefore beneficial for cells to prevent the initiation of noncoding translation. While the total number of translated non-canonical ORFs outnumbers that of canonical ORFs when considering hundreds of human samples [15], the average translational activity (as measured by ribosome footprints) for each noncoding ORF is much lower than that of canonical ORFs. In addition, in any given sample, the majority (>80%) of ribosome footprints are mapped to canonical coding regions [15]. While the overall low levels of noncoding translation activity can be attributed in part to the relatively low abundance of most lncRNAs and aberrant mRNAs, separate mechanisms also exist to suppress the efficiency of translation in noncanonical ORFs.

The first such mechanism is the retention of many lncRNAs and aberrant mRNAs in the nucleus, preventing them from entering the cytoplasm where translation occurs (Figure 2A) [26,27]. One of the key signals that triggers nuclear retention is the presence of unspliced 5′ splice sites, recognized by the U1 small nuclear ribonucleoprotein (snRNP) [26]. U1 binding sites are present in lncRNAs and aberrant mRNAs due to inefficient splicing or mis-splicing (e.g., intron retention [28]). Another nuclear retention signal is the presence of **Alu elements** [27], which are enriched in lncRNAs and introns of the human genome. Heterogeneous nuclear ribonucleoprotein K (HnRNP K) recognizes a C-rich motif found in Alu elements, leading to nuclear RNA retention [27]. Many lncRNAs and aberrant mRNAs are also rapidly degraded in the nucleus by the nuclear **exosome**, although the mechanism underlying the specificity of exosome targeting remains poorly understood [29] (Figure 2B).

Glossary

Alu element: a highly abundant, primate-specific repeat/transposable element.
Autophagy-lysosome: a pathway that delivers protein aggregates, organelles, and other cytoplasmic components to the lysosome, a specialized organelle that degrades macromolecules to reuse the materials.
CCR4-NOT complex: a multisubunit complex involved in RNA degradation via deadenylation.
Disomes: two collided ribosomes on an mRNA template.
Exosome: a multiprotein intracellular complex capable of degrading various types of RNA.
Intronic polyadenylation: activation of an intronic poly(A) site leading to cleavage and polyadenylation within the intron, truncating the precursor mRNA.
Long noncoding RNAs (lncRNAs): Transcripts longer than 200 nucleotides with little or no protein-coding potential. Many lncRNAs contain one or multiple small ORFs.
Noncanonical ORFs: ORFs not included in commonly used databases of annotated protein-coding genes such as RefSeq and GENCODE.
Noncoding translation: translation occurring in annotated noncoding sequences, including introns, lncRNAs, and both 5′ and 3′ UTRs.
Nonsense-mediated mRNA decay (NMD): an mRNA surveillance pathway that degrades transcripts with premature stop codons or very long 3′ UTRs.
Open reading frames (ORFs): spans of DNA sequence between start and stop codons.
Rare codons/nonoptimal codons: codons used less frequently than other synonymous codons encoding the same amino acid.
Readthrough: the continuation of translation beyond a stop codon, extending into the 3′ UTR.
Ribosome profiling: a sequencing-based approach to map genome-wide footprints of ribosomes in cells.
Ubiquitin-proteasome system (UPS): the major proteolytic system in cells that marks soluble proteins for degradation by tagging substrates with ubiquitin.

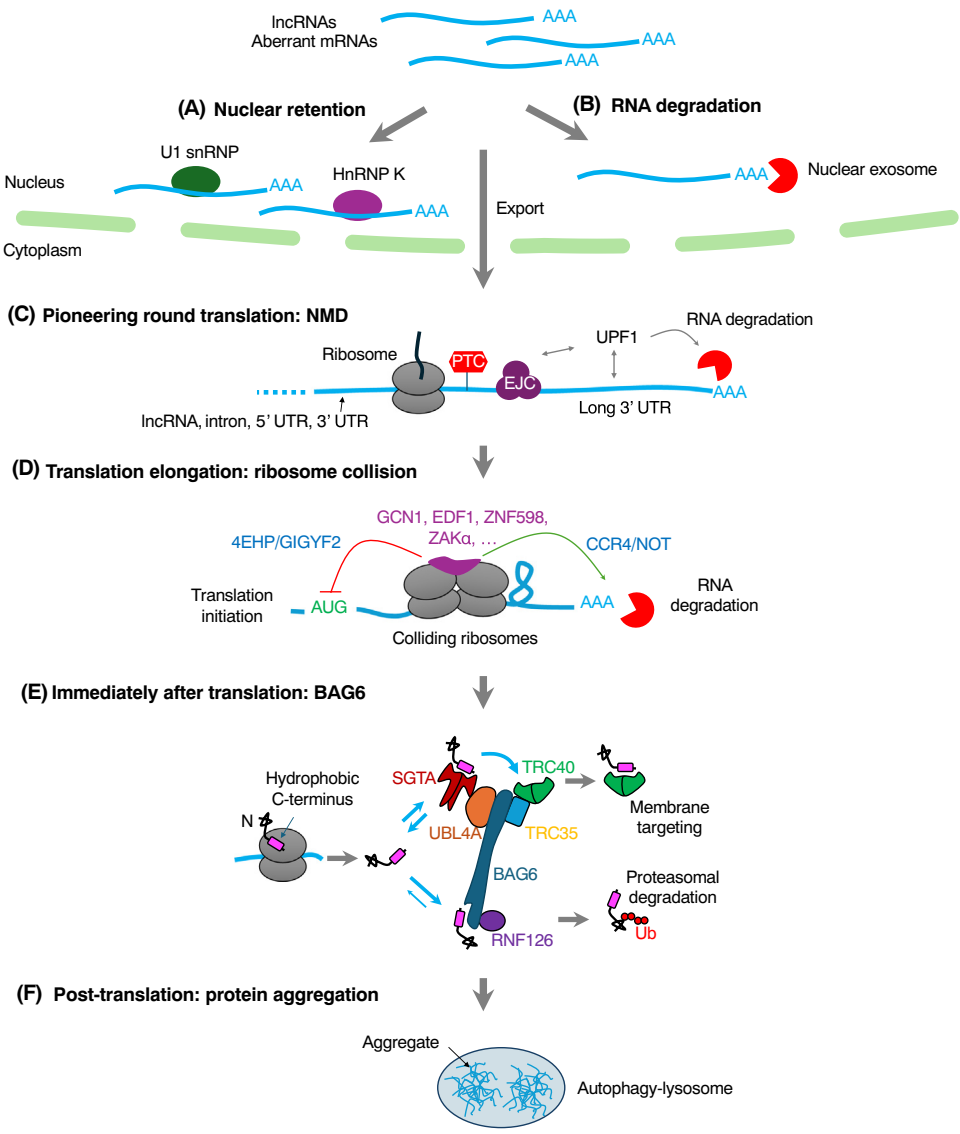

Figure 2. Mechanisms suppressing noncoding translation. (A) Many long noncoding RNAs (lncRNAs) and aberrant mRNAs are sequestered in the nucleus and prevented from being translated in the cytoplasm. (B) lncRNAs and aberrant mRNAs can be degraded in the nucleus by the nuclear exosome. (C) The pioneering round of translation can trigger nonsense-mediated mRNA decay (NMD) via the recognition of a premature termination codon (PTC) upstream of an exon junction complex (EJC) or by the detection of a long 3′ untranslated region (UTR) by UPF1. (D) Noncoding sequences enrich for roadblocks of ribosomes, such as RNA structures, resulting in ribosome stalling and subsequent collision, which can be detected by sensors including GCN1 (purple), leading to various downstream responses, such as CCR4/NOT-mediated RNA degradation and 4EHP/GIGYF2-mediated inhibition of translation initiation. (E) On the completion of translation, noncoding sequences are captured by the BCL2-associated athanogene 6 (BAG6) complex via a hydrophobic C-terminal tail, polyubiquitinated by RNF126, and then degraded by the proteasome. The BAG6 complex also mediates the membrane targeting of tail-anchored (TA) transmembrane proteins that contain a hydrophobic transmembrane domain at the C termini. (F) Aberrant proteins synthesized from noncoding sequences can form protein aggregates, which are often segregated and degraded in autophagy-lysosomes. Abbreviations: HnRNP K, heterogeneous nuclear ribonucleoprotein K; SGTA, small glutamine-rich tetratricopeptide repeat-containing protein α; snRNP, small nuclear ribonucleoprotein; Ub, ubiquitin.

Once escaped to the cytoplasm, most lncRNAs and aberrant mRNAs will inevitably be translated as they often carry a 5′ cap and a 3′ poly(A) tail, similar to canonical mRNAs. This is supported by the observation that when nuclear RNA surveillance is disrupted, lncRNAs accumulate in the cytosol, where they sequester ribosomes and inhibit global mRNA translation [30]. The rate of translation initiation in lncRNAs is potentially limited by nonoptimal 5′ leader sequences, including RNA structures and nonoptimal Kozak sequences at the start codon [31]. Many of these RNAs are expected to be detected and degraded after the pioneering round of translation via the **nonsense-mediated mRNA decay (NMD)** pathway [32] (Figure 2C). Canonical NMD is triggered by co-translational detection of exon junctions more than 50 nucleotides downstream of the stop codon [32]. Such signals are depleted from canonical mRNAs but are likely to be created in mis-spliced mRNAs due to premature stop codons being introduced by introns or frameshifted coding exons. In addition, a long 3′ UTR can also trigger NMD [33]. With a small ORF, lncRNAs are more likely to have splice junctions downstream of the stop codon, as well as a longer 3′ UTR. Similarly, translation in upstream ORFs (uORFs) located in 5′ UTRs often triggers NMD in mammalian cells, as in such instances most of the mRNA is effectively a long 3′ UTR [34].

While NMD is expected to degrade many noncoding translation substrates after the pioneering round of translation, it is rarely 100% efficient [35]. Furthermore, not all transcripts trigger NMD. For instance, mRNAs truncated by intronic polyadenylation (Figure 1B) will not trigger NMD because they do not contain a downstream exon junction. The same applies to mRNAs with the last intron retained. Additional mechanisms are required to suppress translation occurring in these types of noncoding sequences.

Roadblocks for elongating ribosomes in noncoding sequences

Ribosomes frequently encounter roadblocks, resulting in pausing or even stalling during translation elongation. The roadblocks for ribosomes include the presence of **rare or nonoptimal codons** [36] and codon pairs [37], codons encoding proline-rich peptides [38] and other ribosome-arresting peptides [39], poly(A) sequences [40], and RNA structures [41] among others [42]. Given that noncoding sequences have not evolved to be translated, they are likely to contain more obstacles that stall ribosomes than coding sequences. It is thus conceivable that the ribosome itself can sense the difference between coding and noncoding sequences by detecting these roadblocks, which then triggers surveillance mechanisms to terminate translation or degrade the nascent polypeptide. Ribosome stalling often results in a collision between the stalled ribosome and a trailing ribosome, generating a unique interface between the collided ribosomes. This interface allows the recruitment of sensor proteins that initiate a range of possible responses, including recycling the stalled ribosome, degrading the nascent polypeptide, inhibiting translation initiation, and degrading the mRNA (Figure 2D).

One of the sensors of ribosomal collision is the GCN1 protein [42,43]. Müller *et al.* [23] reported evidence supporting a model in which stop codon readthrough results in ribosome collisions in 3′ UTRs, and that the collided ribosomes subsequently recruit GCN1, which in turn recruits the **CCR4-NOT complex** to degrade the affected mRNA (Figure 2D). Supporting this model, selective ribosome profiling showed that GCN1-associated ribosomes, especially **disomes** (two collided ribosomes), are enriched in 3′ UTRs compared with canonical coding regions, presumably due to stochastic stop codon readthrough. The cause of ribosome stalling in 3′ UTRs remains unclear, although it was noted that 3′ UTRs are enriched for nonoptimal codons that are more likely to encode hydrophobic amino acids [23]. Beyond 3′ UTRs, it is unclear whether GCN1 is also recruited to suppress translation in 5′ UTRs, introns, or lncRNAs.

In addition to GCN1, there are several other ribosome collision sensors that are known to activate different surveillance responses. For instance, the ZNF598 and EDF1 proteins can both

independently sense ribosome collisions, leading to ribosome disassembly via the ASC-1 complex [44], and the degradation of the nascent polypeptide via the ribosome-associated quality control (RQC) pathway [45], and the inhibition of translation initiation via 4EHP-GIGYF1/2 [46] (Figure 2D). Under conditions of cellular stress, ZAKα also senses ribosome collisions and inhibits global translation initiation via eIF2α phosphorylation [47]. It remains to be seen whether any of these ribosome collision sensors are involved in the detection and suppression of noncoding translation.

A long queue of ribosomes can form if ribosome stalling is not resolved quickly enough. In human cells, ribosome stalling is usually cleared promptly. Live-cell imaging studies have shown that it takes about 5–8 min to clear ribosomes stalled by poly(A) sequences, one of the strongest roadblocks for ribosomes [48]. However, in about 10% of cases stalling is not cleared for over 30 min, and the resulting queue can contain tens of ribosomes [48]. Ribosome queueing has been proposed to explain how readthrough translation in human AMD1 mRNA inhibits the production of the readthrough protein [49]. The stop codon in the canonical ORF of AMD1 has a relatively high readthrough frequency, supported by a prominent peak of ribosome footprints around the next in-frame stop codon in the 3′ UTR, indicative of stalled ribosomes. It was proposed that a queue of ribosomes could extend up the entire length of the 3′ UTR and eventually reach the canonical stop codon or even the start codon, inhibiting the synthesis of the full-length AMD1 protein [49]. However, multiple lines of experimental evidence in another study argued against this model [24]. Notably, deleting the putative ribosome stalling sequence did not rescue the loss of the readthrough protein [24]. A genome-wide CRISPR screen using the AMD1 readthrough reporter also failed to identify any factors with known roles in ribosome collision sensing and clearing pathways [24]. Instead, this genetic screen identified BAG6-mediated proteasomal degradation as the main driver of AMD1 translation readthrough mitigation (see later). It remains to be seen whether ribosome queueing inhibits other noncoding translation events.

Proteasomal degradation

On the completion of translation, newly synthesized aberrant proteins need to be shielded or degraded as soon as possible to prevent them from interacting with other proteins in the cell. The **ubiquitin-proteasome system (UPS)** is the main machinery for the degradation of aberrant and misfolded proteins in cells. Early studies have suggested that peptides translated from noncoding sequences as a result of stop codon readthrough in yeast are eventually degraded by the UPS [50]. A subsequent study identified the proteasomal degradation of a readthrough mutant of the mouse cFLIP-L protein following ubiquitination by the TRIM21 protein [51]. Similarly, proteasomal degradation was found to be responsible for the loss of mutant readthrough proteins linked to hereditary disorders in the human PNPO and HSD3B2 genes [51]. Stop codon readthrough in the tumor suppressor protein SMAD4, whose loss plays a critical role in the development of pancreatic cancer, has been shown to cause near total loss of the SMAD4 protein via degradation by the UPS system [12]. Loss of the mutant SMAD4 is regulated by the ubiquitination of a highly hydrophobic ten-amino-acid degron located within a 40-amino-acid C-terminal extension in the mutant SMAD4 protein [12]. In addition to readthrough translation in 3′ UTRs, translation of a retained intron in yeast HAC1 mRNA also resulted in a ten-amino-acid C-terminal tail that triggers proteasomal degradation in a manner dependent on the putative SCF ubiquitin ligase F-box protein DUH1 [52]. Beyond these individual examples, it remains unclear whether the UPS plays a broad role in suppressing noncoding translation, and if so, what confers its specificity for noncoding translation products.

The BAG6 complex has emerged as the adaptor that mediates the proteasomal degradation of diverse noncoding translation products (Figure 2E). Two recent studies, one using unbiased

genetic screens in human cells and the other using a biochemical enrichment screen in worms [23,24], both uncovered the involvement of the ribosome-associated BAG6 complex in the degradation of translation readthrough proteins carrying a C-terminal extension encoded by 3′ UTR sequences. In this pathway, the C-terminal extended protein is captured by the BAG6 complex, polyubiquitylated by the E3 ubiquitin ligase RNF126, and subsequently degraded by the proteasome [23,24]. While both screens were performed using translation readthrough reporters, a high-throughput reporter assay comparing translation in thousands of human sequences in both wild-type and BAG6 knockout cells further showed that the BAG6 complex mediates the proteasomal degradation of all types of noncoding sequences, including 3′ UTRs, 5′ UTRs, lncRNAs, and introns, primarily by recognizing hydrophobic regions enriched in noncoding sequences [24].

The BAG6 complex has a previously established role in the triage of tail-anchored (TA) transmembrane proteins [53–56], a class of highly conserved proteins sharing a key feature with noncoding translation products: a hydrophobic C-terminal tail. While most transmembrane proteins contain an N-terminal signaling peptide allowing them to be co-translationally targeted to membranes, TA proteins have to be translated in the cytoplasm and then targeted to membranes after the C-terminal hydrophobic transmembrane domain is released from the ribosome. TA proteins require constant cellular monitoring and chaperoning as soon as translation is completed to prevent exposure of the hydrophobic domain to the cytosol, which is likely to lead to protein aggregation, mislocalization, or faulty interactions [53,56,57]. Two key chaperone proteins, small glutamine-rich tetratricopeptide repeat-containing protein α (SGTA) and BAG6 (also known as BAT3) [53,56] function in close physical proximity to translating ribosomes and both cooperate and antagonize each other in the decision to commit the nascent protein to membrane insertion or to ubiquitination and eventual proteasomal degradation [53,56,58]. In general, SGTA functions to capture and shield the C-terminal hydrophobic domain of the nascent peptide from the cytosol post-translationally, preventing ubiquitination by the E3 ligase RNF126 when recruited by BAG6. Commitment of authentic membrane proteins to membrane insertion involves transfer of the shielded nascent peptide from SGTA to TRC40, both of which are anchored at the C-terminal region of BAG6 via UBL4A and TRC35, respectively. TRC40 then directs the membrane protein for downstream insertion into endoplasmic reticulum (ER) membranes [55]. Alternatively, commitment to proteasomal degradation requires transfer of the nascent peptide from SGTA to the N-terminal region of BAG6, which binds the hydrophobic C-terminal tail and recruits RNF126 to ubiquitinate the targeted protein [59] (Figure 2E). In addition to defective TA proteins, BAG6 has also been implicated in the proteasomal degradation of other proteins carrying a hydrophobic C-terminal tail [60,61]. The mitigation of AMD1 3′ UTR translation strongly depends on BAG6, RNF126, and TRC35, and to a lesser extent on SGTA and UBL4A [24], suggesting that noncoding translation products are potentially recognized as defective or cryptic TA proteins.

A hydrophobic C-terminal tail appears to be a general feature of proteins translated from all types of noncoding sequences, especially when compared with functional coding sequences (excluding those encoding TA proteins) [24]. Three factors contribute to this difference between coding- and noncoding-derived proteins. First, when considering the variety of possible mechanisms of translation in various noncoding sequences, one unifying factor is that all mechanisms result in the C-terminal tail of the translated peptide being encoded by noncoding sequences [24] (Figure 1B, light blue in the nascent peptide linked to the ribosome). Second, noncoding sequences are more likely to encode hydrophobic amino acids [24]. This is due to a prominent bias in the genetic code where almost all hydrophobic amino acids are encoded by U-rich codons [24,62], and noncoding sequences are more U rich than coding sequences, which have evolved to be GC rich [24]. Compared with coding sequences, even random sequences have a much

higher propensity to be U rich and encode more hydrophobic amino acids. This explains why many random sequences cause prominent loss when fused to the C-terminal end of a reporter [22,24,49,63] and suggests that the BAG6 mechanism may act as a failsafe quality control system targeting 'unevolved' protein sequences. Last, hydrophobic amino acids are further depleted towards the C-terminal end of functional proteins (excluding TA proteins), presumably to avoid being targeted by BAG6 for proteasomal degradation [24]. Altogether, a C-terminal hydrophobic tail distinguishes aberrant proteins derived from all types of noncoding sequences and random sequences from functional proteins, allowing the BAG6 complex and potentially other pathways to selectively degrade noncoding translation products.

Protein aggregation and autophagy-lysosome degradation

When they fail to be cleared by the proteasome, misfolded proteins often aggregate with other misfolded proteins in cells [64]. The resulting protein aggregates are resistant to proteasomal degradation [64]. Instead, cells utilize a system known as the **autophagy-lysosome** pathway to segregate and remove protein aggregates from the cytosol [64,65] (Figure 2F). This system involves the sequestration of proteins into membrane-bound compartments known as autophagosomes, which are then fused with lysosomes, followed by digestion of the sequestered proteins by proteolytic enzymes contained in the lysosome [64].

Multiple studies have demonstrated protein aggregation induced by readthrough translation in 3′ UTRs, although it remains unclear whether these aggregates are substrates of the autophagy-lysosome pathway. For instance, nonstop mutations abolishing the stop codon of the gene REEP1 resulted in a C-terminal extension that causes the protein itself or a GFP extension reporter to form aggregate-like puncta in cells [66]. Similarly, a nonstop mutation in the *BRI* gene results in the generation of an amyloid-forming peptide associated with neuronal dysfunction and dementia [67]. In both cases, it is unclear whether those aggregates are localized in and cleared by the lysosome.

Subsequently, a study showed that EGFP extended with peptides encoded by a set of randomly selected human 3′ UTRs resulted in EGFP puncta reminiscent of protein aggregates in the lysosome, although no lysosomal marker was used in the same experiment [63]. Intriguingly, when cells were treated with the lysosomal inhibitor chloroquine, the whole-cell fluorescent intensity did not change for any of the 12 3′ UTR-derived peptides tested [63]. Western blotting of the reporter protein in both soluble and insoluble fractions confirmed that while 3′ UTR-encoded peptides move the reporter from the soluble to insoluble fraction, they did not change the total amount of the reporter protein. Similarly, lysosomal inhibition showed no effect in a high-throughput analysis of ~12 000 human sequences [24], nor did autophagy-lysosomal genes show any significant effect in the genome-wide CRISPR screen that identified BAG6 as a master regulator of noncoding translation mitigation [24]. These results suggested that, while aggregated proteins may eventually be degraded in the autophagosome/lysosome, this pathway does not necessarily shorten the half-life of these proteins. Instead, the autophagosome/lysosome functions more to isolate protein aggregates from other proteins in the cell. Alternatively, lysosomal inhibition may activate other redundant pathways, such as the proteasome, which has been shown to itself be degraded by the lysosome [68].

Concluding remarks

Maintaining a healthy proteome is of paramount importance for all cells and organisms. How cells avoid the production of aberrant proteins from noncoding sequences, which account for 98% of the human genome, is a key question of modern biology. Despite various transcriptome surveillance mechanisms that prevent ribosomes from translating noncoding sequences, noncoding

Outstanding questions

What is the relative contribution of the various mechanisms that suppress noncoding translation in cells?

To what extent do the mechanisms discovered for 3′ UTR readthrough translation mitigation (e.g., GCN1) also function in the mitigation of translation in lncRNAs, introns, and 5′ UTRs?

What sequence and structural features in a protein determine whether it is targeted for membrane insertion or proteasomal degradation in the BAG6 pathway?

What are the endogenous, noncoding-derived substrates of the BAG6 complex?

Is the BAG6 complex dysregulated in any diseases, and if so, are the symptoms related to the accumulation of aberrant proteins resulting from noncoding translation?

What causes ribosomes to stall in 3′ UTRs and other types of noncoding sequences?

Do cells send noncoding-derived proteins to plasma membranes or secrete them to extracellular spaces as mitigation mechanisms?

To what extent does ribosome collision-mediated inhibition of translation initiation play a role in suppressing noncoding translation?

translation is widespread, and potentially contributes to aging and various other diseases. Multiple mitigation mechanisms form a cascade of filters to reduce the production and/or accumulation of aberrant proteins derived from various types of noncoding sequences (Figure 2). Together, these mitigation mechanisms allow cells to faithfully synthesize a functional proteome encoded by a complex genome.

Several important questions remain to be addressed (see Outstanding questions). One of the key limitations of existing studies is that most studies have focused on stop codon readthrough and subsequent 3′ UTR translation. Beyond the BAG6 pathway, it remains unknown whether the other mitigation mechanisms discussed here also apply to translation in lncRNAs, introns, and 5′ UTRs. Furthermore, most studies focused on more or less arbitrarily selected noncoding sequences using reporter assays. Studies of endogenous noncoding translation substrates have been limited to 3′ UTR translation caused by stop codon mutations implicated in cancer and other diseases. Integrative analyses of ribosome profiling and proteomics data will likely uncover endogenous noncoding translation products from introns and lncRNAs. When combined with the unbiased genetic and biochemical approaches that identified the BAG6 complex's role in noncoding translation, future studies on endogenous substrates may lead to a deeper understanding of the mechanisms suppressing noncoding translation as well as its role in physiology and pathophysiology.

Last, it is worth noting that no surveillance mechanism is perfect. The detection of noncoding-derived proteins in mass spectrometry experiments indicates that some noncoding translation products escape all mitigation mechanisms in cells, including those discussed in this review (Figure 2). In some cases, the resulting aberrant protein contributes to cancer [21], highlighting the importance of noncoding translation mitigation mechanisms. By contrast, the imperfection of translation mitigation mechanisms may allow the evolution of new functional proteins, as exemplified by many functional micropeptides encoded by annotated lncRNAs [24]. Understanding how noncoding sequence-derived proteins escape surveillance could provide further insights into noncoding translation mitigation mechanisms.

Acknowledgments

X.W. is supported by the National Institutes of Health (NIH) Director's New Innovator Award (1DP2GM140977), NIH/National Heart, Lung, and Blood Institute (NHLBI) grant 1R01HL171664-01, Pershing Square Sohn Prize for Cancer Research, Pershing Square Foundation MIND Prize, Pew-Stewart Scholar for Cancer Research Award, Glenn Foundation Discovery Award, Impetus Longevity Grants, Million Dollar Bike Ride (MDBR) Grant, Ines Mandl Research Foundation (IMRF) grant, and a grant from the Cure Alzheimer's Fund.

Declaration of interests

X.W. is a member of the Scientific Advisory Board for Epitor Therapeutics. The remaining author has no interests to declare.

References

1. Djebali, S. *et al.* (2012) Landscape of transcription in human cells. *Nature* 489, 101–108
2. Jensen, T.H. *et al.* (2013) Dealing with pervasive transcription. *Mol. Cell* 52, 473–484
3. Wang, E.T. *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476
4. Pan, Q. *et al.* (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415
5. Derti, A. *et al.* (2012) A quantitative atlas of polyadenylation in five mammals. *Genome Res.* 22, 1173–1183
6. Ingolia, N.T. *et al.* (2014) Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* 8, 1365–1379
7. Ji, Z. *et al.* (2015) Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* 4, e08890
8. Weatheritt, R.J. *et al.* (2016) The ribosome-engaged landscape of alternative splicing. *Nat. Struct. Mol. Biol.* 23, 1117–1123
9. Sudmant, P.H. *et al.* (2018) Widespread accumulation of ribosome-associated isolated 3′ UTRs in neuronal cell populations of the aging brain. *Cell Rep.* 25, 2447–2456.e2444
10. Chen, J. *et al.* (2020) Pervasive functional translation of noncanonical human open reading frames. *Science* 367, 1140–1146
11. Calvo, S.E. *et al.* (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7507–7512

12. Dhamija, S. *et al.* (2020) A pan-cancer analysis reveals nonstop extension mutations causing SMAD4 tumour suppressor degradation. *Nat. Cell Biol.* 22, 999–1010
13. Wu, Q. *et al.* (2020) Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J.* 39, e104763
14. Bazin, J. *et al.* (2017) Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. *Proc. Natl. Acad. Sci. U. S. A.* 114, E10018–E10027
15. Yang, H. *et al.* (2024) Widespread stable noncanonical peptides identified by integrated analyses of ribosome profiling and ORF features. *Nat. Commun.* 15, 1932
16. Ruiz-Orera, J. *et al.* (2014) Long non-coding RNAs as a source of new peptides. *Elife* 3, e03523
17. Adusumalli, S. *et al.* (2019) Increased intron retention is a post-transcriptional signature associated with progressive aging and Alzheimer's disease. *Aging Cell* 18, e12928
18. Mazin, P. *et al.* (2013) Widespread splicing changes in human brain development and aging. *Mol. Syst. Biol.* 9, 633
19. Hsieh, Y.C. *et al.* (2019) Tau-mediated disruption of the spliceosome triggers cryptic RNA splicing and neurodegeneration in Alzheimer's disease. *Cell Rep.* 29, 301–316.e310
20. Dvinge, H. and Bradley, R.K. (2015) Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* 7, 45
21. Lee, S.H. *et al.* (2018) Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* 561, 127–131
22. Arribere, J.A. *et al.* (2016) Translation readthrough mitigation. *Nature* 534, 719–723
23. Muller, M.B.D. *et al.* (2023) Mechanisms of readthrough mitigation reveal principles of GCN1-mediated translational quality control. *Cell* 186, 3227–3244.e3220
24. Kesner, J.S. *et al.* (2023) Noncoding translation mitigation. *Nature* 617, 395–402
25. Buttgereit, F. and Brand, M.D. (1995) A hierarchy of ATP-consuming processes in mammalian cells. *Biochem. J.* 312, 163–167
26. Yin, Y. *et al.* (2020) U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* 580, 147–150
27. Lubelsky, Y. and Ulitsky, I. (2018) Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* 555, 107–111
28. Boutz, P.L. *et al.* (2015) Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* 29, 63–80
29. Kilchert, C. *et al.* (2016) The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* 17, 227–239
30. Ogami, K. *et al.* (2017) An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev.* 31, 1257–1271
31. Patraquim, P. *et al.* (2022) Translation and natural selection of micropeptides from long non-canonical RNAs. *Nat. Commun.* 13, 6515
32. Kurosaki, T. *et al.* (2019) Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* 20, 406–420
33. Hogg, J.R. and Goff, S.P. (2010) Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell* 143, 379–389
34. Hurt, J.A. *et al.* (2013) Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Res.* 23, 1636–1650
35. Linde, L. *et al.* (2007) Nonsense-mediated mRNA decay affects nonsense transcript levels and governs response of cystic fibrosis patients to gentamicin. *J. Clin. Invest.* 117, 683–692
36. Meijer, H.A. and Thomas, A.A. (2003) Ribosomes stalling on uORF1 in the *Xenopus* Cx41 5′ UTR inhibit downstream translation initiation. *Nucleic Acids Res.* 31, 3174–3184
37. Gamble, C.E. *et al.* (2016) Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell* 166, 679–690
38. Ivanov, I.P. *et al.* (2018) Polyamine control of translation elongation regulates start site selection on antizyme inhibitor mRNA via ribosome queuing. *Mol. Cell* 70, 254–264.e256
39. Cao, J. and Geballe, A.P. (1996) Coding sequence-dependent ribosomal arrest at termination of translation. *Mol. Cell. Biol.* 16, 603–608
40. Chandrasekaran, V. *et al.* (2019) Mechanism of ribosome stalling during translation of a poly(A) tail. *Nat. Struct. Mol. Biol.* 26, 1132–1140
41. Smirnova, A.M. *et al.* (2024) Stem-loop-induced ribosome queuing in the uORF2/ATF4 overlap fine-tunes stress-induced human ATF4 translational control. *Cell Rep.* 43, 113976
42. Yip, M.C.J. and Shao, S. (2021) Detecting and rescuing stalled ribosomes. *Trends Biochem. Sci.* 46, 731–743
43. Pochopien, A.A. *et al.* (2021) Structure of Gcn1 bound to stalled and colliding 80S ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2022756118
44. Juszkiewicz, S. *et al.* (2020) The ASC-1 complex disassembles collided ribosomes. *Mol. Cell* 79, 603–614.e608
45. Joazeiro, C.A.P. (2019) Mechanisms and functions of ribosome-associated protein quality control. *Nat. Rev. Mol. Cell Biol.* 20, 368–383
46. Meydan, S. and Guydosh, N.R. (2021) A cellular handbook for collided ribosomes: surveillance pathways and collision types. *Curr. Genet.* 67, 19–26
47. Vind, A.C. *et al.* (2020) ZAKα recognizes stalled ribosomes through partially redundant sensor domains. *Mol. Cell* 78, 700–713.e707
48. Goldman, D.H. *et al.* (2021) Live-cell imaging reveals kinetic determinants of quality control triggered by ribosome stalling. *Mol. Cell* 81, 1830–1840.e1838
49. Yordanova, M.M. *et al.* (2018) AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation. *Nature* 553, 356–360
50. Namy, O. *et al.* (2002) Translational readthrough of the PDE2 stop codon modulates cAMP levels in *Saccharomyces cerevisiae*. *Mol. Microbiol.* 43, 641–652
51. Shibata, N. *et al.* (2015) Degradation of stop codon read-through mutant proteins via the ubiquitin-proteasome system causes hereditary disorders. *J. Biol. Chem.* 290, 28428–28437
52. Di Santo, R. *et al.* (2016) The fail-safe mechanism of post-transcriptional silencing of unspliced HAC1 mRNA. *Elife* 5, e20069
53. Wunderley, L. *et al.* (2014) SGTA regulates the cytosolic quality control of hydrophobic substrates. *J. Cell Sci.* 127, 4728–4739
54. Shao, S. *et al.* (2017) Mechanistic basis for a molecular triage reaction. *Science* 355, 298–302
55. Hessa, T. *et al.* (2011) Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* 475, 394–397
56. Mariappan, M. *et al.* (2010) A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* 466, 1120–1124
57. Buchberger, A. *et al.* (2010) Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol. Cell* 40, 238–252
58. Leznicki, P. and High, S. (2020) SGTA associates with nascent membrane protein precursors. *EMBO Rep.* 21, e48835
59. Rodrigo-Brenni, M.C. *et al.* (2014) Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol. Cell* 55, 227–237
60. Minami, R. *et al.* (2010) BAG-6 is essential for selective elimination of defective proteasomal substrates. *J. Cell Biol.* 190, 637–650
61. Zhang, Z. *et al.* (2023) Elucidation of E3 ubiquitin ligase specificity through proteome-wide internal degron mapping. *Mol. Cell* 83, 4191–4192
62. Wolfenden, R.V. *et al.* (1979) Water, protein folding, and the genetic code. *Science* 206, 575–577
63. Kramarski, L. and Arbely, E. (2020) Translational read-through promotes aggregation and shapes stop codon identity. *Nucleic Acids Res.* 48, 3747–3760
64. Kopito, R.R. (2000) Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* 10, 524–530
65. Leeman, D.S. *et al.* (2018) Lysosome activation clears aggregates and enhances quiescent neural stem cell activation during aging. *Science* 359, 1277–1283
66. Bock, A.S. *et al.* (2018) A nonstop variant in REEP1 causes peripheral neuropathy by unmasking a 3′ UTR-encoded, aggregation-inducing motif. *Hum. Mutat.* 39, 193–196
67. Vidal, R. *et al.* (1999) A stop-codon mutation in the BRI gene associated with familial British dementia. *Nature* 399, 776–781
68. Cuervo, A.M. *et al.* (1995) Degradation of proteasomes by lysosomes in rat liver. *Eur. J. Biochem.* 227, 792–800