Trends in Biochemical Sciences

CellPress
REVIEWS

Review

Function and Evolution of Upstream ORFs in Eukaryotes

Hong Zhang,[1,2] Yirong Wang,[1,2] and Jian Lu[1,*]

There is growing interest in the role of translational regulation in cellular homeostasis during organismal development. Translation initiation is the rate-limiting step in mRNA translation and is central to translational regulation. Upstream open reading frames (uORFs) are regulatory elements that are prevalent in eukaryotic mRNAs. uORFs modulate the translation initiation rate of downstream coding sequences (CDSs) by sequestering ribosomes. Over the past several years, genome-wide studies have revealed the widespread regulatory functions of uORFs in different species in different biological contexts. Here, we review the current understanding of uORF-mediated translational regulation from the perspective of functional and evolutionary genomics and address remaining gaps that deserve further study.

Highlights

Recent advances in RNA sequencing and ribosome profiling allow the quantitative study of uORFs at the genomic scale.

Most uORFs in eukaryotic genomes show evidence of translation, and the translational efficiency of uORFs and their impact on the translation of downstream CDSs are influenced by the sequence context of uORFs.

Both positive Darwinian selection and purifying selection have shaped the genome-wide landscapes of uORFs in eukaryotes.

uORFs can dynamically modulate the translation of downstream CDSs in various biological contexts.

uORFs Have an Essential Role in Eukaryotic Translational Regulation

Eukaryotic gene expression is tightly regulated at both the transcriptional and translational levels. Compared with transcriptional regulation, translational control allows more immediate responses to adjust the protein abundance upon cellular signals or environmental stimuli in a variety of biological processes [1–3]. Translation of an mRNA comprises four steps: initiation, elongation, termination, and ribosome recycling. In eukaryotes, translation of most mRNAs is initiated via the cap-dependent scanning mechanism [4–6] (Box 1). Translation initiation is the major step that determines the rate of protein biosynthesis and is regulated by multiple mechanisms [1–3].

u**ORFs** (see Glossary) are short ORFs in the **5′ untranslated regions** (5′ UTRs) of eukaryotic mRNAs the start codons of which are located upstream of the CDS start codons [7]. uORFs have an essential role in modulating the translation initiation of CDSs [3,8] and, with the advent of **ribosome profiling** [9–11], genome-wide studies of translated uORFs have become possible under physiological and pathological conditions and for a range of species [10–18]. These studies have considerably expanded our understanding of the regulatory functions of uORFs. Here, we review the current understanding of the regulatory mechanisms, functional consequences, and evolutionary principles of uORFs at the genomic scale. Finally, we discuss the challenges and gaps in our knowledge of uORF-mediated regulation that remain to be addressed in future studies.

Regulatory Mechanisms of Translation by uORFs

As previously stated, uORFs have essential roles in regulating the translation of CDS in eukaryotes [3,8]. There are several possible scenarios when a pre-initiation complex (PIC; Box 1) scans along a 5′ UTR that contains canonical (beginning with an AUG start codon and ending with a UAA/UAG/UGA stop codon) uORFs (Figure 1A). When the scanning PIC fails to recognize the canonical start codon of a uORF (uAUG), which is located in an unfavorable **Kozak sequence** context, it continues to search for the next start codon (‘leaky scanning’, Figure 1Ai) [19]. If a uAUG is recognized by the PIC, the complete **80S ribosome** will be assembled to translate the uORF and potentially block other scanning PICs [20]. After termination at the uORF stop codon, the **40S small ribosomal subunit** may remain bound to the 5′ UTR and re-initiate translation at the

[1]State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, 100871, China
[2]These authors contributed equally

*Correspondence:
LUJ@pku.edu.cn (J. Lu).

CDS start codon after reacquiring a **ternary complex** (TC) ('re-initiation'; Figure 1Aii). Occasionally, the ribosome may stall at the uORF stop codon and trigger **nonsense-mediated decay** (NMD; 'stall') [21], or both the 40S and **60S ribosomal subunits** may dissociate from the mRNA ('drop-off'; Figure 1Aiii). Either mechanism will impede translation initiation of the downstream CDS. Overall, the translation of canonical uORFs attenuates translation initiation of the annotated CDSs by sequestering ribosomes [3,8].

Based on the positions of the start and stop codons, a uORF can be classified into one of three categories: (i) nonoverlapping uORFs, the stop codons of which are upstream of the AUG start codons of the annotated CDSs (cAUGs); (ii) out-of-frame overlapping uORFs (oORFs), the stop codons of which are located downstream of cAUGs and in different reading frames [16,22]; and (3) N-terminal extensions, which are in fact in-frame overlapping uORFs that share the same stop codon with the annotated CDSs (Figure 1B). The downstream CDS can be engaged by the re-initiating PIC only when the translated uORFs do not overlap with CDSs [23].

Ribosome Profiling Facilitates Genome-Wide Characterization of Functional uORFs

The recently developed ribosome-profiling technique [9–11] has expanded our understanding of the possible regulatory functions of uORFs at the genome-wide level. In a ribosome profiling experiment, cells are first lysed in buffer containing cycloheximide (CHX), which prevents ribosome elongation (Figure 2Ai). Then the cell lysate is treated with RNase [or micrococcal nuclease (MNase)] to digest mRNA regions that are not covered by ribosomes, and the resultant **monosomes** are separated from ribosome subunits and polysomes in a sucrose gradient via ultracentrifugation (Figure 2Aii). Finally, ribosome-protected mRNA fragments (RPFs), usually of ~30 nucleotides in length, are released from the ribosomes and subjected to normal **RNA sequencing** (RNA-seq) procedures (Figure 2Aiii) [24,25]. The RPFs accurately indicate the exact mRNA regions that are being translated by ribosomes. Given that each RPF corresponds to a translating ribosome, the translation rate of an ORF (uORF or CDS) should be proportional to the density of RPFs along that ORF. Thus, one can measure the translational efficiency of an ORF by contrasting RPF density with mRNA abundance. Ribosome profiling has since been applied for the genome-wide characterization of uORFs that are associated with ribosomes in a range of species, including yeast [10], maize [12], *Arabidopsis* [13], flies [14,15], zebrafish [16,17], mice [11], and humans [18]. Collectively, these studies demonstrate that uORFs are prevalently translated in eukaryotic cells.

To provide further evidence that uORF-associated RPFs result from translation events, several groups have modified the ribosome-profiling method to detect the initiating ribosomes after treating cells with harringtonine [11,14] or lactimidomycin (LTM) [26,27]. Harringtonine and LTM inhibit the first cycle of translation elongation, thus stalling the ribosomes at translation initiation sites (Figure 2B) [28]. These studies have confirmed that thousands of uORFs are translated [11,14,26]. Moreover, these studies have demonstrated that uAUGs generally show lower ribosomal occupancy peaks compared with cAUGs, because the latter are located in more-optimized sequence contexts than the former. Ribosome profiling also reveals that many near-cognate (beginning with non-AUG start codons) uORFs are translated [11,27,29,30], which is further supported by a recent study that analyzed protein N termini through mass spectrometry and found many translation initiation events from non-AUG start codons in 5′ UTRs [31]. Overall, ribosome-profiling techniques have enabled us to capture the genome-wide translation of uORFs with high sensitivity and accuracy.

Genomic Features That Affect the Repressiveness of uORFs in the Translation of CDSs

It is well established that, at the genomic scale, genes containing uORFs exhibit significantly lower translational efficiencies in CDS regions compared with genes without uORFs [14,16,17,32,33].

Glossary

5′ untranslated regions (5′ UTRs): mRNA regions that are upstream of the start codon of the main ORFs.
40S small ribosomal subunit: the small subunit of a eukaryotic ribosome. '40S' denotes its sedimentation coefficient during centrifugation in Svedberg units.
60S ribosomal subunit: the large subunit of a eukaryotic ribosome. It has a sedimentation coefficient of 60S.
80S ribosome: the eukaryotic ribosome, which has a sedimentation coefficient of 80S.
Alternative splicing: a process in which the same nascent RNA molecules are processed into different transcripts.
***cis*-regulatory elements (CREs):** sequence elements that only regulate the same mRNA where they are located.
Fixed: the frequency of an allele reaches 100% in the population.
Kozak sequence: nucleotides flanking the start codon of an ORF (usually from –6 to +4 given that the first nucleotide of ORF start codon is +1). The consensus Kozak sequence of CDSs in vertebrates is GCCRCCAUGG (where R can be A or G).
miRNAs: small noncoding RNAs that form a complex with Argonaute proteins and bind to target sites by seed-pairing. miRNAs are usually ~22 nucleotides in length.
Monosome: complex formed by a single ribosome and the associated mRNA fragment.
Nonsense-mediated decay (NMD): a surveillance pathway that degrades aberrant transcripts with premature stop codons.
Open reading frame (ORF): a sequence of triplets encoding amino acids. An ORF begins with a start codon and is bounded by a stop codon.
Polymorphic uORFs: uORFs that are present in some but not all individuals in the population.
Ribosomal tunnel: a passage in the large ribosomal subunit where the nascent peptide chain exits from the ribosome.
Ribosome profiling: a recently developed technique that could separate ribosome-protected RNA fragments for sequencing. It is widely used in studies of mRNA translation.
RNA-sequencing (RNA-seq): a high-throughput method for determining the sequences of RNA molecules in the transcriptome.

Deletion of uORFs by genome editing further confirmed the hypothesis that uORFs inhibit the translation of downstream CDSs primarily by sequestering ribosomes [34]. For example, a previous study identified 35 735 uORFs in *Drosophila melanogaster* and showed that these uORFs were not evenly distributed across genes: 49.4% genes did not have any uORF, and 15.2% and 35.4% genes had one uORF and at least two uORFs, respectively [14]. Extensive ribosome profiling and mRNA-seq experiments conducted during the life cycle of *D. melanogaster* demonstrated that 32 224 (90.2%) out of 35 735 expressed uORFs showed evidence of translation in at least one developmental stage of *D. melanogaster* [14]. It is also estimated that the translational efficiency of CDS was 8.38–30.4% lower for genes containing a single translated uORF and 18.4–60.7% lower for genes with multiple translated uORFs in *D. melanogaster* [14]. These results suggest that uORFs cumulatively inhibit the translation of the annotated CDSs. Interestingly, cAUGs are generally located in unfavored sequence contexts if a 5′ UTR has multiple uORFs, suggesting that uORFs are used by specific genes to ensure a low basal translation level via double-negative controls [35].

Previous studies demonstrated that contextual characteristics can affect the ribosomal occupancies of uORFs and, hence, the repressive efficiencies of uORFs in the translation of the annotated CDSs [11,14,16,17,32,33,36–38]. uAUGs are generally located in disfavored Kozak sequence contexts (Figure 3A) compared with cAUGs [14,17,37], and uORFs with higher translational efficiencies tend to exhibit more preferred sequence contexts around their start codons [14,17]. Although the length of a uORF usually has a limited effect on its repressive efficiency, a uORF with a start codon that is distant from the 5′ cap or a stop codon that is close to the cAUG tends to be more repressive [14,16,17,39]. Notably, uORFs showing out-of-frame overlap with downstream CDSs are more repressive because translational re-initiation of the downstream CDS is unlikely once the out-of-frame overlapping uORF in the same RNA molecule is translated [16,22]. Moreover, the sequence composition of uORFs can be optimized for ribosome association efficiency, and the higher coding potential of a uORF is associated with higher translational efficiency of the uORF [14]. The importance of the genomic features of uORFs is also manifested by their evolutionary patterns. Since a uAUG is the most important definitive feature of a uORF, uORFs with more conserved start codons are more repressive [14,39]. Notably, after multiple regression analysis of the translational efficiency of CDSs with various genomic features of uORFs in *D. melanogaster*, it was found that the most important features influencing the repressiveness of uORFs in downstream CDSs were the optimized Kozak context around uAUGs, high conservation levels of uAUGs, and long distances between a uAUG and the 5′ cap [14].

The Origins and Evolution of uORFs in Eukaryotic Genomes

Kozak first posited that uORFs overall are deleterious and should be depleted in 5′ UTRs because they impair translation of the downstream CDSs [19]. Moreover, protein translation is a highly energy-consuming process [40], and excessive uORF translation will lead to a tremendous burden of energy provision in cells. Accordingly, genome-wide analyses have demonstrated that there are significantly fewer ATG triplets in 5′ UTRs than expected by chance in a range of species [35,41,42], which suggested that uORFs tend to be depleted from 5′ UTR due to their deleterious effects. Moreover, the mutations that generate **polymorphic uORFs** in *D. melanogaster* exhibit significantly lower derived allele frequencies than neutral mutations, suggesting that they are generally selected against [14].

By contrast, the uORFs that are preserved in genomes are usually evolutionarily more conserved than expected under the assumption of neutral evolution [42,43]. In line with these observations, the polymorphic uORFs of human populations resulting from the creation of new uORFs or abolition of pre-existing uORFs are potentially associated with human diseases (Box 2). Intriguingly, it was shown for the first time that a considerable fraction of uORFs that have recently been fixed in

Ternary complex (TC): the complex formed by eIF2, GTP, and Met-tRNA$_i$. It is required not only for the assembly of the pre-initiation complex, but also for re-initiation.
Three-nucleotide periodicity: a pattern that repeats every three nucleotides. Here, it refers to the 'high-low-low' pattern of RPF 5′ end coverage in a codon.
Transposable elements (TEs): genomic elements that could move into new locations in the genomes.

Box 1. Scanning Mechanism of Eukaryotic Translation Initiation

Most eukaryotic mRNAs are translated in a cap-dependent manner (Figure I) [4–6]. Before translation initiation, GTP-bound eukaryotic initiation factor 2 (eIF2) first binds Met-tRNA$_i$ to form a ternary complex (TC). Then, TC binds the 40S small ribosomal subunit for assembly into the 43S preinitiation complex (PIC), with the assistance of eIF1, eIF1A, eIF3, and eIF5. Additionally, the cap-binding protein eIF4E binds to the 5′ cap of mRNA and recruits eIF4G and eIF4A. eIF4G activates the ATP-dependent helicase activity of eIF4A, which unwinds secondary structures near the 5′ cap of the mRNA. Next, the 43S PIC binds to the pre-activated 5′ end of the mRNA, scans the mRNA base by base, and searches for the AUG triplet complementary to the anticodon of Met-tRNA$_i$. Upon successful recognition of the AUG start codon, the 60S subunit is recruited to assemble the complete 80S ribosome to initiate translation. During elongation, the ribosome processes the ORF codon by codon and the nascent peptide will emerge from peptide exit tunnel (Figure I). Upon termination, the newly synthesized peptide will be released, and the ribosomal subunits will be recycled.

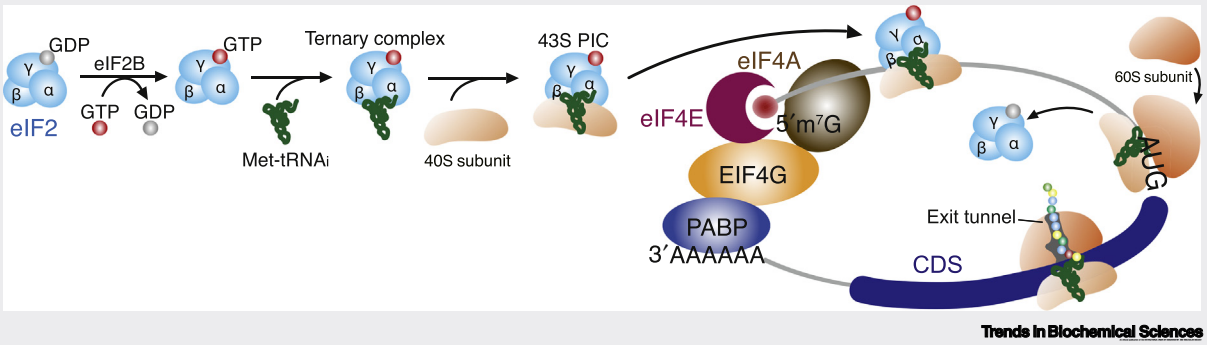


Figure I. The Scanning Model of Eukaryotic Translation Initiation. Eukaryotic initiation factor 2 (eIF2), GTP, and methionyl initiator transfer RNA (Met-tRNA$_i$) assemble into a ternary complex (TC), which binds the 40S subunit to form a pre-initiation complex (PIC). The PIC binds the 5′ end of mRNAs activated by the eIF4F complex (eIF4G, eIF4E, and eIF4A) and scans from the 5′ end to the 3′ end in search of AUG codons. Upon successful recognition of the AUG start codon, GTP is hydrolyzed to GDP and dissociates from the PIC with eIF2. The 60S subunit joins the complex to form a complete ribosome and translate the open reading frame (ORF). The ribosomal exit tunnel is a passage in the large ribosomal subunit where the nascent peptide chain exits from the ribosome.

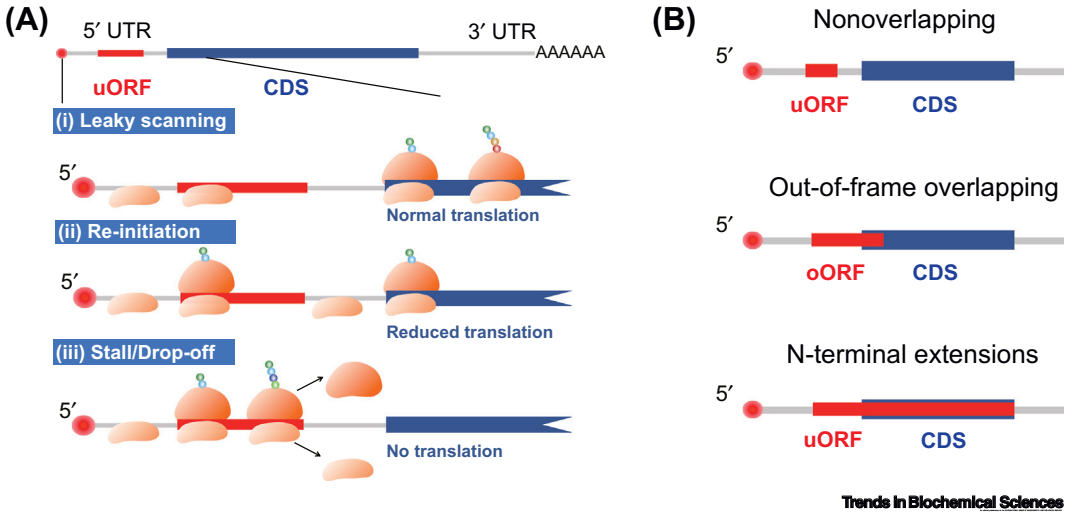


Figure 1. Upstream Open Reading Frame (uORF)-Mediated Translational Regulation. (A) The three scenarios when a pre-initiation complex (PIC) scans along a uORF-containing 5′ untranslated region (5′ UTR). (i) The scanning PIC fails to recognize the start codon of the uORF and continues to search for next start codon ('leaky scanning'). In this case, the normal translation of the downstream coding sequence (CDS) is not affected. (ii) The PIC successfully recognizes the start codon of the uORF and assembles into a full ribosome for translation. After termination at the uORF stop codon, the 40S subunit remains bound to the mRNA and re-initiates translation at the CDS start codon ('re-initiation'). The translation of the uORF will interfere with other scanning ribosomes [20], leading to reduced CDS translation. (iii) Upon translation termination at the uORF stop codon, the ribosome may stall and trigger nonsense-mediated decay (NMD) ('stall'), or dissociate from the mRNA ('drop off') without translating the CDS. (B) Classification of uORFs based on the positions of their start codons and stop codons.

the *D. melanogaster* population are advantageous and favored by natural selection [14]. Hence, a unifying model of uORF evolution has been proposed in which newly formed uORFs that are deleterious will be selected against and segregate at low frequencies in the population, whereas beneficial newly formed uORFs will be rapidly **fixed** in populations under positive Darwinian selection. After fixation, functional uORFs will be maintained by natural selection during evolution, due to functional constraints (Figure 3B). Notably, positive selection is facilitated by the large effective population size of *D. melanogaster*, which makes natural selection more efficient [44]. The evolutionary forces governing the evolution of uORFs in organisms with smaller effective population sizes, such as humans [44], remain to be further explored.

In addition to point mutations, the insertion of **transposable elements** (TEs) can generate new uORFs [45,46]. Reporter assays indicated that TE-derived uORFs in genes such as *RPP38* and *GJB3* are functional [45,46]. Ribosome profiling data also support the translation of TE-derived uORFs in *RPP38* and *GJB3* (Figure 3C). In humans, ~10% (3992 of 39 786) of uORFs are derived from TEs [46]. Most TE-derived uORFs result from Alu, L1, MIR, and L2 elements, which is consistent with their prevalence in the human genome [46]. It is well documented that TEs are rich

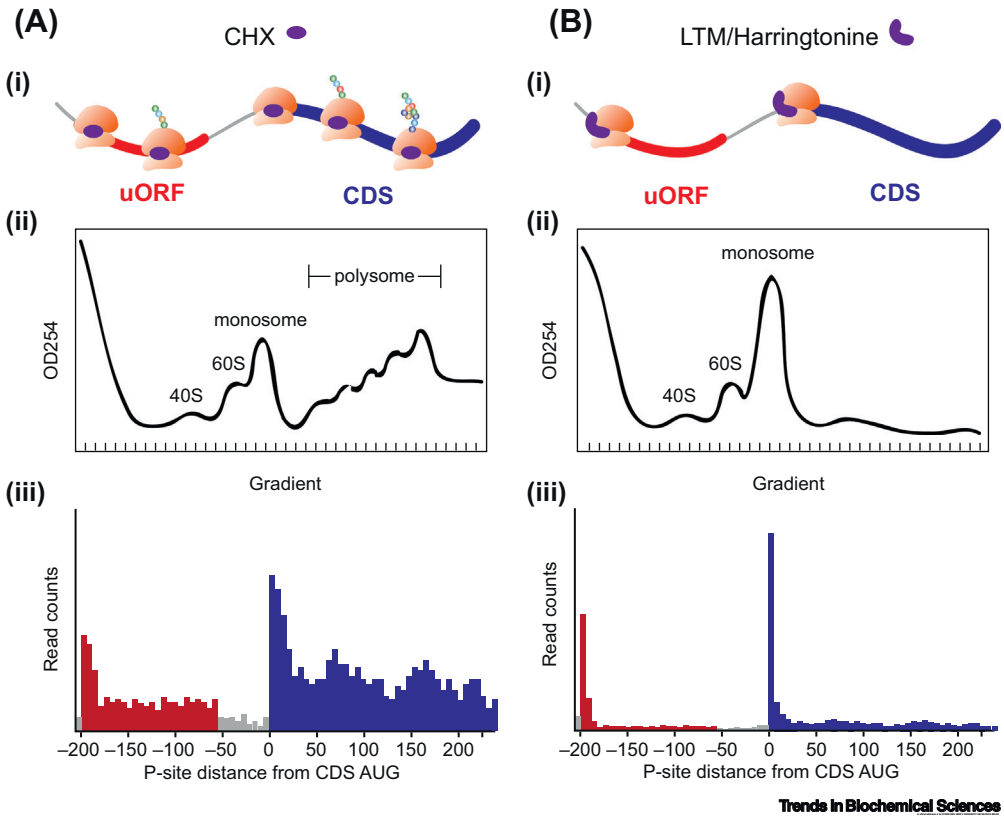


Figure 2. Effect of Cycloheximide (CHX) (A) and Lactimidomycin (LTM) (B) on Ribosome Profiles. (A) CHX binds to the exit site of a ribosome and inhibits translational elongation (i) [28]. Therefore, polysomes are well preserved by CHX treatment before digestion (ii). After digestion with either RNase I or micrococcal nuclease (MNase), monosomes are separated by ultracentrifugation, and ribosome-protected fragments (RPFs) are released from monosomes and sequenced (iii). If the samples are digested with the commonly used RNase I, the 5′ end location of RPFs in the mRNA usually shows **three-nucleotide periodicity**, because ribosomes process coding sequences (CDSs) codon by codon [9,10]. However, three-nucleotide periodicity is compromised due to cutting bias if MNase is used to digest ribosome-bound mRNAs [14,15]. (B) LTM or harringtonine only inhibits the first cycle of translation elongation (i) [28]. Therefore, after LTM or harringtonine treatment, only ribosomes at translation initiation sites are associated with mRNAs, leading to pronounced monosomes before digestion (ii). With subsequent purification and sequencing of RPFs, most reads are enriched at initiation sites (iii). Abbreviation: uORF, upstream open-reading frame.

sources of transcriptional regulatory motifs [47,48]. These new results suggest that TE domestication also has an important role in translational regulation.

uORF-Mediated Translation Regulation during Development

Since **alternative splicing** or altered transcriptional initiation is prevalent across the developmental stages of eukaryotes [49], it is possible that the uORF-mediated regulation is exploited to fine-tune the translational program during development [50]. For example, during meiosis in budding yeast, kinetochore inactivation occurs via reduction of Ndc80 abundance, which is achieved by transcribing an alternative, longer *Ndc80* mRNA isoform that is translationally inhibited by the uORFs [51]. Additionally, through the simultaneous quantification of transcriptomes, translatomes, and proteomes during the meiotic differentiation of budding yeasts, Cheng *et al.* [52] demonstrated that pervasive translational reprogramming of the CDS

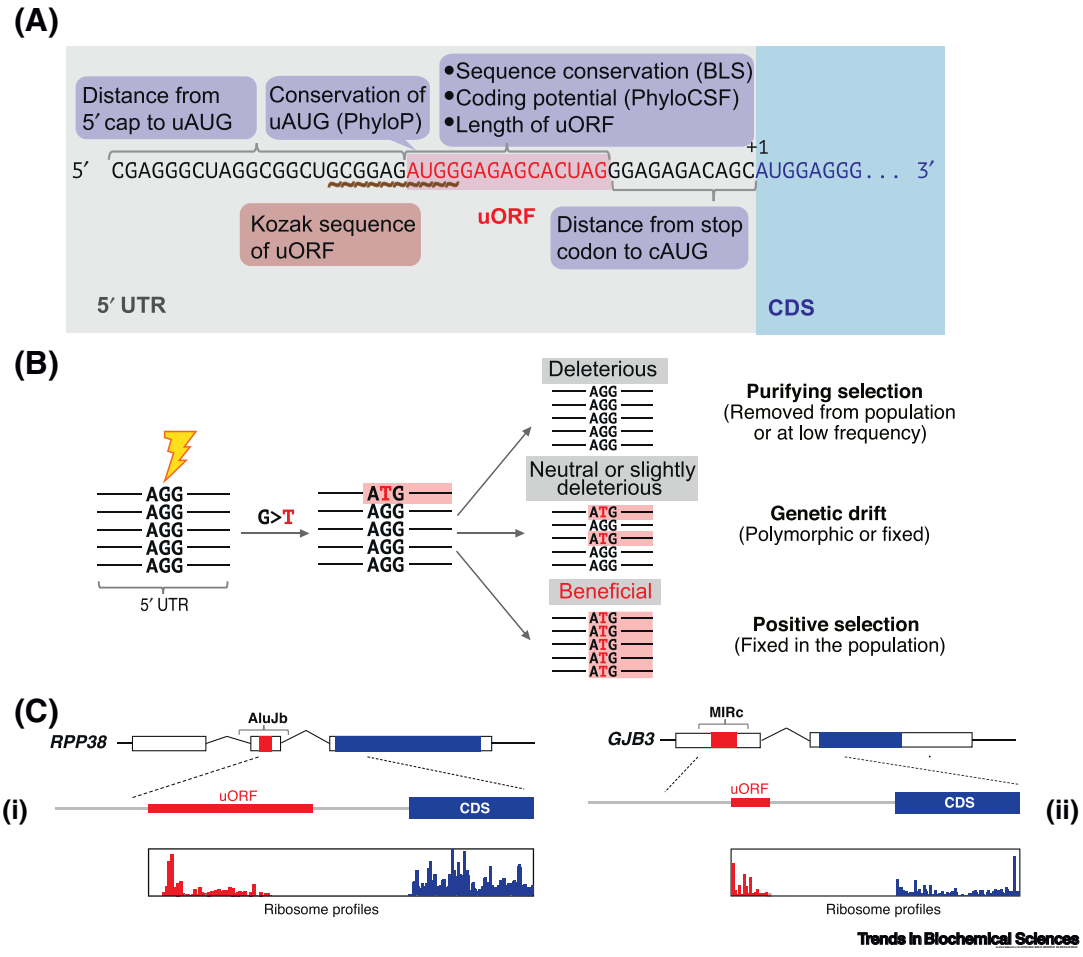


Figure 3. Genomic Features and Evolutionary Principles of Upstream Open Reading Frames (uORFs). (A) Illustration of different features of uORFs that influence their translation and repression efficiencies on the downstream coding sequences (CDSs). PhyloP is a frequently used index for measuring the conservation of individual nucleotide sites based on multiple sequence alignment [86]. (B) A model of uORF evolution. Mutations frequently generate novel uORFs (uAUGs) in 5′ untranslated regions (5′ UTRs). For example, a new uORF created by G>T mutation in an individual might be deleterious, neutral, or advantageous. The uORF will be removed by natural selection or persist in the population at low frequencies if highly detrimental or will randomly drift in the population if neutral or slightly deleterious. However, it will be favored by natural selection and become fixed in the population rapidly if beneficial. (C) Example of uORFs derived from transposable elements (TEs) that are supported by ribosome-protected fragments (RPFs). Exon2 of *RPP38* is contributed by an *AluJb* element [45] (i). The translation of the *AluJb*-derived uORF in exon2 is demonstrated by RPFs. In another example, a fragment in exon2 of *GJB3* is derived from an *MIRc* element [46] (ii). A uORF in the *MIRc*-derived fragment is also translated. The RPF coverage for both examples was obtained from the GWIPS-viz database [87]. Adapted from [14] (B).

Box 2. uORFs and Human Diseases

The number of uORFs varies among humans, and mutations that create new uORFs or disrupt pre-existing uORFs can cause diseases [39,88–91]. For instance, a point mutation in the 5′ UTR of *CDKN2A* introduces an oORF, which leads to reduced production of the CDKN2A protein and a predisposition for melanoma [92] (Figure IA). Moreover, a point mutation in an intronic splicing donor site of *THPO* leads to exon skipping (ΔE), which eliminates a uORF located in that exon and causes two overlapping uORFs to be fused with the downstream CDS (Figure IB). Although the ΔE transcript encodes a different N-terminal signal peptide from the wild-type protein and the sequence of mature thrombopoietin (TPO) is unaffected, the ΔE transcript produces significantly more TPO than the wild-type allele and causes thrombocythemia in patients who carry this mutation [93]. In addition to these well-known disease-related uORF-altering mutations, genome-wide scanning of variations in human populations or tumor samples has identified many uORF-generating or uORF-disrupting mutations [39,89,94]. While the translational changes in downstream CDSs have been verified experimentally in some cases, the causal relationships between these mutations and the associated diseases remain to be further explored.

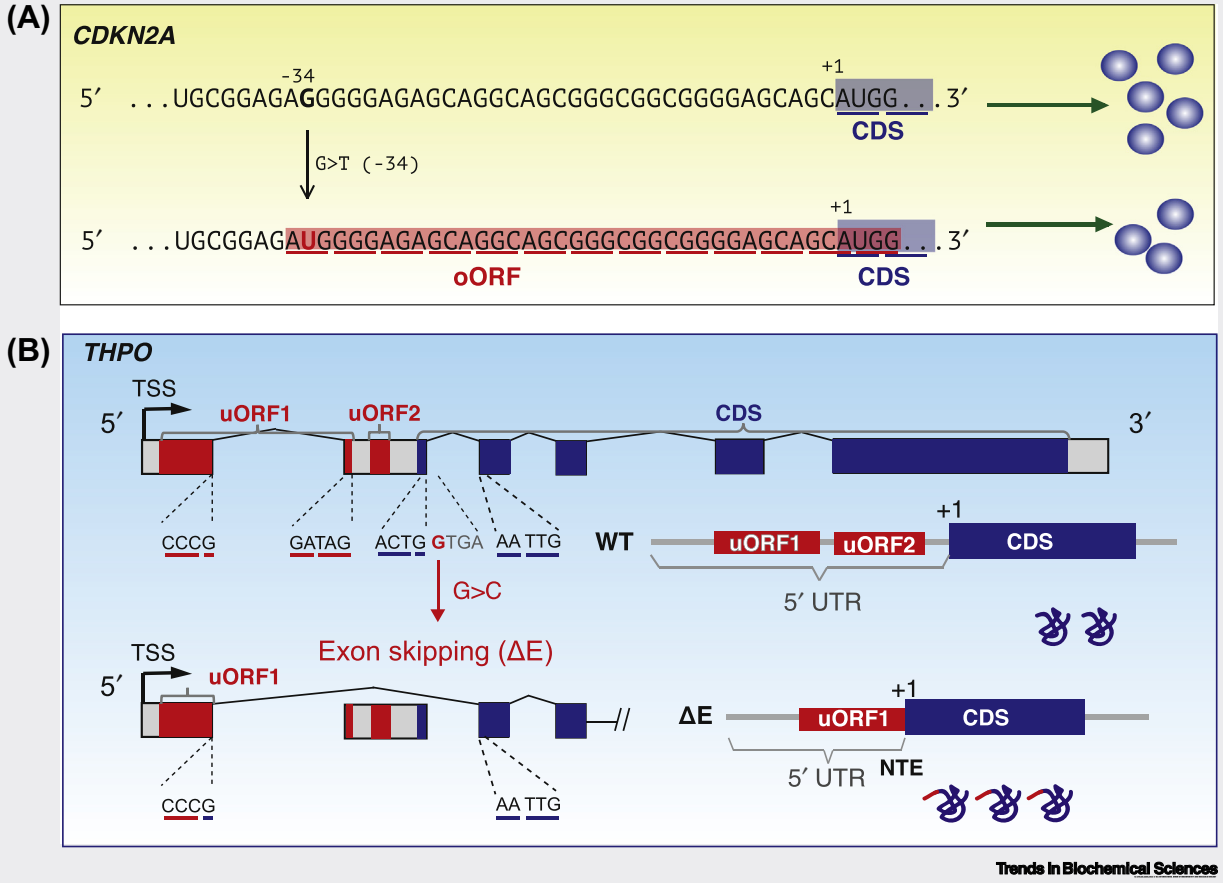
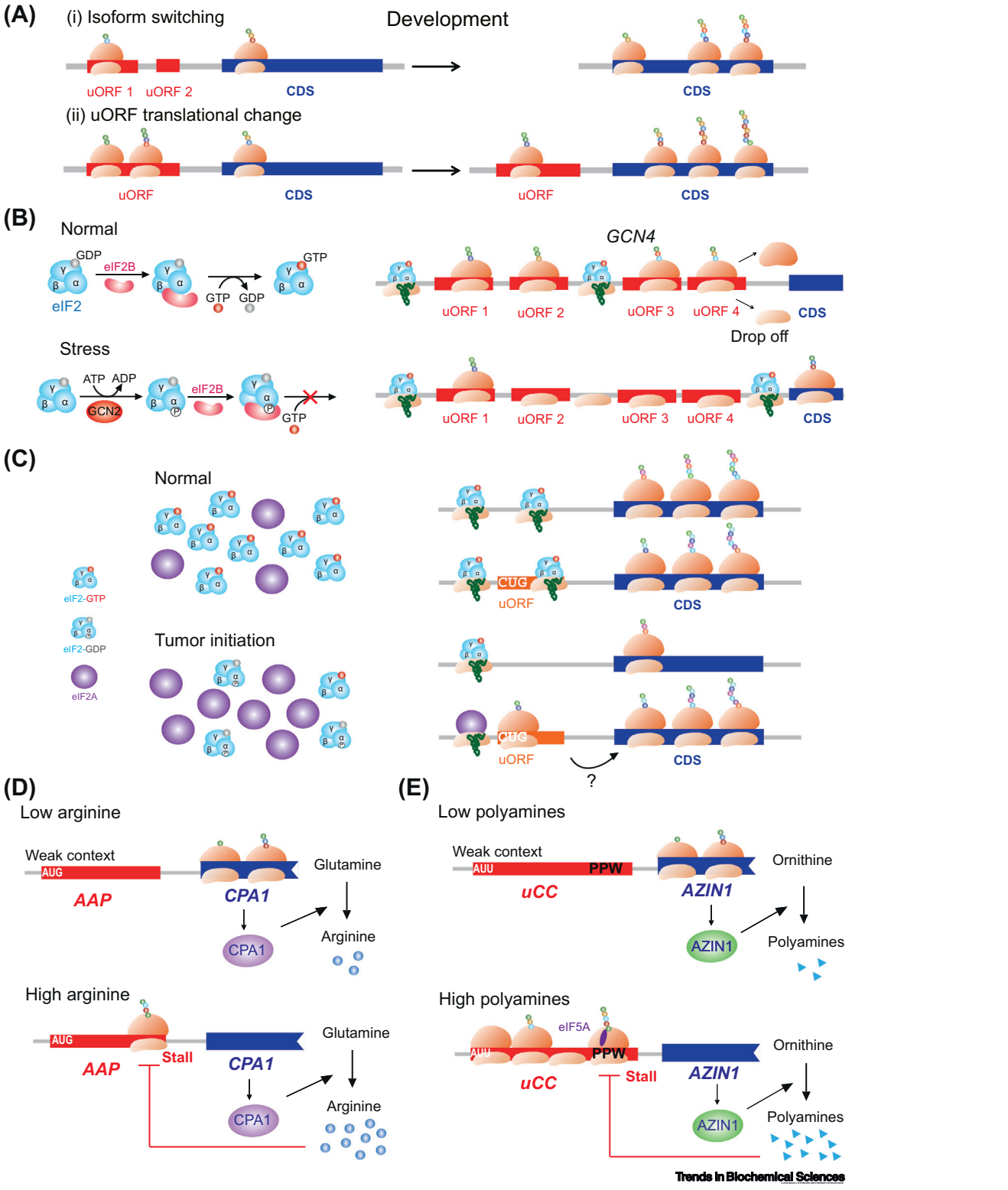


Figure I. Mutations Generating or Disrupting Upstream Open Reading Frames (uORFs) That Cause Human Disease. (A) A G>T point mutation at position –34 creates an overlapping uORF (oORF) in the 5′ UTR of *CDKN2A* mRNA, which decreases the translation of the downstream coding sequence (CDS) [92]. (B) A G>C point mutation at a splicing donor site of *THPO* mRNA leads to skipping of the preceding exon (ΔE) during mRNA splicing. This mutation abolishes uORF2 in wild-type mRNA and turns the nonoverlapping uORF1 wild-type mRNA into an in-frame overlapping uORF. As a result, the *THPO* CDS in the ΔE transcript is efficiently translated with an N-terminal extension (the red segment in the protein sequence), which is finally removed during post-translational maturation. The increased level of TPO leads to thrombocythemia [93]. The genomic sequences around exon boundaries are displayed to illustrate the influence of exon-skipping. The CDS start sites are indicated with '+1'. Codons in uORFs or CDSs are denoted with underscores. Abbreviations: TSS, transcription start site; UTR, untranslated region.

is modulated by switching between mRNA transcript isoforms that differ in the content of uORFs. Furthermore, genome-wide maps of ribosomal occupancy in uORFs and CDSs during the major developmental stages of *D. melanogaster* were recently generated [14]. These results showed



(See figure legend at the bottom of the next page.)

that, during *D. melanogaster* development, changes in the translational efficiencies of uORFs, as well as the inclusion and/or exclusion of uORFs, are frequently exploited to inversely influence the translation of downstream CDSs (Figure 4A) [14]. In *Arabidopsis thaliana*, transcription from alternative transcription start sites (TSSs) is induced to evade uORF-mediated repression after exposure to blue light [53]. In contrast to other model organisms, *Caenorhabditis elegans* uses *trans*-splicing to remove uAUGs in native 5′ UTRs and improve the translational efficiency of a subset of mRNAs [54]. Together, these studies suggest that uORF-mediated regulation has essential roles in controlling developmental programs.

Functional Roles of uORFs in Stress Responses

Interestingly, uORFs are significantly enriched in stress-response genes and control the translation of certain master regulators of stress responses [55,56]. Eukaryotic initiation factor 2 (eIF2), which is a heterotrimer comprising α, β, and γ subunits, forms a TC with GTP and methionyl initiator transfer RNA (Met-tRNA$_i$) as the first step in translation initiation [6]. This TC then binds to the 40S ribosome subunit, forming the PIC. Upon successful AUG recognition through PIC scanning, the GTPase-activating protein eIF5 stimulates the hydrolysis of eIF2-bound GTP, which leads to eIF2-GDP dissociation and the 60S subunit joining (Box 1) [6]. To reform TC, eIF2-GDP must be converted back to eIF2-GTP with the aid of the guanine nucleotide exchange factor eIF2B. Under cellular stress, kinases such as GCN2 or PERK phosphorylate eIF2α (eIF2α~P) to inhibit the activity of eIF2B, leading to reduced TC abundance (Figure 4B) [57]. Hence, it will take longer for a post-termination scanning 40S ribosome to reacquire a new TC and become competent for re-initiation in stressed cells. This delay causes leaky scanning of downstream uORFs that can typically be translated by re-initiating ribosomes. For example, in yeast, GCN4 controls the transcription of amino acid biosynthesis genes in response to amino acid starvation [58]. Translation of the four uORFs of the *GCN4* mRNA represses the translation of the CDS under normal conditions. Under starvation, the TC level is reduced due to the phosphorylation of eIF2 by GCN2. Hence, although uORF1 is translated, the reinitiating ribosomes bypass uORFs 2–4 and re-initiate translation at the *GCN4* CDS to increase the GCN4 protein level in starved cells (Figure 4B) [10,58].

In mammals, ATF4 regulates the transcription of many genes in the stress adaptation of mammalian cells. eIF2α~P phosphatase GADD34 and CReP are also highly expressed during stress to control protein homeostasis [59]. Similar to *GCN4*, their ORFs are translated by regulated leaky scanning of the second uORFs in 5′ UTR during stress [60,61]. In *Arabidopsis*, the translation of the *TBF1* CDS is inhibited by two uORFs under normal conditions but activated upon immune induction [62]. Taken together, these results provide valuable insights into the regulatory roles of uORFs in stress responses and provide guidelines for genetic engineering [63].

Alternative translation initiation at noncanonical uORFs has been demonstrated to be an important mechanism of translational regulation during stress responses [64,65] or tumor initiation

Figure 4. Upstream Open Reading Frames (uORFs) Dynamically Modulate Coding Sequence (CDS) Translation. (A) Modulation of CDS translation via the inclusion or exclusion of uORFs through isoform switching (i) or changes in the translational efficiency of uORFs (ii) during development. (B) uORFs mediate the stress-specific translation of the *GCN4* CDS in yeast. Under normal conditions, eukaryotic initiation factor 2 (eIF2)-GDP is quickly converted to eIF2-GTP by eIF2B to reform the ternary complex (TC). Re-initiating 40S ribosomes after uORF1 and uORF2 could receive new TCs to translate uORF3 and uORF4 and subsequently dissociate from the mRNA, leaving the CDS untranslated. Under stress conditions, GCN2 phosphorylates eIF2α and traps eIF2-GDP in an 'unproductive' state with eIF2B, leading to reduced TC [57]. Therefore, re-initiating ribosomes will spend a longer time scanning along mRNA and bypass uORF3 and uORF4 before reacquiring TCs, leading to CDS translation. (C) During tumor initiation, reduced abundance of TCs causes global translation suppression of most mRNAs. However, eIF2A abundance increases and promotes the translation of CUG/GUG/AUG uORFs, leading to upregulated translation of downstream CDSs through mechanisms that are not yet clear. (D) The translation of *CPA1* is not inhibited by *AAP* when the arginine level is low. A high concentration of arginine causes ribosome stalling at the stop codon of *AAP*, which blocks other ribosomes and induces nonsense-mediated decay (NMD) to reduce the production of CPA1. (E) When the level of polyamines is low, the upstream conserved coding region (uCC) has little effect on the translation of *AZIN1*. A high concentration of polyamines induces ribosome stalling at the PPW motif by interfering with the function of eIF5A, which also causes ribosome queuing along uCC [73].

[66]. eIF2A is an alternative $tRNA_i$ carrier that can bind and deliver Met-tRNA or Leu-tRNA for initiation at the AUG/CUG/UUG codon [64,67,68]. When cAUG-initiated translation is compromised during endoplasmic reticulum (ER) stress or tumor formation, the eIF2A-dependent translation of many uORFs (mostly noncanonical uORFs) in specific mRNAs promotes the translation of downstream CDSs (Figure 4C) [64–67]. For example, *BiP* encodes a chaperone protein that is important for ER homeostasis and has two non-AUG uORFs (–190 UUG and –61 CUG). Under normal conditions, canonical eIF2α-mediated translation initiation is responsible for the translation of the *BiP* CDS. During ER stress, upregulated eIF2A initiates the translation of non-AUG uORFs, which further promotes the translation of the downstream *BiP* CDS [65]. Exactly how the eIF2A-mediated translation of non-AUG uORFs promotes the translation of downstream CDSs remains to be further explored. It is possible that, after translational termination at uORFs, re-initiating ribosomes contribute to the translation of CDSs in these mRNAs. Overall, these results suggest that the eIF2A-mediated translation initiation of uORFs is a common mechanism used by mammalian cells under abnormal conditions.

Notably, the nascent peptide chains of uORFs can interact with the **ribosomal tunnel** (Box 1) to induce ribosome stalling, which further modulates the translation of downstream CDSs. For example, the fungal arginine attenuator peptide (AAP) is encoded by a uORF in the 5′ UTR of the *CPA1* mRNA (Figure 4D). The start codon of *AAP* is located in a disfavored context and seldom translated [69]. When the level of arginine is low, the translation of *CPA1* is not inhibited by *AAP*. Nevertheless, when the arginine level is high, the translation of *AAP* causes stalling of ribosomes at the stop codon of *AAP* since arginine might induce the changes in conformation or the environment of AAP nascent peptides in the ribosome tunnel [70], which inhibits translation re-initiation at the *CPA1* CDS and triggers NMD of *CPA1* mRNA [69,71]. Since CPA1 promotes arginine biogenesis from glutamine, the production of arginine is controlled by a negative feedback loop. Similar uORF-mediated translational regulation is also observed for the ortholog of *CPA1* in *Neurospora crassa*, *arg-2*, suggesting that this mechanism is evolutionarily conserved [72].

The translation of antizyme inhibitor 1 (AZIN1), a protein required for polyamine synthesis from ornithine in animals, is also regulated by a uORF called the upstream conserved coding region (uCC) [73,74]. The start codon of uCC is AUU, which is located in a disfavored Kozak context (Figure 4E). When the cellular polyamine level is low, uCC has a negligible effect on the translation of *AZIN1* CDS. However, when the concentration of polyamines is high, ribosomes would be stalled at the PPW tripeptide motif near the stop codon of uCC because polyamines compete for binding sites with the translation factor eIF5A, which prevents the translation of *AZIN1* CDS [73,75]. Moreover, ribosome stalling can also cause ribosome queuing, which increases the initiation rate at uCC by positioning the scanning ribosomes near the start codon of uCC, further augmenting the inhibition of CDS translation [73] (Figure 4E). Hence, the feedback control mediated by the interactions between uORF translation and metabolites might be important for maintaining cellular metabolic homeostasis.

Crosstalk between uORFs and Other *cis*-Regulatory Elements

In addition to uORFs, many ***cis*-regulatory elements** (CREs) of mRNAs also influence mRNA translation [3], such as the target sites of **miRNAs**, the binding sites of RNA-binding proteins [76], secondary structures, mRNA modifications, and the polyadenylated [poly(A)] tails of mRNAs [77]. For example, since the translation is predominantly controlled by poly(A)-tail length in early embryos of *Drosophila* [78], the repressive effects of uORFs on translation regulation might be overwhelmed by the activating effects of poly(A)-tails in 0–2 h embryos [14]. It has also been demonstrated that translational re-initiation and dynamic changes in N^6-methyladenosine (m^6A) modification in uORF2 of *ATF4* cooperate to promote *ATF4* translation under stress [79]. In addition, two recent studies showed that downstream secondary structures in 5′ UTR could retard the

movement of scanning PICs, which promotes the translation initiation of uORFs with near-cognate start codons or located in suboptimal context [80,81]. We are now beginning to understand the possible crosstalk between different regulatory mechanisms. However, further studies are required to decipher the possible interactions (synergistic, opposing, or additive) between uORFs and other categories of CREs in different cellular contexts.

How Many uORFs Encode Functional Peptides?

Although most eukaryotic uORFs show evidence of translation as revealed by ribosome profiling, how many of these uORFs encode functional peptides remains enigmatic. It is reported that codon usage in uORFs is more similar to the random triplet frequency in 5′ UTRs than that in CDSs [14,82]. Moreover, analyses of codon substitutions in uORFs suggest that they are unlikely to encode conserved peptides [14,16], and only a relatively small number of uORFs might encode peptides that are evolutionarily conserved [83–85]. Collectively, these observations support the hypothesis that most uORFs are designed not to encode bioactive peptides [42], but to impede the translation initiation of downstream CDSs by sequestering ribosomes.

Concluding Remarks and Future Perspectives

Our understanding of the regulatory functions of uORFs has been greatly expanded during the past years. Most uORFs in eukaryotic genomes have been found to be translated, and the translation of uORFs has been shown to modulate the translation of downstream CDSs by sequestering ribosomes. Combining functional genomics and evolutionary analyses enables us to not only identify the genomic features that affect the repressive efficiency of uORFs, but also to dissect the evolutionary forces governing the genome-wide distribution of uORFs in eukaryotic genomes. Nevertheless, many fundamental questions regarding uORF-mediated regulation remain (see Outstanding Questions), and it is exciting that we are beginning to decipher the regulatory mechanisms and consequences of uORFs at a genomic scale.

Outstanding Questions

What are the evolutionary principles underlying the compositions of uORFs in eukaryotic genomes at the micro- and macroevolutionary scales?

How do uORFs dynamically modulate the translation of downstream CDSs during various biological processes, such as stress responses and tumorigenesis?

How do uORFs interact with other *cis*- and *trans*-regulatory elements in modulating mRNA translation?

To what extent are non-AUG uORFs translated and what are their biological functions?

What is the relationship between uORF translation and nonsense-mediated decay?

What is the mechanism by which uORFs enable leaky scanning or re-initiation?

Acknowledgments

We thank Shu-Bing Qian and Wenfeng Qian for critical reading of this manuscript and constructive suggestions. This work was supported by grants from the National Natural Science Foundation of China (No. 91731301 and 91431101), the Ministry of Science and Technology of the People's Republic of China (2016YFA0500800) and the Peking-Tsinghua Center for Life Sciences to J.L.

References

1. Jackson, R.J. *et al.* (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113–127
2. Sonenberg, N. and Hinnebusch, A.G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731–745
3. Hinnebusch, A.G. *et al.* (2016) Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science* 352, 1413–1416
4. Shirokikh, N.E. and Preiss, T. (2018) Translation initiation by cap-dependent ribosome recruitment: recent insights and open questions. *Wiley Interdiscip. Rev. RNA* 9, e1473
5. Hinnebusch, A.G. (2017) Structural insights into the mechanism of scanning and start codon recognition in eukaryotic translation initiation. *Trends Biochem. Sci.* 42, 589–611
6. Hinnebusch, A.G. (2014) The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* 83, 779–812
7. Ruiz-Orera, J. and Alba, M.M. (2019) Translation of small open reading frames: roles in regulation and evolutionary innovation. *Trends Genet.* 35, 186–198
8. Morris, D.R. and Geballe, A.P. (2000) Upstream open reading frames as regulators of mRNA translation. *Mol. Cell. Biol.* 20, 8635–8642
9. Ingolia, N.T. *et al.* (2012) The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550
10. Ingolia, N.T. *et al.* (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223
11. Ingolia, N.T. *et al.* (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802
12. Lei, L. *et al.* (2015) Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *Plant J.* 84, 1206–1218
13. Hsu, P.Y. *et al.* (2016) Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 113, E7126–E7135
14. Zhang, H. *et al.* (2018) Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during *Drosophila* development. *PLoS Biol.* 16, e2003903
15. Dunn, J.G. *et al.* (2013) Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife* 2, e01179
16. Johnstone, T.G. *et al.* (2016) Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.* 35, 706–723
17. Chew, G.L. *et al.* (2016) Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat. Commun.* 7, 11663

18. Stumpf, C.R. *et al.* (2013) The translational landscape of the mammalian cell cycle. *Mol. Cell* 52, 574–582
19. Kozak, M. (1981) Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* 9, 5233–5252
20. Andreev, D.E. *et al.* (2018) TASEP modelling provides a parsimonious explanation for the ability of a single uORF to derepress translation during the integrated stress response. *eLife* e32563, 7
21. Uchiyama-Kadokura, N. *et al.* (2014) Polyamine-responsive ribosomal arrest at the stop codon of an upstream open reading frame of the AdoMetDC1 gene triggers nonsense-mediated mRNA decay in *Arabidopsis thaliana*. *Plant Cell Physiol.* 55, 1556–1567
22. Torrance, V. and Lydall, D. (2018) Overlapping open reading frames strongly reduce human and yeast STN1 gene expression and affect telomere function. *PLoS Genet.* 14, e1007523
23. Gunisova, S. *et al.* (2018) Please do not recycle! Translation reinitiation in microbes and higher eukaryotes. *FEMS Microbiol. Rev.* 42, 165–192
24. Hrdlickova, R. *et al.* (2017) RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA* 8, e1364
25. Wang, Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63
26. Lee, S. *et al.* (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2424–E2432
27. Gao, X. *et al.* (2015) Quantitative profiling of initiating ribosomes *in vivo*. *Nat. Methods* 12, 147–153
28. Garreau de Loubresse, N. *et al.* (2014) Structural basis for the inhibition of the eukaryotic ribosome. *Nature* 513, 517–522
29. Erhard, F. *et al.* (2018) Improved Ribo-seq enables identification of cryptic translation events. *Nat. Methods* 15, 363–366
30. Spealman, P. *et al.* (2018) Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. *Genome Res.* 28, 214–222
31. Na, C.H. *et al.* (2018) Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. *Genome Res.* 28, 25–36
32. Arribere, J.A. and Gilbert, W.V. (2013) Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res.* 23, 977–987
33. Wethmar, K. *et al.* (2016) Comprehensive translational control of tyrosine kinase expression by upstream open reading frames. *Oncogene* 35, 1736–1742
34. Zhang, H. *et al.* (2018) Genome editing of upstream open reading frames enables translational control in plants. *Nat. Biotechnol.* 36, 894–898
35. Rogozin, I.B. *et al.* (2001) Presence of ATG triplets in 5′ untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics* 17, 890–900
36. Calviello, L. *et al.* (2016) Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods* 13, 165–170
37. Schleich, S. *et al.* (2014) DENR-MCT-1 promotes translation reinitiation downstream of uORFs to control tissue growth. *Nature* 512, 208–212
38. Fujii, K. *et al.* (2017) Pervasive translational regulation of the cell signalling circuitry underlies mammalian development. *Nat. Commun.* 8, 14443
39. 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
40. Li, G.-W. *et al.* (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635
41. Lynch, M. *et al.* (2005) The evolution of transcription-initiation sites. *Mol. Biol. Evol.* 22, 1137–1146
42. Neafsey, D.E. and Galagan, J.E. (2007) Dual modes of natural selection on upstream open reading frames. *Mol. Biol. Evol.* 24, 1744–1751
43. Churbanov, A. *et al.* (2005) Evolutionary conservation suggests a regulatory function of AUG triplets in 5′-UTRs of eukaryotic genes. *Nucleic Acids Res.* 33, 5512–5520
44. Charlesworth, B. (2009) Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10, 195–205
45. Shen, S. *et al.* (2011) Widespread establishment and regulatory impact of Alu exons in human genes. *Proc. Natl. Acad. Sci. U. S. A.* 108, 2837–2842
46. Kitano, S. *et al.* (2018) Transposable elements shape the human proteome landscape via formation of cis-acting upstream open reading frames. *Genes Cells* 23, 274–284
47. Chuong, E.B. *et al.* (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86
48. Rebollo, R. *et al.* (2012) Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* 46, 21–42
49. Baralle, F.E. and Giudice, J. (2017) Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437
50. Resch, A.M. *et al.* (2009) Evolution of alternative and constitutive regions of mammalian 5′UTRs. *BMC Genomics* 10, 162
51. Chen, J. *et al.* (2017) Kinetochore inactivation by expression of a repressive mRNA. *eLife* 6, e27417
52. Cheng, Z. *et al.* (2018) Pervasive, coordinated protein-level changes driven by transcript isoform switching during meiosis. *Cell* 172, 910–923
53. Kurihara, Y. *et al.* (2018) Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 115, 7831–7836
54. Yang, Y.F. *et al.* (2017) Trans-splicing enhances translational efficiency in C. elegans. *Genome Res.* 27, 1525–1535
55. Sidrauski, C. *et al.* (2015) The small molecule ISRIB reverses the effects of eIF2α phosphorylation on translation and stress granule assembly. *eLife* 4, e05033
56. Andreev, D.E. *et al.* (2015) Translation of 5′ leaders is pervasive in genes resistant to eIF2 repression. *eLife* 4, e03971
57. Kashiwagi, K. *et al.* (2016) Crystal structure of eukaryotic translation initiation factor 2B. *Nature* 531, 122–125
58. Hinnebusch, A.G. (2005) Translational regulation of GCN4 and the general amino acid control of yeast. *Annu. Rev. Microbiol.* 59, 407–450
59. Carrara, M. *et al.* (2017) Decoding the selectivity of eIF2alpha holophosphatases and PPP1R15A inhibitors. *Nat. Struct. Mol. Biol.* 24, 708–716
60. Young, S.K. *et al.* (2015) Ribosome reinitiation directs gene-specific translation and regulates the integrated stress response. *J. Biol. Chem.* 290, 28257–28271
61. Vattem, K.M. and Wek, R.C. (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11269–11274
62. Xu, G. *et al.* (2017) Global translational reprogramming is a fundamental layer of immune regulation in plants. *Nature* 545, 487–490
63. Xu, G. *et al.* (2017) uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature* 545, 491–494
64. Kim, J.H. *et al.* (2011) eIF2A mediates translation of hepatitis C viral mRNA under stress conditions. *EMBO J.* 30, 2454–2464
65. Starck, S.R. *et al.* (2016) Translation from the 5′ untranslated region shapes the integrated stress response. *Science* 351, aad3867
66. Sendoel, A. *et al.* (2017) Translation from unconventional 5′ start sites drives tumour initiation. *Nature* 541, 494–499
67. Starck, S.R. *et al.* (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* 336, 1719–1723
68. Kearse, M.G. and Wilusz, J.E. (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.* 31, 1717–1731
69. Werner, M. *et al.* (1987) The leader peptide of yeast gene CPA1 is essential for the translational repression of its expression. *Cell* 49, 805–813
70. Wu, C. *et al.* (2012) Arginine changes the conformation of the arginine attenuator peptide relative to the ribosome tunnel. *J. Mol. Biol.* 416, 518–533
71. Gaba, A. *et al.* (2005) Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Mol. Cell* 20, 449–460

72. Wang, Z. and Sachs, M.S. (1997) Ribosome stalling is responsible for arginine-specific translational attenuation in *Neurospora crassa*. *Mol. Cell. Biol.* 17, 4904–4913
73. 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
74. Ivanov, I.P. *et al.* (2008) uORFs with unusual translational start codons autoregulate expression of eukaryotic ornithine decarboxylase homologs. *Proc. Natl. Acad. Sci. U. S. A.* 105, 10079–10084
75. Schuller, A.P. *et al.* (2017) eIF5A functions globally in translation elongation and termination. *Mol. Cell* 66, 194–205
76. Gerstberger, S. *et al.* (2014) A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845
77. Subtelny, A.O. *et al.* (2014) Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508, 66–71
78. Eichhorn, S.W. *et al.* (2016) mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos. *eLife* 5, e16955
79. Zhou, J. *et al.* (2018) N(6)-methyladenosine guides mRNA alternative translation during integrated stress response. *Mol. Cell* 69, 636–647
80. Guenther, U.P. *et al.* (2018) The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs. *Nature* 559, 130–134
81. Murat, P. *et al.* (2018) RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.* 19, 229
82. Liu, M.J. *et al.* (2013) Translational landscape of photomorphogenic *Arabidopsis*. *Plant Cell* 25, 3699–3710
83. Mackowiak, S.D. *et al.* (2015) Extensive identification and analysis of conserved small ORFs in animals. *Genome Biol.* 16, 179
84. Aspden, J.L. *et al.* (2014) Extensive translation of small open reading frames revealed by poly-ribo-seq. *eLife* 3, e03528
85. van der Horst, S. *et al.* (2019) Novel pipeline identifies new upstream ORFs and non-AUG initiating main ORFs with conserved amino acid sequences in the 5′ leader of mRNAs in *Arabidopsis thaliana*. *RNA* 25, 292–304
86. Pollard, K.S. *et al.* (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20, 110–121
87. Michel, A.M. *et al.* (2014) GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res.* 42, D859–D864
88. von Bohlen, A.E. *et al.* (2017) A mutation creating an upstream initiation codon in the SOX9 5′ UTR causes acampomelic campomelic dysplasia. *Mol. Genet. Genomic Med.* 5, 261–268
89. Schulz, J. *et al.* (2018) Loss-of-function uORF mutations in human malignancies. *Sci. Rep.* 8, 2395
90. Barbosa, C. *et al.* (2013) Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet.* 9, e1003529
91. Cenik, C. *et al.* (2015) Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res.* 25, 1610–1621
92. Liu, L. *et al.* (1999) Mutation of the CDKN2A 5′ UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat. Genet.* 21, 128–132
93. Wiestner, A. *et al.* (1998) An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. *Nat. Genet.* 18, 49–52
94. McGillivray, P. *et al.* (2018) A comprehensive catalog of predicted functional upstream open reading frames in humans. *Nucleic Acids Res.* 46, 3326–3338