SURVEY AND SUMMARY

Death of a dogma: eukaryotic mRNAs can code for more than one protein

Hélène Mouilleron[1,2,†], Vivian Delcourt[1,2,3,†] and Xavier Roucou[1,2,*]

[1]Department of biochemistry, Université de Sherbrooke, Sherbrooke, Quebec J1E 4K8, Canada, [2]PROTEO, Quebec Network for Research on Protein Function, Structure, and Engineering, Quebec, Canada and [3]Inserm U-1192, Laboratoire de Protéomique, Réponse Inflammatoire, Spectrométrie de Masse (PRISM), Université de Lille 1, Cité Scientifique, 59655 Villeneuve D'Ascq, France

ABSTRACT

mRNAs carry the genetic information that is translated by ribosomes. The traditional view of a mature eukaryotic mRNA is a molecule with three main regions, the 5′ UTR, the protein coding open reading frame (ORF) or coding sequence (CDS), and the 3′ UTR. This concept assumes that ribosomes translate one ORF only, generally the longest one, and produce one protein. As a result, in the early days of genomics and bioinformatics, one CDS was associated with each protein-coding gene. This fundamental concept of a single CDS is being challenged by increasing experimental evidence indicating that annotated proteins are not the only proteins translated from mRNAs. In particular, mass spectrometry (MS)-based proteomics and ribosome profiling have detected productive translation of alternative open reading frames. In several cases, the alternative and annotated proteins interact. Thus, the expression of two or more proteins translated from the same mRNA may offer a mechanism to ensure the co-expression of proteins which have functional interactions. Translational mechanisms already described in eukaryotic cells indicate that the cellular machinery is able to translate different CDSs from a single viral or cellular mRNA. In addition to summarizing data showing that the protein coding potential of eukaryotic mRNAs has been underestimated, this review aims to challenge the single translated CDS dogma.

BACKGROUND—THE SINGLE FUNCTIONAL ORF PERSPECTIVE OF EUKARYOTIC mRNAS

The general vision of a typical eukaryotic mature mRNA is a monocistronic molecule with a tripartite structure: a single translated ORF or CDS is flanked by 5′ and 3′ UTRs (Figure 1A). Although bicistronic mRNAs were detected in plants and some aspects of the translational mechanisms for these mRNAs have been elucidated (1), this review will mainly focus on animal mRNAs. The single CDS concept mainly stems from the canonical cap dependent scanning mechanism model for the selection of translation initiation sites in eukaryotic mRNAs. This model has provided a fundamental framework for the study of the regulation of translation initiation in eukaryotes and is supported by a substantial body of data (2–4). Briefly, a 43S preinitiation complex binds to the cap structure at the 5′ end and scans the 5′ UTR to arrest at a translation initiation site (TIS). The large 60S subunit then joins to form a fully functional 80S ribosome and polypeptide synthesis starts. In contrast to bacterial ribosomes which can bind to internal binding sites in polycistronic mRNAs, the cap-dependent scanning mechanism as currently visioned is not compatible with the translation of several CDSs in the same mRNA molecule. Consequently, it was concluded well before the pre-genomic era that each mature eukaryotic mRNA is monocistronic and is translated into a single polypeptide (5).

This dominant view of a single functional ORF intensified with the implementation of computational pipelines for automated annotation of genomes. Contemporary approaches to identify protein-coding genes in large eukaryotic genomes and transcriptomes generally use combinations of three methods: (i) statistical information, including codon usage; (ii) splice sites and sequence similarity to previously identified proteins and genes; and (iii) experimen-
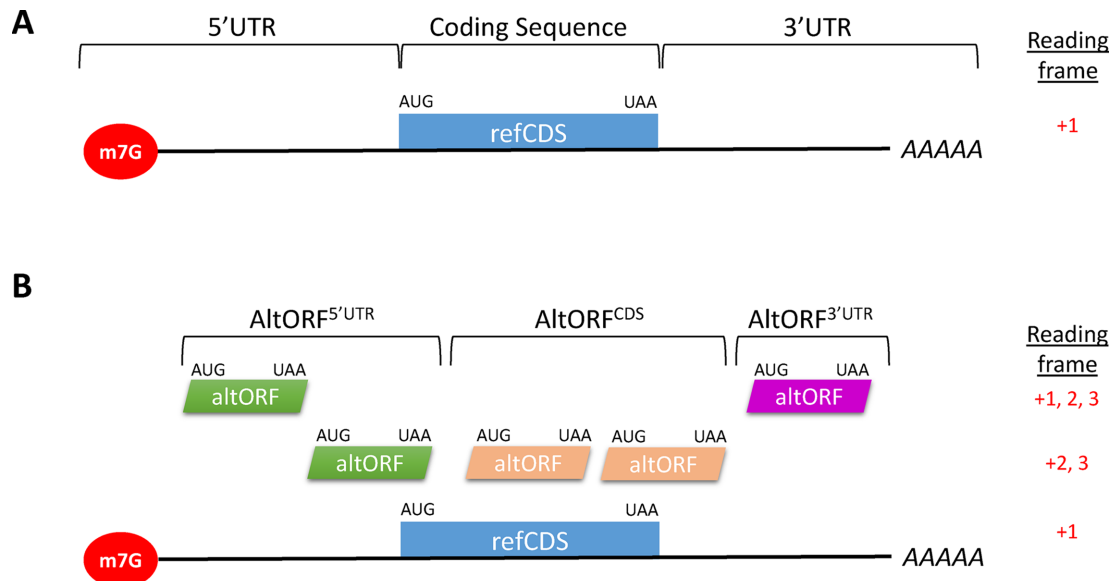
Figure 1. The typical tripartite structure of a eukaryotic mRNA with a single annotated or reference coding sequence or refCDS (**A**, single CDS dogma), or with possible alternative ORFs with an initiation codon located in the 5′ UTR, the CDS, or the 3′ UTR. (**B**). AltORFs$^{5'UTR}$ may overlap refCDSs in a different reading frame. AltORFs^{CDS} may extend into 3′ UTRs. In general, refCDSs are longer than altORFs. Generally, all annotated mRNAs have a refCDS. An mRNA may have no, one or several altORFs. Abbreviations: AAAAA, polyA tail; altORF, alternative open reading frame; AUG, translation initiation codon; refCDS, annotated or reference protein coding sequence; m7G, 7-methyl-guanosine cap; UAA, stop codon (only one possible stop codon is shown for simplicity).

tal evidence of transcript-derived sequences of cDNAs or expressed sequence tags (6–12). Fortunately, vast resources of cDNAs, ESTs, and protein sequences have been assembled in public databases and are invaluable for the analysis of large genomes (13). The refinement of the predictions of protein-coding genes using cDNAs and ESTs has clearly improved the accuracy of the annotation of CDSs within large genomes (14). Reviews and comparisons of several methods for the identification of protein-coding regions identification methods are available (15–21). In the absence of experimental data on gene expression or gene homology, several programs can be used to make ab initio gene predictions (22–25).

Methods to specifically predict CDSs in the transcriptome have also been developed and used for the characterization of large ESTs, cDNAs and RNA-seq collections (26–32). Similar to approaches used with genomic DNA sequences, CDSs annotation in high-throughput transcriptomic data include statistical and similarities metrics.

Typically, all computational calculations described above identify a single functional ORF or CDS per locus or transcript with a statistically significant signature of a protein-coding region. In this review, annotated CDSs are termed reference CDSs (refCDSs). When no substantial similarity is generated, the longest ORF is predicted to be the single most probable CDS and becomes the refCDS. Generally, a cutoff of 100 codons is applied for a significant CDS despite increasing evidence that peptides translated from short non-annotated CDSs have important functions (33–35). These rules, which exclude 'small' CDSs are part of the annotation guidelines used by virtually all publicly available gene sets, including AceView (36), GENCODE (37), RefSeq (38), En-

sembl (39), VEGA (40), CCDS (41) and have exacerbated the single CDS dogma (Figure 1).

Establishing a comprehensive catalog of protein-coding genes and CDSs in the genome of eukaryotes remains a fundamental objective in modern biology and medicine, and the objective of identifying one complete refCDS for each gene is already an ambitious project (15). By reducing the search space in genome-wide analyses, the single refCDS principle greatly simplifies the annotation of CDSs and protein-coding genes but ignores additional CDSs and substantially underestimates the coding potential of genomes.

A SECOND LOOK AT ORFs WITHIN EUKARYOTIC mRNAS

Alternative ORFs versus reference CDSs

There are three reading frames for a transcript, and a typical mature mRNA may contain several alternative ORFs (altORFs) in addition to the refCDS (Figure 1). In this article, altORFs are defined as potential protein-coding sequences that are completely different from refCDSs in the same transcript. Proteins translated from altORFs have a different amino acid sequence and are not isoforms of the annotated proteins. AltORFs may be divided into three classes according to the location of the alternative versus the annotated TIS (Figure 1). Here, refCDSs are present in the +1 reading frame. AltORFs$^{5'UTR}$ include altORFs with a TIS in the 5′ UTRs in any of the 3 reading frames. AltORFs that extend into refCDSs in the +2 or +3 reading frames are also considered to be AltORFs$^{5'UTR}$. ORFs initiating upstream of the canonical initiation codon in the +1 reading frame code for long isoforms of the annotated proteins and will not be considered as altORFs in this article. AltORFs$^{5'UTR}$

are also termed upstream ORFs. They repress the translation of refCDSs and are believed to be mainly translational regulatory elements (42–44). AltORFs^{CDS} include altORFs with a TIS inside the refCDSs in the +2 or +3 reading frame. AltORFs^{CDS} are also termed overlapping ORFs (45). AltORFs$^{3'UTR}$ are localized inside 3′ UTRs.

As mentioned above, refCDSs are usually the longest CDSs (>100 codons), and altORFs are generally shorter. There are however exceptions when the first discovered protein translated from a specific mRNA does not turn out to be the largest one, such as the RPP14 mRNA. RPP14 refCDS is translated into a 124 amino acid subunit of the essential human ribonuclease P, and an altORF$^{3'UTR}$ is translated into a 168 amino acid 3-hydroxyacyl-thioester dehydratase HsHTD2 which is important in mitochondrial fatty acid synthesis (46). The presence of both ORFs in the human RPP14 mRNA is clear evidence of a functional bicistronic human mRNA.

ORFs within long non-coding RNAs

There is growing evidence that some annotated long noncoding RNAs (lncRNAs) are translated (47–52) and should in principle be classified as mRNAs. The homo sapiens apelin receptor early endogenous ligand lncRNA is a striking example of a previous lncRNA (RefSeq record, NR_038825.1) that recently changed status to mRNA (NM_001297550.1). This transcript codes for a conserved 32 amino acid hormone with critical function for cell movement and cardiovascular development (35,53). Similar to mRNAs, lncRNAs may be an important source of altORFs, but this review focuses on currently annotated protein-coding transcripts or mRNAs.

In silico detection of altORFs

The serendipitous discovery that a few human genes express proteins from two different CDSs within a single mRNA prompted several laboratories to predict candidate altORFs (45,54,55). Three initial bioinformatics genome-wide studies predicted altORFs^{CDS} in mammalian transcripts (56–58). These studies used sequential filters with different stringencies, including a cut off size (150 or 500 nucleotides), conservation between species and the presence of a strong Kozak signal around the predicted TIS. Yet, they failed to predict several experimentally validated altORFs and subsequent research used less stringent filters and predicted 17,096 altORFs^{CDS} in human transcripts (59). These predictions were later applied to all classes of altORFs and in other eukaryotes (60). The human transcriptome contains 83,886 potential altORFs with a minimum size of 40 codons while the current human proteome contains about 52,000 annotated proteins (RefSeq release 72). A recently developed database specifically facilitates the detection of conserved altORFs$^{5'UTR}$ completely located within 5′ UTRs (61). Overall, these analyses clearly detected potential protein-coding altORFs and revealed their widespread presence in different transcriptomes. These predictions likely underestimate the number of altORFs since the presence of an AUG initiation codon is an important criteria in the computational detection of altORFs. Yet, experimental evidence and evidence from evolutionary studies convincingly demonstrate that translation initiation does not always start at AUG codons (48,62–64).

EXPERIMENTAL EVIDENCE FOR THE TRANSLATION OF altORFs

Non-large scale approaches

Over the past two decades, several refCDS/altORF doublets have been discovered in mammals. These include INK4a/ARF (55), histone H4/OGP (65), XLalphas/ALEX (45,66), RPP14/HsHTD2 (46), PrP/altPrP (67), ATXN1/altATXN1 (68), $A_{2A}R$/uORF5 (69), MKKS/uMKKS1 and uMKKS2 (70) and $AT_{1a}R$/PEP7 (71). Remarkably and perhaps not coincidentatly, some alternative proteins functionally interact with their respective reference proteins. XLalphas/ALEX co-localize in plasma membrane ruffles and directly interact, and ALEX negatively regulates the activity of the G-protein XLalphas subunit (45,72). ATXN1/altATXN1 co-localize in nuclear foci and directly interact (68). The stimulation of $A_{2A}R$ by adenosine increases the expression of uORF5 through the activation of protein kinase A (69). PEP7 blocks the beta-arrestin dependent signaling pathway of $AT_{1a}R$ (71). We suggest that this may be a frequent or even common occurrence. This would provide a mechanism for ensuring that two or more proteins are always expressed together.

The co-expression of refCDS/altORF^{CDS} doublets in cultured cells transfected with an expression vector containing the refCDS is a significant result (67,68). First, it demonstrates that mechanisms for the translation of overlapping ORFs are constitutive and are an intrinsic feature of mammalian cells. The serendipitous discovery of the translation of HsHTD2 altORF$^{3'UTR}$ in yeast cells transformed with an expression plasmid also indicates that translation mechanisms for altORFs are conserved in eukaryotes (46). Second, transfection of refCDSs in cultured cells for functional investigations of reference proteins is a common technique in research laboratories. The undetected co-expression of refCDS/altORF^{CDS} doublets indicate that the interpretation of some experimental results may have to be questioned (section Implications below).

Some human cancer specific antigens that are silent in normal tissues are also translated from altORFs^{CDS} (73–78). Such tumor-specific antigens are promising targets for the development of immunotherapies but the function of these proteins in cancer has not been characterized yet.

Large scale approaches

Detection of alternative proteins by mass spectrometry. Two main challenges have hampered the detection of alternative proteins by mass spectrometry (MS). First, protein sequence databases are central to the success of MS-based protein identification. UniProt Knowledgebase (79) and the NCBI Reference Sequence collection (38) are among the most widely used databases. Alternative proteins are not annotated in these databases and, as such, cannot be detected and remain invisible to MS users. We addressed this issue by creating a database containing the protein sequence of

all predicted altORFs. A total of 1259 novel alternative proteins were detected in different human cell lines, tissues and fluids (60) (Table 1 and Figure 2A).

Second, altORFs are shorter than refCDSs and small proteins in general are more challenging to detect by MS. In an approach specifically designed for the detection of non-annotated small proteins, a total of more than 200 polypeptides of less than 150 amino acids long were detected in three independent studies performed in human K562 cells and tissues (80–82). The coding sequences were searched against public or in-house built cDNA libraries and matched to all three classes of altORFs.

Detection of altORFs being translated by ribosome profiling. Ribosome profiling is an emerging technique that provides genome-wide information on different aspects of translation in vivo (83–85). Briefly, ribosome-bound mRNAs are isolated and treated with a nuclease. The resulting ribosome-protected RNA fragments or footprints are identified by high throughput sequencing. This novel technology can map TISs and regions within transcripts that are translated, and has already provided major novel insights into the complexity of translated sequences and the mechanisms of protein synthesis and translational control (84,86). Importantly, ribosome profiling detects functional TISs independently of annotated CDSs and is thus completely unbiased towards the detection of annotated CDSs or alternative CDSs.

In mouse embryonic fibroblast cells and in human HEK293 cells, the majority of mRNAs contain more than one TIS and >50% of detected TISs map to altORFs$^{5'UTR}$ and altORFs^{CDS}. A small proportion of TIS are detected in 3′ UTRs (48,62). *Ab initio* predictions of transcriptome-wide TISs using a neural network trained on ribosomal profiling data generated in a human monocytic cell line independently confirmed these results (63). In violation of the single CDS dogma, these initial striking observations on the widespread translation of altORFs attest to the presence of an unanticipated mechanism for protein diversity. Extensive translation of altORFs have been reported in different organisms and under various experimental settings (47,51,87–92) (Table 1 and Figure 2A). Functional annotated and alternative TISs detected by ribosomal profiling are now mapped and easily searchable in several databases and online genome browsers (93–96).

Few altORFs$^{3'UTR}$ were detected in the studies cited above, but this is not surprising given that the ribosome profiling technology used was very inefficient at detecting ribosomes in the 3′ UTRs (97). Recently, ribosome profiling detected 3′ UTR translation in yeast cells, but TISs were not investigated (98).

TRANSLATION MECHANISMS FOR ALTERNATIVE PROTEINS

AltORFs are clearly not receiving sufficient attention in genome annotations, and the translation of altORFs does not comply with the single CDS rule. Yet, a large number of altORFs are translated and cellular translational mechanisms must operate which allow more than one protein to be translated from a single mRNA species.

The scanning mechanism for initiation of translation predicts that altORFs$^{5'UTR}$ TISs, strategically located in 5′ UTRs are detected prior to the annotated TISs, particularly if they have a strong Kozak sequence (2,4,99). For such mRNAs, altORFs$^{5'UTR}$ (also termed upstream ORFs) are translated first. The translation of refCDSs relies on a reinitiation mechanism highly dependent on the length of altORFs$^{5'UTR}$, the distance and structural constraints between altORFs$^{5'UTR}$ and the refCDSs (43,100–105). Two factors essential for translation reinitiation of refCDSs after translation of altORFs$^{5'UTR}$ with a strong Kozak sequence were recently identified in drosophila cells (106). Based on results obtained with the human immunodeficiency virus type 1 tat mRNA, it was predicted that altORFs$^{5'UTR}$ longer than 55 codons would prevent reinitiation (105).

Nevertheless, SLC35A4 contains 11 ORFs upstream of the refCDS, and translation of the 102 codon long 11th upstream ORF was independently detected by several groups by ribosomal profiling and MS (60,107,108). Clearly, other mechanisms may operate for the translation of such altORFs and for altORFs^{CDS} and altORFs$^{3'UTR}$ for which several upstream TISs must be bypassed before ribosomes reach the altORF TIS. One such possible mechanism is leaky scanning. Leaky scanning allows for ribosomes to scan through TISs without initiating translation and results in the translation of downstream ORFs (100). An optimal Kozak context strongly reduces but does not completely prevent leaky scanning in all mRNAs. A large scale analysis of 22 208 human mRNAs indicates that only 37.4% of mRNAs have an annotated TIS with an optimal Kozak sequence. The majority of human mRNAs are thus predicted to undergo leaky scanning of annotated TISs and produce alternative proteins (109). The osteogenic growth peptide represents a fascinating example of a biological active molecule translated from an altORF^{CDS} by leaky scanning (65,109). The corresponding refCDS encodes the small 103 amino acids and extremely well conserved histone H4, yet it shelters an altORF. Several CDSs in polycistronic viral RNAs also use a leaky scanning mechanism for their translation in mammalian cells (110). Overall, leaky scanning is a well-recognized translational mechanism which allows ribosomes to reach downstream altORFs^{CDS} and altORFs$^{3'UTR}$ TISs in cellular mRNAs.

In addition to leaky scanning, viruses have evolved other strategies to make eukaryotic cells translate several CDSs in polycistronic mRNAs using non-canonical mechanisms (110–113). Internal ribosomal entry sites (IRES) are structured RNA sequences able to recruit the translation machinery under conditions in which cap-dependent translation is compromised. Although the function of IRES in animal and plant viral RNAs is well established, the presence of putative IRES in cellular mRNAs is disputed in the literature and it would be too premature to invoke this mechanism for the translation of altORFs (114–116). Ribosome shunting involves conventional cap-dependent initiation, but the scanning ribosome bypasses a large structured region of the mRNA to reach downstream TISs (110). Two human mRNAs use ribosome shunting for the translation of their refCDSs, HSP70 and cIAP2 (117,118). Ribosome shunting would provide a possible mechanism allowing scanning ribosomes to reach altORFs^{CDS} and
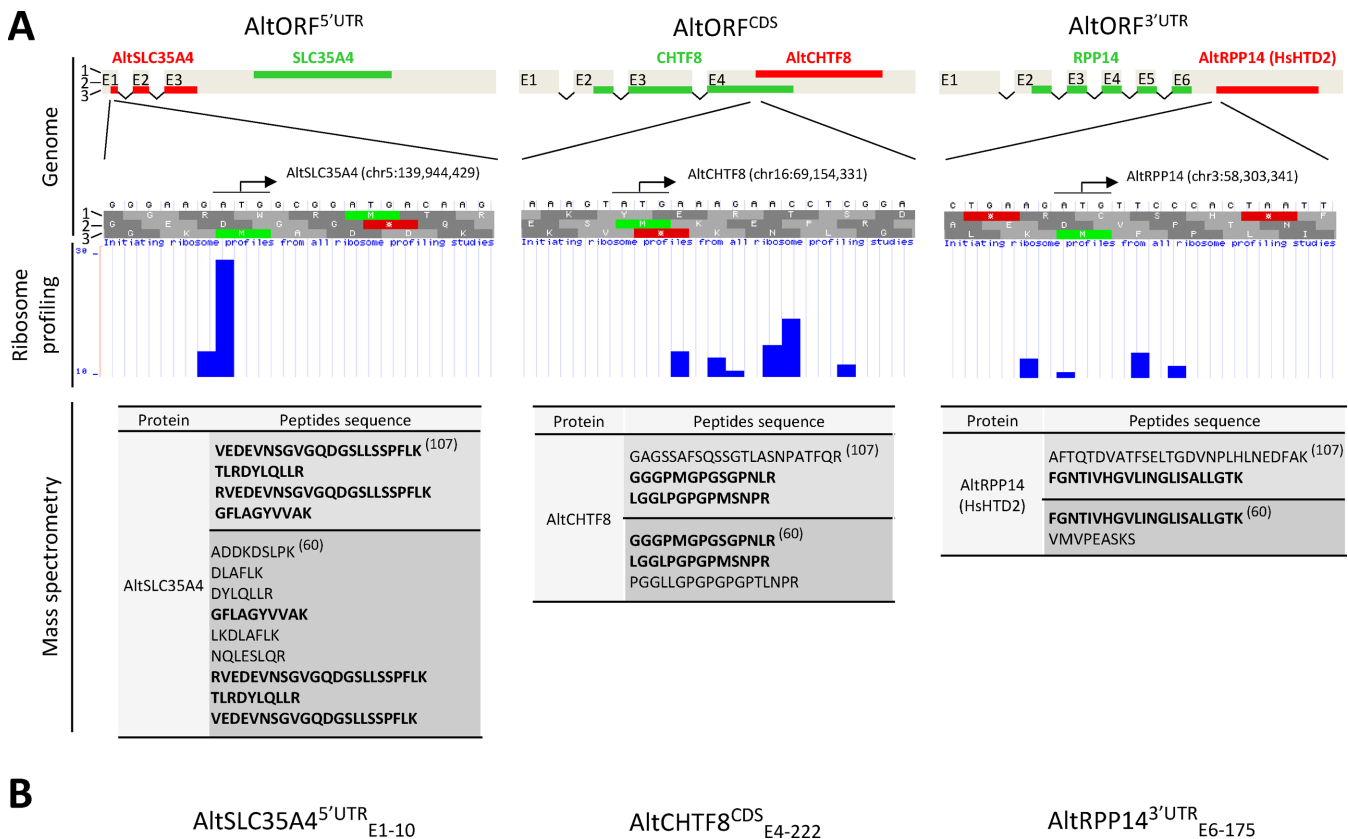
Figure 2. Ribosome profiling and MS are common techniques to detect alternative proteins in large scale studies. (**A**) AltSLC35A4, altCHTF8 and altRPP14(HTD2) are examples of alternative proteins detected by independent laboratories and at least by two different techniques. A schematic view of the genomic structures for the three human genes is shown with the chromosomal coordinates for the three alternative TISs. Exons (gray), refCDSs (green), altORFs (red), and the three reading frames are indicated. The genomic bases around the alternative TISs and the amino acids translated in the three reading frames are also indicated (image from the online genome browser GWIPS-viz (93)). Ribosome profiling data retrieved from GWIPS-viz indicate the position of initiating ribosome profiles from 4 profiling studies (62,63,90,136). Initiating ribosomes are clearly detected around altORFs TISs. The MS data show the peptide sequence of peptides detected in two independent studies, with peptides detected in both studies shown in bold (60,107). (**B**) Proposed nomenclature for altORFs. For example, altSLC35A4$^{5'UTR}_{E1-10}$ indicates an altORF in the SLC35A4 gene; it is located in the 5′ UTR, and the TIS starts at the 10th base of exon 1.

Table 1. Examples of detected alternative proteins in the literature

altORF$^{5'UTR}$	altORF^{CDS}	altORF$^{3'UTR}$
AltSMCR7L (60,62,107,108)	AltCHTF8 (60,62,107)	HsHTD2 (46,60,89,107)
AltSLC35A4 (60,62,107,108)	AltHNRPUL1 (60,107)	AltSF1 (60,107)
uORF5 (69)	AltATXN1 (68)	
PEP7 (71)	AltPrP (67)	
uMKKS1 and 2 (70)		

altORFs$^{3'UTR}$. Ribosome shunting depends on ribosomal protein S25 (119). Stable ribosomal protein S25 knockdown cells are viable and thus it should be possible to verify the function of that protein in the translation of altORFs.

3′ UTRs are a large source of potential altORFs and a majority of alternative proteins detected by MS are encoded in 3′ UTRs (60). A recent ribosome profiling investigation revealed a high abundance of ribosomes in 3′ UTRs of drosophila and human cells but this study did not provide evidence of actual 3′ UTR translation (97). In yeast cells, ribosome profiling and MS demonstrated 3′ UTR translation through an Rli1-dependent mechanism (98).

Does the translation of altORFs necessarily violate the cap-dependent scanning model? This model for the selec-

tion of TISs in eukaryotic mRNAs has provided a fundamental framework for the study of the regulation of translation initiation in eukaryotes, and is supported by a substantial body of data (2–4). A stringent scanning mechanism states that 5′ proximal ORFs are preferentially translated and downstream ORFs are thus unlikely to be translated. More than 60% of human mRNAs do not have an annotated TIS with an optimal Kozak sequence (109), and half of human mRNAs have at least one altORF$^{5'UTR}$ (43). Thus, a strict scanning mechanism could not sustain the translation of refCDSs in a large portion of human mRNAs. For that reason, leaky scanning and reinitiation mechanisms have been incorporated into the scanning model (5,100,120). In addition, reinitiation mecha-

nisms are particularly efficient in mammalian cells. The mechanism of translation for bicistronic GDF1/CERS1, SNRPN/SNURF and RPP14/HsHTD2 was not investigated but is likely to occur by leaky scanning and/or reinitation (46,121,122). Functional polycistronic mRNAs were discovered in metazoans (123–126). For one of them, translation clearly occurs via leaky scanning (123). Overall, there is strong evidence that translation of altORFs in eukaryotic mRNAs does not violate but requires some already demonstrated plasticity of the scanning rule.

In addition to the widely accepted scanning model, mechanisms such as ribosomal tethering, clustering, internal initiation and RNA looping have been proposed for the recruitment of ribosomes to TIS (127–129). For example, two structural elements in the histone H4 mRNA refCDS tether the cap and the 43S preinitiation complex, respectively, and detection of the reference TIS is scanning-independent (128). Importantly, these scanning-independent mechanisms bypass 5′ TISs and allow the recognition of downstream TISs, and possibly the translation of altORFs^{CDS} and altORFs$^{3'UTR}$. These studies were mostly performed with model mRNAs and provided proof of principle that scanning-independent mechanisms can operate. Yet, it will be important to determine if these mechanisms are involved in the translation of cellular mRNAs.

TRANSLATION OF ALTERNATIVE PROTEINS: IMPLICATIONS

The observation that expression of AltORFs may well be a widespread phenomenon has vast implications regarding the way the structure and function of genes are investigated, and the way 'omics' data are interpreted. In mammalian mRNAs, most altORFs with a minimum size of 40 codons are distributed either in 3′ UTRs (46%) or inside refORFs (41%) (60). The presence of a large number of altORFs nested inside refCDSs has the potential to generate results which are difficult to interpret since plasmid driven refCDSs expression in cultured cells or transgenic animals may result in the unnoticed co-expression of altORFs (60,67,68,130). Conversely, knocking down mRNAs could inhibit the expression of both annotated and alternative proteins (60,67,68). The human coagulation factor IX contains an altORF^{CDS} encoding a 64 amino acid protein. Unfortunately, the co-expression of this alternative protein from a transgenic therapeutic cassette elicits an unwanted cytotoxic T lymphocyte response, resulting in the death of cells expressing the transgene (130). The authors suggested that alternative TISs should be inactivated in transgenes used for therapeutic purposes.

Large scale screening assays, including yeast 2-hybrid assays, result in the identification of out-of-frame clones that are systematically rejected. Yet, some may represent positive hits such as the interaction between BRCA1 and altMRVI1 (60).

Other potential implications particularly for overlapping refCDSs/altORFs are more speculative but surely diserve further investigations. Functional polymorphisms affecting the overlapping XLalphas/ALEX refCDS/altORF^{CDS} doublet associate with G signaling hyperfunction in the platelets of patients (72). Deregulated signaling results in part from a decreased interaction between the G-protein XLalphas subunit and ALEX. In addition, synonymous nucleotide substitutions in coding regions or silent mutations are implicated in a number of pathological manifestations (131). A fraction of synonymous mutations in the refCDS may produce a nonsynonymous mutation in an overlapping altORF^{CDS} and alter the corresponding alternative protein.

CONCLUSION AND FUTURE PERSPECTIVES

Eukaryotic mRNAs contain a single refCDS but usually contain several ORFs. The combination of the traditional view of a monocistronic mRNA with computational approaches for genome-wide annotations rejected altORFs. Unfortunately, the single CDS dogma has artificially limited our view of the coding capacity of mRNAs and has prevented the discovery of alternative proteins despite some clues in the literature over the years (132). Recently, a large and rapidly growing body of evidence has provided conclusive experimental evidence for the translation of alternative proteins in addition to the annotated protein from the same mRNA (133).

A new field of investigation is emerging and will have to address several issues. The contribution of alternative proteins to the translational landscape will need to be determined and will likely take advantage of the combination of ribosome profiling and MS-based proteomics (134). In order to demonstrate the functional relevance of alternative proteins, it will be important to specifically inactivate altORFs without affecting refCDSs using gene editing techniques now available (135). New tools such as specific antibodies, updated databases to keep track of the discovery of new transcripts, and new methods to study small proteins will have to be generated. Functional relationships between reference and alternative proteins expressed from the same gene may help identify a new layer of regulation of protein activity (45,68,72).

In parallel to the growing evidence for the translation of non-annotated CDSs, the coding potential of genomes should be updated and the nomenclature for alternative ORFs should be standardized. These unconventional ORFs are termed altORFs (60), overlapping ORFs (56), uORFs (42), sORFs (33), and smORFs (91). The nomenclature for the corresponding translational products is also confusing, and there is a clear need to adopt a standard classification. sORFs and smORFs are shorter than 100 codons and do not represent all non-annotated ORFs. Similarly, uORFs (altORFs$^{5'\,UTR}$) and overlapping ORFs (altORFs^{CDS}) do not include ORFs located in 3′ UTRs. We suggest the term alternative ORFs or altORFs to distinguish all the currently non-annotated ORFs from the annotated CDSs (Figure 1). As indicated above, altORFs may be divided into three classes, altORF$^{5'UTR}$, altORF^{CDS}, and altORF$^{3'UTR}$ according to the localization of the TIS in the tripartite structure of the longest mRNA. Since several altORFs may be present within one mRNA, we propose to add the exon containing the alternative TIS and the position of the first base of the TIS relative to the first base of this exon (Figure 2B). AltORFs in non-coding RNAs may be termed altORFs^{nc}.

REFERENCES

1. Matsuda,D. and Dreher,T.W. (2006) Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA. *RNA*, **12**, 1338–1349.
2. Kozak,M. (1989) The scanning model for translation: an update. *J. Cell Biol.*, **108**, 229–241.
3. Hinnebusch,A.G. (2014) The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.*, **83**, 779–812.
4. Kozak,M. (1978) How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell*, **15**, 1109–1123.
5. Kozak,M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene*, **234**, 187–208.
6. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
7. Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F. *et al.* (2000) The genome sequence of Drosophila melanogaster. *Science*, **287**, 2185–2195.
8. Gibbs,R.A., Weinstock,G.M., Metzker,M.L., Muzny,D.M., Sodergren,E.J., Scherer,S., Scott,G., Steffen,D., Worley,K.C., Burch,P.E. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.
9. Curwen,V., Eyras,E., Andrews,T.D., Clarke,L., Mongin,E., Searle,S.M.J., Clamp,M., Wellcome,T., Genome,T. and Broad,T. (2004) The Ensembl automatic gene annotation system. *Genome Res.*, **14**, 942–950.
10. Haas,B.J., Salzberg,S.L., Zhu,W., Pertea,M., Allen,J.E., Orvis,J., White,O., Buell,C.R. and Wortman,J.R. (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.*, **9**, R7-2008-9-1-r7.
11. Madupu,R., Brinkac,L.M., Harrow,J., Wilming,L.G., Bohme,U., Lamesch,P., Hannick,L.I., Böhme,U., Lamesch,P. and Hannick,L.I. (2010) Meeting report: a workshop on Best Practices in Genome Annotation. *Database (Oxford).*, baq001.
12. Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H. *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60, 770 full-length cDNAs. *Nature*, **420**, 563–573.
13. Birney,E., Clamp,M. and Hubbard,T. (2002) Databases and tools for browsing genomes. *Annu. Rev. Genomics Hum. Genet.*, **3**, 293–310.
14. Imanishi,T., Itoh,T., Suzuki,Y.Y., O'Donovan,C., Fukuchi,S., Koyanagi,K.O., Barrero,R.A., Tamura,T., Yamaguchi-Kabata,Y., Tanino,M. *et al.* (2004) Integrative annotation of 21, 037 human genes validated by full-length cDNA clones. *PLoS Biol.*, **2**, e162.
15. Brent,M.R. (2005) Genome annotation past, present, and future: how to define an ORF at each locus. *Genome Res.*, **15**, 1777–1786.
16. Zhang,M.Q. (2002) Computational prediction of eukaryotic protein-coding genes. *Nat. Rev.*, **3**, 698–709.
17. Mathe,C., Sagot,M.-F., Schiex,T. and Rouze,P. (2002) Current methods of gene prediction, their strengths and weaknesses. *Nucleic Acids Res.*, **30**, 4103–4117.
18. Rogic,S., Mackworth,A.K. and Ouellette,F.B. (2001) Evaluation of gene-finding programs on mammalian sequences. *Genome Res.*, **11**, 817–832.
19. Frith,M.C., Bailey,T.L., Kasukawa,T., Mignone,F., Kummerfeld,S.K., Madera,M., Sunkara,S., Furuno,M., Bult,C.J., Quackenbush,J. *et al.* (2006) Discrimination of Non-Protein-Coding Transcripts from Protein-Coding mRNA RIB. *RNA Biol.*, **3**, 40–48.
20. Yandell,M. and Ence,D. (2012) A beginner's guide to eukaryotic genome annotation. *Nat. Rev.*, **13**, 329–342.
21. Brent,M.R. (2007) How does eukaryotic gene prediction work? *Nat. Biotechnol.*, **25**, 883–885.
22. Dunham,I., Shimizu,N., Roe,B.A., Chissoe,S., Hunt,A.R., Collins,J.E., Bruskiewich,R., Beare,D.M., Clamp,M., Smink,L.J. *et al.* (1999) The DNA sequence of human chromosome 22. *Nature*, **402**, 489–495.
23. Salamov,A.A. and Solovyev,V.V. (2000) Ab initio gene finding in Drosophila genomic DNA. *Genome Res.*, **10**, 516–522.
24. Majoros,W.H., Pertea,M. and Salzberg,S.L. (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics*, **20**, 2878–2879.
25. Yada,T., Takagi,T., Totoki,Y., Sakaki,Y. and Takaeda,Y. (2003) Digit: a Novel Gene Finding Program By Combining Gene-Finders. *Biocomput. 2003 - Proc. Pacific Symp.*, **8**, 375–387.
26. Min,X.J., Butler,G., Storms,R. and Tsang,A. (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.*, **33**, W677–W680.
27. Hatzigeorgiou,A.G., Fiziev,P. and Reczko,M. (2001) DIANA-EST: a statistical analysis. *Bioinformatics*, **17**, 913–919.
28. Furuno,M., Kasukawa,T., Saito,R., Adachi,J., Suzuki,H., Baldarelli,R., Hayashizaki,Y. and Okazaki,Y. (2003) CDS annotation in full-length cDNA sequence. *Genome Res.*, **13**, 1478–1487.
29. Iseli,C., Jongeneel,C.V., Bucher,P. and Jongeneel,V. (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol*. Vol. **1999**, pp. 138–148.
30. Fukunishi,Y. and Hayashizaki,Y. (2001) Amino acid translation program for full-length cDNA sequences with frameshift errors. *Physiol. Genomics*, **5**, 81–87.
31. Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R., Wakamatsu,A., Hayashi,K., Sato,H., Nagai,A. *et al.* (2004) Complete sequencing and characterization of 21, 243 full-length human cDNAs. *Nat. Genet.*, **36**, 40–45.
32. Lin,M.F., Jungreis,I. and Kellis,M. (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*, **27**, i275–i282.
33. Andrews,S.J. and Rothnagel,J.a (2014) Emerging evidence for functional peptides encoded by short open reading frames. *Nat. Rev. Genet.*, **15**, 193–204.
34. Ramamurthi,K.S. and Storz,G. (2014) The small protein floodgates are opening; now the functional analysis begins. *BMC Biol.*, **12**, 96.
35. Chng,S.C., Ho,L., Tian,J. and Reversade,B. (2013) ELABELA: a hormone essential for heart development signals via the apelin receptor. *Dev. Cell*, **27**, 672–680.
36. Thierry-Mieg,D. and Thierry-Mieg,J. (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.*, **7**(Suppl. 1), S12–S14.
37. Harrow,J., Frankish,A., Gonzalez,J.M., Tapanari,E., Diekhans,M., Kokocinski,F., Aken,B.L., Barrell,D., Zadissa,A., Searle,S. *et al.* (2012) GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.*, **22**, 1760–1774.
38. Pruitt,K.D., Brown,G.R., Hiatt,S.M., Thibaud-Nissen,F., Astashyn,A., Ermolaeva,O., Farrell,C.M., Hart,J., Landrum,M.J., McGarvey,K.M. *et al.* (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.*, **42**, D756–D763.
39. Cunningham,F., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S. *et al.* (2014) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.
40. Harrow,J.L., Steward,C.A., Frankish,A., Gilbert,J.G., Gonzalez,J.M., Loveland,J.E., Mudge,J., Sheppard,D., Thomas,M., Trevanion,S. *et al.* (2014) The vertebrate genome annotation browser 10 years on. *Nucleic Acids Res.*, **42**, D771–D779.
41. Farrell,C.M., O'Leary,N.A., Harte,R., Loveland,J.E., Wilming,L.G., Wallin,C., Diekhans,M., Barrell,D., Searle,S.M.J.,

Aken,B. *et al.* (2014) Current status and new features of the Consensus Coding Sequence database. *Nucleic Acids Res.*, **42**, 865–872.

42. Wethmar,K., Barbosa-Silva,A., Andrade-Navarro,M.A. and Leutz,A. (2014) uORFdb–a comprehensive literature database on eukaryotic uORF biology. *Nucleic Acids Res.*, **42**, D60–D67.

43. Calvo,S.E., Pagliarini,D.J. and Mootha,V.K. (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.

44. Wethmar,K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip. Rev. RNA*, **5**, 765–778.

45. Klemke,M., Kehlenbach,R.H. and Huttner,W.B. (2001) Two overlapping reading frames in a single exon encode interacting proteins - A novel way of gene usage. *EMBO J.*, **20**, 3849–3860.

46. Autio,K.J., Kastaniotis,A.J., Pospiech,H., Miinalainen,I.J., Schonauer,M.S., Dieckmann,C.L. and Hiltunen,J.K. (2008) An ancient genetic link between vertebrate mitochondrial fatty acid synthesis and RNA processing. *FASEB J.*, **22**, 569–578.

47. Ingolia,N.T., Brar,G.A., Stern-Ginossar,N., Harris,M.S., Talhouarne,G.J.S.S., Jackson,S.E., Wills,M.R. and Weissman,J.S. (2014) Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.*, **8**, 1365–1379.

48. Ingolia,N.T., Lareau,L.F. and Weissman,J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.

49. Ruiz-Orera,J. and Messeguer,X. (2014) Long non-coding RNAs as a source of new peptides. *eLife*, **3**, 03523.

50. Anderson,D.M., Anderson,K.M., Chang,C.-L., Makarewich,C.A., Nelson,B.R., McAnally,J.R., Kasaragod,P., Shelton,J.M., Liou,J., Bassel-Duby,R. *et al.* (2015) A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, **160**, 595–606.

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

52. Mackowiak,S.D., Zauber,H., Bielow,C., Thiel,D., Kutz,K., Calviello,L., Mastrobuoni,G., Rajewsky,N., Kempa,S., Selbach,M. *et al.* (2015) Extensive identification and analysis of conserved small ORFs in animals. *Genome Biol.*, **16**, 179.

53. Pauli,A., Norris,M.L., Valen,E., Chew,G.-L., Gagnon,J.A., Zimmerman,S., Mitchell,a., Ma,J., Dubrulle,J., Reyon,D. *et al.* (2014) Toddler: an embryonic signal that promotes cell movement via apelin receptors. *Science*, **343**, 1248636.

54. Yoshida,H., Matsui,T., Yamamoto,A., Okada,T. and Mori,K. (2001) XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell*, **107**, 881–891.

55. Quelle,D.E., Zindy,F., Ashmun,R.A. and Sherr,C.J. (1995) Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell*, **83**, 993–1000.

56. Chung,W.-Y., Wadhawan,S., Szklarczyk,R., Pond,S.K. and Nekrutenko,A. (2007) A first look at ARFome: dual-coding genes in mammalian genomes. *PLoS Comput. Biol.*, **3**, e91.

57. Ribrioux,S., Brüngger,A., Baumgarten,B., Seuwen,K. and John,M.R. (2008) Bioinformatics prediction of overlapping frameshifted translation products in mammalian transcripts. *BMC Genomics*, **9**, 122.

58. Xu,H., Wang,P., Fu,Y., Zheng,Y., Tang,Q., Si,L., You,J., Zhang,Z., Zhu,Y., Zhou,L. *et al.* (2010) Length of the ORF, position of the first AUG and the Kozak motif are important factors in potential dual-coding transcripts. *Cell Res.*, **20**, 445–457.

59. Vanderperre,B., Lucier,J.-F. and Roucou,X. (2012) HAltORF: a database of predicted out-of-frame alternative open reading frames in human. *Database (Oxford).*, bas025.

60. Vanderperre,B., Lucier,J.-F., Bissonnette,C., Motard,J., Tremblay,G., Vanderperre,S., Wisztorski,M., Salzet,M., Boisvert,F.-M. and Roucou,X. (2013) Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS One*, **8**, e70698.

61. Skarshewski,A., Stanton-Cook,M., Huber,T., Al Mansoori,S., Smith,R., Beatson,S.A. and Rothnagel,J.A. (2014) uPEPperoni: an online tool for upstream open reading frame location and analysis of transcript conservation. *BMC Bioinformatics*, **15**, 36.

62. Lee,S., Liu,B., Lee,S., Huang,S.-X., Shen,B. and Qian,S.-B. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E2424–E2432.

63. Fritsch,C., Herrmann,A., Nothnagel,M., Szafranski,K., Huse,K., Schumann,F., Schreiber,S., Platzer,M., Krawczak,M., Hampe,J. *et al.* (2012) Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.*, **22**, 2208–2218.

64. Ivanov,I.P., Firth,A.E., Michel,A.M., Atkins,J.F. and Baranov,P. V. (2011) Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.*, **39**, 4220–4234.

65. Bab,I., Smith,E., Gavish,H., Attar-Namdar,M., Chorev,M., Chen,Y.C., Muhlrad,A., Birnbaum,M.J., Stein,G. and Frenkel,B. (1999) Biosynthesis of osteogenic growth peptide via alternative translational initiation at AUG85 of histone H4 mRNA. *J. Biol. Chem.*, **274**, 14474–14481.

66. Abramowitz,J., Grenet,D., Birnbaumer,M., Torres,H.N. and Birnbaumer,L. (2004) XLalphas, the extra-long form of the alpha-subunit of the Gs G protein, is significantly longer than suspected, and so is its companion Alex. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 8366–8371.

67. Vanderperre,B., Staskevicius,A.B., Tremblay,G., McCoy,M., O'Neill,M.A., Cashman,N.R. and Roucou,X. (2011) An overlapping reading frame in the PRNP gene encodes a novel polypeptide distinct from the prion protein. *FASEB J.*, **25**, 2373–2386.

68. Bergeron,D., Lapointe,C., Bissonnette,C., Tremblay,G., Motard,J. and Roucou,X. (2013) An out-of-frame overlapping reading frame in the ataxin-1 coding sequence encodes a novel ataxin-1 interacting protein. *J. Biol. Chem.*, **288**, 21824–21835.

69. Lee,C.-f., Lai,H.-L., Lee,Y.-C., Chien,C.-L. and Chern,Y. (2014) The A2A Adenosine Receptor Is a Dual Coding Gene: A novel mechanism of gene usage and signal transduction. *J. Biol. Chem.*, **289**, 1257–1270.

70. Akimoto,C., Sakashita,E., Kasashima,K., Kuroiwa,K., Tominaga,K., Hamamoto,T. and Endo,H. (2013) Translational repression of the McKusick-Kaufman syndrome transcript by unique upstream open reading frames encoding mitochondrial proteins with alternative polyadenylation sites. *Biochim. Biophys. Acta - Gen. Subj.*, **1830**, 2728–2738.

71. Yosten,G, Liu,J., Ji,H., Sandberg,K., Speth,R. and Samson,W.K. (2015) A 5′-Upstream short open reading frame encoded peptide regulates angiotensin type 1a receptor production and signaling via the beta-arrestin pathway. *J. Physiol.*, doi:10.1113/JP270567.

72. Freson,K., Jaeken,J., Van Helvoirt,M., de Zegher,F., Wittevrongel,C., Thys,C., Hoylaerts,M.F., Vermylen,J. and Van Geet,C. (2003) Functional polymorphisms in the paternally expressed XLalphas and its cofactor ALEX decrease their mutual interaction and enhance receptor-mediated cAMP formation. *Hum. Mol. Genet.*, **12**, 1121–1130.

73. Wang,R.F., Johnston,S.L., Zeng,G., Topalian,S.L., Schwartzentruber,D.J., Rosenberg,S.A., Suzanne,L., Schwartzentruber,D.J., Steven,A., Topalian,S.L. *et al.* (1998) A breast and melanoma-shared tumor antigen: T cell responses to antigenic peptides translated from different open reading frames. *J. Immunol.*, **161**, 3598–3606.

74. Wang,R.-F.F., Parkhurst,M.R., Kawakami,Y., Robbins,P.F. and Rosenberg,S.A. (1996) Utilization of an alternative open reading frame of a normal gene in generating a novel human cancer antigen. *J. Exp. Med.*, **183**, 1131–1140.

75. Ronsin,C., Chung-Scott,V., Poullion,I., Aknouche,N., Gaudin,C. and Triebel,F. (1999) A non-AUG-defined alternative open reading frame of the intestinal carboxyl esterase mRNA generates an epitope recognized by renal cell carcinoma-reactive tumor-infiltrating lymphocytes in situ. *J. Immunol.*, **163**, 483–490.

76. Rosenberg,S.A, Tong-On,P., Li,Y., Riley,J.P., El-gamil,M., Parkhurst,M.R. and Robbins,P.F. (2002) Identification of BING-4 cancer antigen translated from an alternative open reading frame of a gene in the extended MHC class II region using lymphocytes from

a patient with a durable complete regression following immunotherapy. *J. Immunol.*, **168**, 2402–2407.

77. Oh,S., Terabe,M., Pendleton,C.D., Bhattacharyya,A., Bera,T.K., Epel,M., Reiter,Y., Phillips,J., Linehan,W.M., Kasten-Sportes,C. *et al.* (2004) Human CTLs to wild-type and enhanced epitopes of a novel prostate and breast tumor-associated protein, TARP, lyse human breast cancer cells. *Cancer Res.*, **64**, 2610–2618.

78. Slager,E.H., Borghi,M., van der Minne,C.E., Aarnoudse,C.A., Havenga,M.J.E., Schrier,P.I., Osanto,S. and Griffioen,M. (2003) CD4+ Th2 cell recognition of HLA-DR-restricted epitopes derived from CAMEL: a tumor antigen translated in an alternative open reading frame. *J. Immunol.*, **170**, 1490–1497.

79. Magrane,M. and Consortium U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database*, **2011**, bar009.

80. Oyama,M., Itagaki,C., Hata,H., Suzuki,Y., Izumi,T., Natsume,T., Isobe,T. and Sugano,S. (2004) Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs. *Genome Res.*, **14**, 2048–2052.

81. Slavoff,S.A, Mitchell,A.J., Schwaid,A.G., Cabili,M.N., Ma,J., Levin,J.Z., Karger,A.D., Budnik,B.a, Rinn,J.L. and Saghatelian,A. (2013) Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.*, **9**, 59–64.

82. Ma,J., Ward,C.C., Jungreis,I., Slavoff,S.A., Schwaid,A.G., Neveu,J., Budnik,B.A., Kellis,M. and Saghatelian,A. (2014) Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *J. Proteome Res.*, **13**, 1757–1765.

83. Ingolia,N.T., Ghaemmaghami,S., Newman,J.R.S. and Weissman,J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.

84. Ingolia,N.T. (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.*, **15**, 205–213.

85. Brar,G.A. and Weissman,J.S. (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.*, **16**, 651–664.

86. Michel,A.M. and Baranov,P.V. (2013) Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip. Rev. RNA*, **4**, 473–490.

87. de Klerk,E., Fokkema,I.F.A.C., Thiadens,K.A.M.H., Goeman,J.J., Palmblad,M., den Dunnen,J.T., von Lindern,M. and 't Hoen,P.A.C. (2015) Assessing the translational landscape of myogenic differentiation by ribosome profiling. *Nucleic Acids Res.*, **43**, 4408–4428.

88. Brar,G.A., Yassour,M., Friedman,N., Regev,A., Ingolia,N.T. and Weissman,J.S. (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*, **335**, 552–557.

89. Andreev,D.E., O'Connor,P.B., Zhdanov,A.V., Dmitriev,R.I., Shatsky,I.N., Papkovsky,D.B. and Baranov,P.V. (2015) Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol.*, **16**, 90.

90. Gao,X., Wan,J., Liu,B., Ma,M., Shen,B. and Qian,S.-B. (2015) Quantitative profiling of initiating ribosomes in vivo. *Nat. Methods*, **12**, 147–153.

91. Aspden,J.L., Eyre-Walker,Y.C., Philips,R.J., Amin,U., Mumtaz,M.A.S., Brocard,M. and Couso,J.-P. (2014) Extensive translation of small ORFs revealed by Poly-Ribo-Seq. *Elife*, doi:10.7554/eLife.03528.

92. Michel,A.M., Choudhury,K.R., Firth,A.E., Ingolia,N.T., Atkins,J.F. and Baranov,P. V. (2012) Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res.*, **22**, 2219–2229.

93. Michel,A.M., Fox,G., M Kiran,A., De Bo,C., O'Connor,P.B.F., Heaphy,S.M., Mullan,J.P.A., Donohue,C.A., Higgins,D.G. and Baranov,P.V. (2014) GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res.*, **42**, D859–D864.

94. Wan,J. and Qian,S.B. (2014) TISdb: A database for alternative translation initiation in mammalian cells. *Nucleic Acids Res.*, **42**, 845–850.

95. Crappé,J., Ndah,E., Koch,A., Steyaert,S., Gawron,D., De Keulenaer,S., De Meester,E., De Meyer,T., Van Criekinge,W., Van Damme,P. *et al.* (2014) PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. *Nucleic Acids Res.*, **43**, e29.

96. Xie,S.-Q., Nie,P., Wang,Y., Wang,H., Li,H., Yang,Z., Liu,Y., Ren,J. and Xie,Z. (2015) RPFdb: a database for genome wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res.*, doi:10.1093/nar/gkv972.

97. Miettinen,T.P. and Björklund,M. (2015) Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3′ untranslated regions. *Nucleic Acids Res.*, **43**, 1019–1034.

98. Young,D.J., Guydosh,N.R., Zhang,F., Hinnebusch,A.G. and Green,R. (2015) Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3′ UTRs In Vivo. *Cell*, **162**, 872–884.

99. Kozak,M. (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.*, **115**, 887–903.

100. Kozak,M. (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, **299**, 1–34.

101. Pöyry,T.A.A., Kaminski,A., Connell,E.J., Fraser,C.S. and Jackson,R.J. (2007) The mechanism of an exceptional case of reinitiation after translation of a long ORF reveals why such events do not generally occur in mammalian mRNA translation. *Genes Dev.*, **21**, 3149–3162.

102. Skabkin,M.A.M., Skabkina,O.V.O., Hellen,C.T.C.U.T. and Pestova,T.V.T. (2013) Reinitiation and other unconventional posttermination events during eukaryotic translation. *Mol. Cell*, **51**, 249–264.

103. Pöyry,T.A.A., Kaminski,A. and Jackson,R.J. (2004) What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes Dev.*, **18**, 62–75.

104. Kozak,M. (2001) Constraints on reinitiation of translation in mammals. *Nucleic Acids Res.*, **29**, 5226–5232.

105. Luukkonen,B.G., Tan,W. and Schwartz,S. (1995) Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *J. Virol.*, **69**, 4086–4094.

106. Schleich,S., Strassburger,K., Janiesch,P.C., Koledachkina,T., Miller,K.K., Haneke,K., Cheng,Y.-S., Küchler,K., Stoecklin,G., Duncan,K.E. *et al.* (2014) DENR-MCT-1 promotes translation re-initiation downstream of uORFs to control tissue growth. *Nature*, **512**, 208–212.

107. Kim,M.-S., Pinto,S.M., Getnet,D., Nirujogi,R.S., Manda,S.S., Chaerkady,R., Madugundu,A.K., Kelkar,D.S., Isserlin,R., Jain,S. *et al.* (2014) A draft map of the human proteome. *Nature*, **509**, 575–581.

108. Andreev,D.E., O'Connor,P.B., Fahey,C., Kenny,E.M., Terenin,I.M., Dmitriev,S.E., Cormican,P., Morris,D.W., Shatsky,I.N. and Baranov,P. V (2015) Translation of 5′ leaders is pervasive in genes resistant to eIF2 repression. *Elife*, **4**, e03971.

109. Smith,E., Meyerrose,T.E., Kohler,T., Namdar-Attar,M., Bab,N., Lahat,O., Noh,T., Li,J., Karaman,M.W., Hacia,J.G. *et al.* (2005) Leaky ribosomal scanning in mammalian genomes: significance of histone H4 alternative translation in vivo. *Nucleic Acids Res.*, **33**, 1298–1308.

110. Firth,A.E. and Brierley,I. (2012) Non-canonical translation in RNA viruses. *J. Gen. Virol.*, **93**, 1385–1409.

111. Racine,T. and Duncan,R. (2010) Facilitated leaky scanning and atypical ribosome shunting direct downstream translation initiation on the tricistronic S1 mRNA of avian reovirus. *Nucleic Acids Res.*, **38**, 7260–7272.

112. Guerrero,S., Batisse,J., Libre,C., Bernacchi,S., Marquet,R. and Paillart,J.-C. (2015) HIV-1 replication and the cellular eukaryotic translation apparatus. *Viruses*, **7**, 199–218.

113. Fitzgerald,K.D. and Semler,B.L. (2009) Bridging IRES elements in mRNAs to the eukaryotic translation apparatus. *Biochim. Biophys. Acta*, **1789**, 518–528.

114. Kozak,M. (2005) A second look at cellular mRNA sequences said to function as internal ribosome entry sites. *Nucleic Acids Res.*, **33**, 6593–6602.

115. Jackson,R.J. (2013) The Current Status of Vertebrate Cellular mRNA IRESs. *Cold Spring Harb. Perspect. Biol.*, **5**, a011569.

116. Gilbert,W.V. (2010) Alternative ways to think about cellular internal ribosome entry. *J. Biol. Chem.*, **285**, 29033–29038.

117. Yueh,A. and Schneider,R.J. (2000) Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA. *Genes Dev.*, **14**, 414–421.

118. Sherrill,K.W. and Lloyd,R.E. (2008) Translation of cIAP2 mRNA is mediated exclusively by a stress-modulated ribosome shunt. *Mol. Cell. Biol.*, **28**, 2011–2022.

119. Hertz,M.I., Landry,D.M., Willis,A.E., Luo,G. and Thompson,S.R. (2013) Ribosomal protein S25 dependency reveals a common mechanism for diverse internal ribosome entry sites and ribosome shunting. *Mol. Cell. Biol.*, **33**, 1016–1026.

120. Wang,X.-Q. and Rothnagel,J.A. (2004) 5′ -untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. *Nucleic Acids Res.*, **32**, 1382–1391.

121. Lee,S. (1991) Expression of growth/differentiation factor 1 in the nervous system: conservation of a bicistronic structure. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 4250–4254.

122. Gray,T., Saitoh,S. and Nicholls,R.D. (1999) An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 5616–5621.

123. Kanamori,Y., Hayakawa,Y., Matsumoto,H., Yasukochi,Y., Shimura,S., Nakahara,Y., Kiuchi,M. and Kamimura,M. (2010) A eukaryotic (insect) tricistronic mRNA encodes three proteins selected by context-dependent scanning. *J. Biol. Chem.*, **285**, 36933–36944.

124. Kondo,T., Hashimoto,Y., Kato,K., Inagaki,S., Hayashi,S. and Kageyama,Y. (2007) Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.*, **9**, 660–665.

125. Savard,J., Marques-Souza,H., Aranda,M. and Tautz,D. (2006) A segmentation gene in tribolium produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell*, **126**, 559–569.

126. Galindo,M.I., Pueyo,J.I., Fouix,S., Bishop,S.A. and Couso,J.P. (2007) Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.*, **5**, e106.

127. Chappell,S.A., Edelman,G.M. and Mauro,V.P. (2006) Ribosomal tethering and clustering as mechanisms for translation initiation. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 18077–18082.

128. Martin,F., Barends,S., Jaeger,S., Schaeffer,L., Prongidi-Fix,L. and Eriani,G. (2011) Cap-Assisted Internal Initiation of Translation of Histone H4. *Mol. Cell*, **41**, 197–209.

129. Paek,K.Y., Hong,K.Y., Ryu,I., Park,S.M., Keum,S.J., Kwon,O.S. and Jang,S.K. (2015) Translation initiation mediated by RNA looping. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 1041–1046.

130. Li,C., Goudy,K., Hirsch,M., Asokan,A., Fan,Y., Alexander,J., Sun,J., Monahan,P., Seiber,D., Sidney,J. *et al.* (2009) Cellular immune response to cryptic epitopes during therapeutic gene transfer. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 10770–10774.

131. Hunt,R.C., Simhadri,V.L., Iandoli,M., Sauna,Z.E. and Kimchi-Sarfaty,C. (2014) Exposing synonymous mutations. *Trends Genet.*, **30**, 308–321.

132. Kochetov,A.V. (2008) Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. *BioEssays*, **30**, 683–691.

133. Landry,C.R., Zhong,X., Nielly-Thibault,L. and Roucou,X. (2015) Found in translation: functions and evolution of a recently discovered alternative proteome. *Curr. Opin. Struct. Biol.*, **32**, 74–80.

134. Koch,A., Gawron,D., Steyaert,S., Ndah,E., Crappé,J., De Keulenaer,S., De Meester,E., Ma,M., Shen,B., Gevaert,K. *et al.* (2014) A proteogenomics approach integrating proteomics and ribosome profiling increases the efficiency of protein identification and enables the discovery of alternative translation start sites. *Proteomics*, **14**, 2688–2698.

135. Gaj,T., Gersbach,C.A. and Barbas,C.F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.*, **31**, 397–405.

136. Stern-Ginossar,N., Weisburd,B., Michalski,A., Le,V.T.K., Hein,M.Y., Huang,S.-X., Ma,M., Shen,B., Qian,S.-B., Hengel,H. *et al.* (2012) Decoding human cytomegalovirus. *Science*, **338**, 1088–1093.