

uORFdb—a comprehensive literature database on eukaryotic uORF biology

Klaus Wethmar^{1,2,*}, Adriano Barbosa-Silva³, Miguel A. Andrade-Navarro³ and Achim Leutz^{1,4}

¹Max Delbrück Center for Molecular Medicine (MDC), Cell Differentiation and Tumorigenesis, Robert-Rössle-Strasse 10, D-13092 Berlin, Germany, ²Hematology, Oncology and Tumor Immunology, Helios Klinikum Berlin-Buch, Schwanebecker Chaussee 50, D-13125 Berlin, Germany, ³Max Delbrück Center for Molecular Medicine (MDC), Computational Biology and Data Mining, Robert-Rössle-Strasse 10, D-13092 Berlin, Germany and ⁴Humoldt-University, Department of Biology, Invalidenstrasse 43, D-10115 Berlin, Germany

Received August 22, 2013; Revised September 25, 2013; Accepted September 26, 2013

ABSTRACT

Approximately half of all human transcripts contain at least one upstream translational initiation site that precedes the main coding sequence (CDS) and gives rise to an upstream open reading frame (uORF). We generated uORFdb, publicly available at <http://cbdm.mdc-berlin.de/tools/uorfdb>, to serve as a comprehensive literature database on eukaryotic uORF biology. Upstream ORFs affect downstream translation by interfering with the unrestrained progression of ribosomes across the transcript leader sequence. Although the first uORF-related translational activity was observed >30 years ago, and an increasing number of studies link defective uORF-mediated translational control to the development of human diseases, the features that determine uORF-mediated regulation of downstream translation are not well understood. The uORFdb was manually curated from all uORF-related literature listed at the PubMed database. It categorizes individual publications by a variety of denominators including taxon, gene and type of study. Furthermore, the database can be filtered for multiple structural and functional uORF-related properties to allow convenient and targeted access to the complex field of eukaryotic uORF biology.

INTRODUCTION

Ribosome profiling of the yeast, mouse and human transcriptomes uncovered high rates of translation beyond the borders of annotated main protein-coding sequences (CDSs) (1–4). Most of these non-protein-coding translational hot spots are localized within the transcript leader sequence of mRNAs (4), where upstream AUG codons or

alternative upstream initiation codons give rise to upstream open reading frames (uORFs). The presence of uORFs, which may overlap or terminate upstream of the main protein CDS, affects downstream initiation efficiency and the translation rate of the respective protein (Figure 1).

The regulatory potential of uORFs has first been described in the 1980s (5); however, only recently, ribosome profiling and a growing list of physiological and medical implications attributed an increased level of biological significance to uORF-mediated translational control (6–9). For example, germ line mutations resulting in the de novo generation or functional activation of uORFs in two prominent tumor suppressor genes (CDKN2A and CDKN1B) were associated with the development of hereditary melanoma and multiple endocrine neoplasia syndrome (MEN4), respectively (9,10).

The vast majority of experiments focused on the functional analysis of AUG-initiated uORFs by luciferase reporter assays and mostly demonstrated inhibitory effects on downstream translation. Exceptionally, uORFs can also mediate the paradoxical induction of downstream protein translation under unfavorable global translational conditions, as intensively studied for the yeast transcription factor GCN4 in response to nutrient stresses (11). A multitude of other uORF-related regulatory functions (12,13),

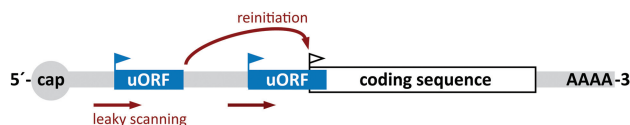


Figure 1. Model of a uORF-containing transcript. Two uORFs (blue boxes) precede the main ORF of the CDS (white box). Ribosomes may initiate at the CDS initiation codon (white flag) after leaky scanning through both uORF initiation codons (blue flags), or may reinitiate after translating the first uORF and leaky scanning through the second uORF initiation site. Ribosomes translating the second overlapping uORF will not be available for translation of the CDS.

*To whom correspondence should be addressed. Tel: +49 30940 62809 or 12199; Fax: +49 30940 63298; Email: k.wethmar@mdc-berlin.de

e.g. uORF-directed selection of downstream translational initiation sites in mammalian key-regulatory transcription factors (14) or the uORF-mediated integration of small molecule concentrations determining downstream translational activity (15–18), may hold abundant novel therapeutic target sites for medical application.

Owing to the overwhelming number and transcript-specific variability of uORF-related properties and functions, the biology of uORFs is far from being understood. With uORFdb, we generated a comprehensive browsable literature database on eukaryotic uORF biology to provide a rapid, targeted and convenient overview of this developing field.

LITERATURE REVIEW AND GENERATION OF THE DATABASE

Since February 2010, we applied a Boolean search for ‘upstream open reading’ or ‘uORF’ or ‘uORFs’ or ‘upstream initiation’ or ‘uAUG’ or ‘small open reading’ or ‘sORF’ or ‘upORF’ or ‘ribosome profiling’ to the NCBI PubMed database at <http://www.ncbi.nlm.nih.gov/pubmed>. On 15 July 2013, this search returned 981 publications. All abstracts were curated manually to eliminate non-related accidental hits. Furthermore, only publications investigating eukaryotic or viral transcripts/uORFs were included, while bacterial data were omitted.

Most importantly, during the curation process, we identified a number of numerical, structural, sequential and cofactor-related properties that were recurrently associated with uORF-mediated regulatory functions. All references were screened and indexed for these newly defined function-related categories. Additionally, publications were categorized by the type of article, by the taxon and by the gene investigated. Wherever required, full-text articles were analyzed to extract missing information according to the uORFdb denominators. All information was collected to build a publicly available browsable database at <http://cbdm.mdc-berlin.de/tools/uorfdb/>.

The initial release of uORFdb provided links to 467 uORF-related references covering a wide range of species/taxa and genes (Table 1). The comprehensive literature survey performed to generate uORFdb revealed that only ~100 of the >10 000 human protein-coding

genes that produce uORF-bearing transcripts have been investigated for uORF-mediated translational control mechanisms. The proportion of analyzed uORF genes for other species is even lower, e.g. ~0.4% for mouse and yeast, and ~0.1% for rat.

Considering the universal prevalence of one or more uORF(s) in ~50% of mRNAs in mammalian transcriptomes, together with the recently proven high rate of uORF-mediated translational activity (4,7), the number of reports on functionally important uORFs is likely to rapidly increase within the next decade of research.

FEATURES OF THE DATABASE

The uORFdb is intended to facilitate convenient and targeted access to the complex field of uORF-mediated translational control mechanisms by a web-based query tool. Making use of manually curated data derived from a review of all PubMed-listed uORF-related literature, users may query uORFdb by three options:

I) Query uORF bibliography by gene or taxon.

A free-text input field at the query page of the web interface allows flexible search inputs, including gene name, gene symbol, gene alias, NCBI Gene/GenBank ID, taxon or taxon common name to identify uORF-related references for a specific gene or taxon.

II) Query uORF bibliography by uORF-related properties.

An individual user-specific literature compilation for one or multiple uORF-related properties can be generated by simple one-click selections of the respective categories on the query page.

III) Query uORF bibliography by manuscript category.

Users may limit returned references to specific manuscript categories, including protocols, review articles and studies characterized by the type of the experimental method applied.

After querying uORFdb, an output page (Figure 2) returns a table summarizing all categories met or addressed by the respective publications. Wherever possible, the output table provides the taxon, official gene symbol and accession number for individual uORF-bearing genes or transcripts, along with links to the corresponding records in the NCBI’s Entrez Gene or Nucleotide databases for further sequence analysis (19). Selection fields next to each reference in the output table allow users to directly display an individual set of abstracts at the PubMed web page for further reading. Query results, as well as the complete content of uORFdb, may be downloaded from the output page and downloads page, respectively.

Table 1. Content of uORFdb v1.0

Taxon	References	Genes
Human	166	103
Yeast	85	15
Mouse	66	43
Virus	50	17
Arabidopsis	28	14
Rat	21	16
Others	52	47

The table summarizes the number of references per taxon and the number of analyzed genes per taxon contained within the initial release of uORFdb. Note that reviews and other manuscript categories, which are not restricted to specific transcripts or taxa are not represented by this table.

TECHNICAL SPECIFICATIONS

The uORFdb is presented as a Web site developed using PHP programming language (version 5.3.2, www.php.net).

uORFdb

- a comprehensive literature database on eukaryotic uORF biology

QUERY RESULTSLoad selected abstracts at

Selection list	PubMed ID	Taxon	Official Symbol	Entrez Gene ID	uORFs per Transcript	Alternative Promoters	Alternative Splicing	Tissue Specific uORFs	Distance from 5'-CAP	CDS Overlap	RNA Secondary Structure	CDS Repression	Ribosome Load	Translational Status	Regulatory Sequence Motif	Disease-related uORFs	Acquired Mutations /SNPs
<input type="checkbox"/>	21289205	Human	ADIPOR1	51094	4	-	✓	✓	-	-	-	✓	-	-	-	✓	-
<input type="checkbox"/>	1400498	Human	AMD1	262	1	-	-	✓	-	-	-	✓	✓	-	-	-	-
<input type="checkbox"/>	8027046	Human	AMD1	262	1	-	-	✓	✓	✗	-	✓	-	-	-	-	-
<input type="checkbox"/>	21873270	Human	DDIT3	1649	1	-	-	✓	-	-	-	✓	-	✓	✓	-	-
<input type="checkbox"/>	17591614	Human	ELK1	2002	>1	-	✓	✓	-	-	✓	✓	✓	-	-	-	-
<input type="checkbox"/>	10216954	Human	ERBB2	2064	1	-	-	✗	-	-	-	✓	✓	-	-	-	-
<input type="checkbox"/>	23806560	Human	ESR2	2100	>1	✓	✓	✓	-	-	-	-	-	-	-	-	✓
<input type="checkbox"/>	12487779	Human	GRAP2	9402	4	✓	✓	✓	-	-	-	✓	-	-	-	-	-
<input type="checkbox"/>	8051167	Human	KLF9	687	10	-	-	✓	-	-	-	✓	-	-	-	-	-
<input type="checkbox"/>	10523842	Human	MDM2	4193	2	✓	-	✓	-	-	-	✓	-	-	-	-	-
<input type="checkbox"/>	15558014	Human	RARB	5915	>1	✓	-	✓	-	-	-	✓	-	-	-	-	-
<input type="checkbox"/>	16325371	Human	SMCP	4184	>1	-	-	✓	-	-	-	-	-	-	-	-	-

download displayed data.

Figure 2. Example of a uORFdb query result. At the first release of uORFdb, filtering for 'human' data and selecting 'tissue specificity' from the uORF-related property section returned 12 PubMed IDs, linked to the respective abstracts. The table summarizes all categories met or addressed within the returned publications. The green check marks indicate 'positive' evidence for a regulatory function of the respective uORF-related property, or that the respective manuscript category is met. The red X symbol identifies 'negative' evidence for a regulatory function of the respective uORF-related property (e.g. PMID 8027046 contains evidence that the overlap of the uORF with the CDS does not influence CDS translation). Users may sort the output table for a category of choice by clicking on the column header (a white arrowhead indicates active sorting). By checking the selection fields on the left, users may select an individual set of abstracts for bulk display via the 'PubMed' button above the table. A yellow funnel sign in the header of the table marks active filters. Links to NCBI Gene or GenBank entries allow further sequence analysis, and query results may be downloaded for local use via a link below the table.

net). On selection of desired filters by the database users, a server-side PHP script builds a correspondent SQL query and executes it on the MySQL system where uORFdb data is stored (MySQL Server version 5.1.61). Matching records are fetched from the MySQL query result and populated into a HTML table to be displayed at the user's browser.

The following section will provide short explanatory and summarizing paragraphs on the individual categories of uORFdb:

DETERMINANTS OF uORF PRESENCE OR ABSENCE

- Alternative promoters
- Alternative splicing
- Tissue-specific uORFs

While AUG is the best conserved trinucleotide within the transcript leader sequence of human and mouse (7), the general prevalence of uAUGs is lower than expected by normal distribution (20). These observations argue for the functional importance of uAUGs and for an evolutionary

negative selection, respectively. In specific cases, the presence or absence of one or several uORF(s) is dependent on the transcript variant produced by transcription initiation from alternative promoters or due to alternative splicing. For example, the predominant usage of an alternative promoter within the oncogene MDM-2 in tumor cells results in the production of a transcript variant lacking exon1 and two inhibitory uORFs, leading to increased translation of MDM-2 protein (21). Tissue-specific presence and functional importance of uORFs have been reported for a number of human and mouse genes including AdipoR1, where a gain of two translational repressive uORFs in a splicing-derived alternative transcript in muscle tissue is implicated in whole-body insulin sensitivity and glucose tolerance (22).

- Non-AUG uORFs

In a recent study using global translational initiation sequencing (4), 54% of human transcripts displayed one or more translational initiation site(s) preceding the CDS. Surprisingly, about three-fourths of upstream translation was initiated by near-cognate, non-AUG initiation

codons, further relativizing the classical 'first-AUG'-role. Nevertheless, uAUG codons appeared to be functionally most effective in repressing CDS translation. To date, only two publications analyzing human BIRC2 and yeast GCN4 have been focusing on non-AUG uORF functions at the individual transcript level (23,24).

STRUCTURAL AND SEQUENCE-DEPENDENT uORF PROPERTIES

• Number • Length • Distance from 5'-cap • Distance from uORF-stop to CDS • CDS overlap • RNA secondary structure

Many publications investigated the importance of structural and sequence-dependent uORF properties in mediating translational regulation. The impact of uORF number, length and position within the transcript leader sequence has most intensively been studied in the classical model for uORF-mediated translational control, the yeast GCN4 transcript (11,25) and in a series of mutational experiments performed by M. Kozak, reviewed in (26). The repression of downstream translation appears to be positively correlated with the number of uORFs per transcript, the length of the uORF and the distance between the 5'-cap structure and the uORF initiation codon. Furthermore, translational repression correlates negatively with the distance between the uORF-stop and the CDS initiation site and is even more profound when the uORF overlaps the CDS initiation codon. Together, the experiments suggest a dynamic regulatory model, where indispensable initiation cofactors detach gradually from ribosomes during the elongation phase of uORF translation, but may be reloaded to allow reinitiation at the CDS.

FUNCTIONAL CONSEQUENCES OF uORF-MEDIATED TRANSLATIONAL CONTROL

• CDS repression • CDS induction • Start site selection

Most uORFs analyzed to date repress translation of the subsequent initiation site(s) and inhibit/diminish translation of the main protein. Post-uORF initiation at the CDS initiation codon may occur after leaky scanning of ribosomes across the uORF initiation codon or by reinitiation if the uORF-stop codon precedes the CDS (26). Despite a generally repressive function on downstream translation, several exceptions have been described, including human DDIT3 (15), mouse Atf4 and yeast GCN4 (11), where translation of specific uORFs or a certain alignment of subsequent uORFs mediate enhanced CDS initiation. Furthermore, uORF-directed start site selection can result in the production of N-terminally distinct protein isoforms that harbor unique biological functions, as demonstrated for CEBPA and CEBPB transcription factors (14,27,28).

• Nonsense-mediated decay • mRNA destabilization

Nonsense-mediated decay (NMD) of mRNA is activated when specific cellular surveillance mechanisms detect premature termination of protein translation (29). Such premature termination events may result from the

use of nonsense codons that arise in mature transcripts due to mutations, incorrect splicing or aberrant initiation site selection. Upstream ORFs have been suggested to induce NMD by conferring additional termination codons to the 5'-leader sequence of certain transcripts. Expression profiling in mammalian cells (30), *Caenorhabditis elegans* (31) and yeast (32) revealed an enrichment of uORF-containing transcripts in the fraction of mRNAs that were targeted by NMD. Similarly, another mode of termination-dependent RNA destabilization that is distinct and independent of the common NMD pathway has been reported in yeast (33,34).

• Ribosome load • Ribosome pausing/stalling • Ribosome shunting

Mutational deletion of a uORF can result in increased ribosome load on a given transcript associated with increased translational activity, as observed for human AMD1 (35) and ERBB2 (36). On the contrary, ribosome stalling at the uORF termination codon or pausing of ribosomes on inhibitory uORF structures (37) may hamper CDS translation. In specific cases, such as the *Arabidopsis* transcription factor GBF6, binding of a small molecule cofactor (sucrose) to the nascent uORF-peptide induced stalling of ribosomes at the uORF termination codon and resulted in decreased translational initiation at the CDS (38). Additional examples of uORF-peptides with regulatory small molecules entail the translational repression of mammalian AMD1 by polyamines (39,40) or repression of yeast CPA1 and *Neurospora crassa* Arg2 by arginine (16).

Underlining the multiplicity of uORF-mediated translational regulation, certain uORFs may facilitate enhanced CDS translation by supporting a ribosome shunt across a highly structured and inhibitory transcript leader sequence, as best studied for Cauliflower mosaic virus 35S RNA (41).

CO-REGULATORY EVENTS AFFECTING uORF FUNCTIONS

• Kozak consensus sequence

Whether or not the ternary preinitiation complex recognizes an AUG or non-AUG triplet as a translational start codon is strongly influenced by the nucleotide context surrounding it. Extensive sequence analysis (20,42) as well as mutational analysis (26,43,44) identified crucial nucleotide residues in the context of an AUG triplet that create favorable or unfavorable surroundings for translational initiation. The optimal surrounding sequence for initiation is GCCRCCAUGG (also called optimal Kozak consensus sequence; R representing a purine base; most important residues underlined). Initiation sequence contexts are frequently classified as strong (both critical residues match the consensus sequence), adequate/intermediate (either residue -3 or +4 matches) or weak (both critical residues do not match) (45). If the AUG codon is surrounded by a strong context, virtually all scanning ribosomes will stop and initiate

translation. When the surrounding context is weak, many ribosomes may scan past the AUG codon and instead initiate at one further downstream. Since the quality of the Kozak consensus sequence is not the only determinant of translation initiation efficiency, the mere evaluation of the surrounding nucleotides does not permit the precise prediction of initiation.

- Translational status

Regulation through uORFs may integrate the overall translation status of a cell and adjust the translation rate of important regulatory proteins. This was first described in a series of experiments on the yeast transcription factor GCN4, where four subsequent uORFs control the paradoxical translation initiation of the main protein while global translation is shut down (11,46–48). Briefly, under favorable translational conditions with high levels of the eIF2–GTP–Met-tRNA^{Met} ternary complex, a fraction of the ribosomes that translate the GCN4 uORF1 reinitiate at the inhibitory uORF4, detach from the mRNA at the uORF4-stop codon and thus inhibit translation of GCN4. Under starving conditions, low availability of the ternary complex causes delayed restoration of a functional pre-initiation ribosomal complex after translation of uORF1. This results in leaky scanning across the uORF4 initiation codon and permits translation of the GCN4 CDS only after prolonged post-termination scanning.

Similar mechanisms depending on the translational status of a cell have been described for the mammalian transcription factors ATF4 (49), ATF5 (50), CEBPA and CEBPB (14), and the macrophage receptor protein CD36 (51).

- Termination (context)

The sequence context surrounding a uORF termination codon may determine the reinitiation efficiency at downstream initiation sites. In particular, stable interactions between the terminating ribosome and the RNA, or stable base pairing of the RNA alone may cause ribosomal pausing or mediate premature mRNA decay (34,52).

- uORF RNA/peptide sequence • Regulatory sequence motif • Cofactor/ribosome interaction

Specific RNA sequences may influence CDS translation by forming stable secondary structures, by binding to a regulatory cofactor or by direct interaction with the translating ribosome. Furthermore, uORF-encoded peptides may induce ribosome stalling and inhibit downstream translation on binding of their respective small molecule interactors, as demonstrated for the sucrose control peptide of *Arabidopsis* GBF6 (38) or the arginine attenuator peptide of *Neurospora* ARG2 (53). For other transcripts, including the HHV-5 gp48 mRNA (54), the DNA damage-inducible transcript 3 (DDIT3/CHOP/CEBP ζ) (55) and the vasopressin V1b receptor (56), translational repression by uORF-encoded peptides has been described without detailed analysis of the mechanism involved. A subset of ~200 human uORFs was suggested to encode unique functional peptides based on a high degree of amino acid sequence conservation (57).

Except for the Kozak consensus sequence, to date only few uORF-related co-regulatory RNA sequence motifs

have been identified. The most prominent example was described for *Drosophila* msl-2, where a protein interaction RNA-motif facilitates binding of the cofactor protein SXL that enhances uORF initiation and thereby represses translation of the CDS (58). In yeast GCN4, reinitiation-promoting elements have been identified surrounding uORF1, which interact with eukaryotic initiation factor 3a to facilitate downstream reinitiation (59). Recently, the h-subunit of eIF3 was found to promote reinitiation after translation of a reinitiation-permissive uORF (60). To what extent ‘specialized ribosomes’ interact with uORFs and other *cis*-regulatory RNA elements to regulate translation awaits investigation (61).

MEDICAL IMPACT

- Disease-related uORFs • Acquired mutations • SNPs

Defects in uORF-mediated translational control may result in the development of human disease. Loss of a uORF in a mutation-related alternative splicing product of the thrombopoietin gene drives enhanced translation of thrombopoietin and causes hereditary thrombocytosis (62). The roles of uORF-related mutations in CDKN2A and CDKN1B for cancer development were mentioned above (9,10). Marie Unna hereditary hair loss is caused by a variety of mutations altering a uORF within the hairless homolog (HR) transcript, resulting in increased expression of hairless homolog protein (8). Additional uORF-altering mutations were identified by computational analysis of the Human Gene Mutation Database (7). Diseases with a confirmed implication of uORF mutations include Cystic fibrosis (CFTR) (63), the van der Woude syndrome (IRF6), hereditary pancreatitis (SPINK1), familial hypercholesterolemia (LDLR) and some others (7). Furthermore, the expression of the beta secretase BACE1, related to Alzheimer’s disease (64), or the transmembrane receptor tyrosine kinase ERBB2, related to breast cancer (65), is at least partially controlled by uORFs. Whether deregulated uORF-mediated translational control is the crucial pathogenic event in these latter cases remains to be established.

Despite few unequivocal cases at this time, it is evident that uORF mutations may be involved in a wide variety of diseases, including malignancies, metabolic or neurologic disorders and inherited syndromes. Considering that many important regulatory proteins, including cell surface receptors, tyrosine kinases and transcription factors, act in a dose-dependent fashion and possess uORFs, we speculate that a substantial number of as yet unexplained pathologies will be traced back to uORF mutations altering expression levels of such key regulatory genes.

MANUSCRIPT CATEGORIES

- Mouse models • Ribosome profiling • Bioinformatics/arrays/screens • Proteomics

To date, two genetically altered mouse models have been generated, confirming the pathogenic role of loss-of-uORF mutations in HR resulting in Marie Unna

hereditary hypotrichosis in humans (66) and validating the physiological importance of the CEBPB uORF in cellular differentiation and proliferation (6), respectively.

Recent progress in computational and sequencing-based technologies and the development of the ribosome profiling method (3) have generated a large amount of information on uORF localization, initiation codon usage and uORF function in response to altered translational conditions (2). Nevertheless, it is yet not possible to predict whether a uORF is translated or has a regulatory role from sequence information only.

Proteomic studies have identified a number of potentially functional uORF-encoded peptides in human cells (67,68). In the human K562 cell line, 40% of small ORF-encoded peptides detected by mass spectrometry originated from transcript leader sequences (69).

OUTLOOK AND FURTHER DEVELOPMENT OF uORFdb

The uORFdb is intended to grow concomitantly to the publication of novel uORF-related literature in respect to the number of references listed and the amount of categorized uORF-related properties. We aim to constantly improve the quality and completeness of indexing applied to individual references and invite users to send feedback, additions and corrections via the contact page of the uORFdb Web site.

ACKNOWLEDGEMENTS

The authors thank Julia Schulz for critical reading of the manuscript.

FUNDING

Deutsche Krebshilfe e.V., Bonn, Germany [110525 to K.W. and A.L.]. Funding for open access charge: Deutsche Krebshilfe e.V., Bonn, Germany [110525 to K.W. and A.L.].

Conflict of interest statement. None declared.

REFERENCES

- Guo,H., Ingolia,N.T., Weissman,J.S. and Bartel,D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.
- Hsieh,A.C., Liu,Y., Edlind,M.P., Ingolia,N.T., Janes,M.R., Sher,A., Shi,E.Y., Stumpf,C.R., Christensen,C., Bonham,M.J. *et al.* (2012) The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature*, **485**, 55–61.
- Ingolia,N.T., Ghaemmaghami,S., Newman,J.R. and Weissman,J.S. (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
- Lee,S., Liu,B., 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. USA*, **109**, E2424–E2432.
- Kozak,M. (1984) Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. *Nucleic Acids Res.*, **12**, 3873–3893.
- Wethmar,K., Begay,V., Smink,J.J., Zaragoza,K., Wiesenthal,V., Dorken,B., Calkhoven,C.F. and Leutz,A. (2010) C/EBPbetaDeltauORF mice—a genetic model for uORF-mediated translational control in mammals. *Genes Dev.*, **24**, 15–20.
- 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. USA*, **106**, 7507–7512.
- Wen,Y., Liu,Y., Xu,Y., Zhao,Y., Hua,R., Wang,K., Sun,M., Li,Y., Yang,S., Zhang,X.J. *et al.* (2009) Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. *Nat. Genet.*, **41**, 228–233.
- Occhi,G., Regazzo,D., Trivellin,G., Boaretto,F., Ciato,D., Bobisse,S., Ferasin,S., Cetani,F., Pardi,E., Korbonits,M. *et al.* (2013) A novel mutation in the upstream open reading frame of the CDKN1B gene causes a MEN4 phenotype. *PLoS Genet.*, **9**, e1003350.
- Liu,L., Dilworth,D., Gao,L., Monzon,J., Summers,A., Lassam,N. and Hogg,D. (1999) Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat. Genet.*, **21**, 128–132.
- Hinnebusch,A.G. (2005) Translational regulation of GCN4 and the general amino acid control of yeast. *Annu. Rev. Microbiol.*, **59**, 407–450.
- Somers,J., Poyry,T. and Willis,A.E. (2013) A perspective on mammalian upstream open reading frame function. *Int. J. Biochem. Cell Biol.*, **45**, 1690–1700.
- Morris,D.R. and Geballe,A.P. (2000) Upstream open reading frames as regulators of mRNA translation. *Mol. Cell. Biol.*, **20**, 8635–8642.
- Calkhoven,C.F., Muller,C. and Leutz,A. (2000) Translational control of C/EBPalpha and C/EBPbeta isoform expression. *Genes Dev.*, **14**, 1920–1932.
- Chen,Y.J., Tan,B.C., Cheng,Y.Y., Chen,J.S. and Lee,S.C. (2010) Differential regulation of CHOP translation by phosphorylated eIF4E under stress conditions. *Nucleic Acids Res.*, **38**, 764–777.
- Hood,H.M., Neafsey,D.E., Galagan,J. and Sachs,M.S. (2009) Evolutionary roles of upstream open reading frames in mediating gene regulation in fungi. *Annu. Rev. Microbiol.*, **63**, 385–409.
- Ivanov,I.P., Atkins,J.F. and Michael,A.J. (2010) A profusion of upstream open reading frame mechanisms in polyamine-responsive translational regulation. *Nucleic Acids Res.*, **38**, 353–359.
- Mize,G.J. and Morris,D.R. (2001) A mammalian sequence-dependent upstream open reading frame mediates polyamine-regulated translation in yeast. *RNA*, **7**, 374–381.
- NCBI Resources Coordinators. (2013) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **41**, D8–D20.
- Iacono,M., Mignone,F. and Pesole,G. (2005) uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene*, **349**, 97–105.
- Brown,C.Y., Mize,G.J., Pineda,M., George,D.L. and Morris,D.R. (1999) Role of two upstream open reading frames in the translational control of oncogene mdm2. *Oncogene*, **18**, 5631–5637.
- Ashwal,R., Hemi,R., Tirosh,A., Gordin,R., Yissachar,E., Cohen-Dayag,A., Rosenberg,A., Karasik,A., Bluhner,M. and Kanety,H. (2011) Differential expression of novel adiponectin receptor-1 transcripts in skeletal muscle of subjects with normal glucose tolerance and type 2 diabetes. *Diabetes*, **60**, 936–946.
- Warnakulasuriyarachchi,D., Ungureanu,N.H. and Holcik,M. (2003) The translation of an antiapoptotic protein HIAP2 is regulated by an upstream open reading frame. *Cell Death Differ.*, **10**, 899–904.
- Cairns,V.R., DeMaria,C.T., Poulin,F., Sancho,J., Liu,P., Zhang,J., Campos-Rivera,J., Karey,K.P. and Estes,S. (2011) Utilization of non-AUG initiation codons in a flow cytometric method for efficient selection of recombinant cell lines. *Biotechnol. Bioeng.*, **108**, 2611–2622.
- Abastado,J.P., Miller,P.F., Jackson,B.M. and Hinnebusch,A.G. (1991) Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for GCN4 translational control. *Mol. Cell. Biol.*, **11**, 486–496.

26. Kozak, M. (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, **299**, 1–34.
27. Smink, J.J., Begay, V., Schoenmaker, T., Sterneck, E., de Vries, T.J. and Leutz, A. (2009) Transcription factor C/EBPbeta isoform ratio regulates osteoclastogenesis through MafB. *EMBO J.*, **28**, 1769–1781.
28. Wethmar, K., Smink, J.J. and Leutz, A. (2010) Upstream open reading frames: molecular switches in (patho)physiology. *Bioessays*, **32**, 885–893.
29. Isken, O. and Maquat, L.E. (2007) Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.*, **21**, 1833–1856.
30. Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F. and Dietz, H.C. (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.*, **36**, 1073–1078.
31. Ramani, A.K., Nelson, A.C., Kapranov, P., Bell, L., Gingeras, T.R. and Fraser, A.G. (2009) High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective *Caenorhabditis elegans*. *Genome Biol.*, **10**, R101.
32. He, F., Li, X., Spatrick, P., Casillo, R., Dong, S. and Jacobson, A. (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol. Cell*, **12**, 1439–1452.
33. Vilela, C., Linz, B., Rodrigues-Pousada, C. and McCarthy, J.E. (1998) The yeast transcription factor genes YAP1 and YAP2 are subject to differential control at the levels of both translation and mRNA stability. *Nucleic Acids Res.*, **26**, 1150–1159.
34. Vilela, C., Ramirez, C.V., Linz, B., Rodrigues-Pousada, C. and McCarthy, J.E. (1999) Post-termination ribosome interactions with the 5'UTR modulate yeast mRNA stability. *EMBO J.*, **18**, 3139–3152.
35. Hill, J.R. and Morris, D.R. (1992) Cell-specific translation of S-adenosylmethionine decarboxylase mRNA. Regulation by the 5' transcript leader. *J. Biol. Chem.*, **267**, 21886–21893.
36. Child, S.J., Miller, M.K. and Geballe, A.P. (1999) Translational control by an upstream open reading frame in the HER-2/neu transcript. *J. Biol. Chem.*, **274**, 24335–24341.
37. Kozak, M. (2001) Constraints on reinitiation of translation in mammals. *Nucleic Acids Res.*, **29**, 5226–5232.
38. Rahmani, F., Hummel, M., Schuurmans, J., Wiese-Klinkenberg, A., Smeekens, S. and Hanson, J. (2009) Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiol.*, **150**, 1356–1367.
39. Raney, A., Baron, A.C., Mize, G.J., Law, G.L. and Morris, D.R. (2000) *In vitro* translation of the upstream open reading frame in the mammalian mRNA encoding S-adenosylmethionine decarboxylase. *J. Biol. Chem.*, **275**, 24444–24450.
40. Hanfrey, C., Elliott, K.A., Franceschetti, M., Mayer, M.J., Illingworth, C. and Michael, A.J. (2005) A dual upstream open reading frame-based autoregulatory circuit controlling polyamine-responsive translation. *J. Biol. Chem.*, **280**, 39229–39237.
41. Pooggin, M.M., Futterer, J., Skryabin, K.G. and Hohn, T. (2001) Ribosome shunt is essential for infectivity of cauliflower mosaic virus. *Proc. Natl Acad. Sci. USA*, **98**, 886–891.
42. Pesole, G., Gissi, C., Grillo, G., Licciulli, F., Liuni, S. and Saccone, C. (2000) Analysis of oligonucleotide AUG start codon context in eukaryotic mRNAs. *Gene*, **261**, 85–91.
43. Kozak, M. (1987) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.*, **196**, 947–950.
44. Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283–292.
45. Meijer, H.A. and Thomas, A.A. (2002) Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochem. J.*, **367**, 1–11.
46. Hinnebusch, A.G. (1984) Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl Acad. Sci. USA*, **81**, 6442–6446.
47. Hinnebusch, A.G. (1997) Translational regulation of yeast GCN4. A window on factors that control initiator-tRNA binding to the ribosome. *J. Biol. Chem.*, **272**, 21661–21664.
48. Mueller, P.P. and Hinnebusch, A.G. (1986) Multiple upstream AUG codons mediate translational control of GCN4. *Cell*, **45**, 201–207.
49. Harding, H.P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M. and Ron, D. (2000) Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell*, **6**, 1099–1108.
50. Zhou, D., Palam, L.R., Jiang, L., Narasimhan, J., Staschke, K.A. and Wek, R.C. (2008) Phosphorylation of eIF2 directs ATF5 translational control in response to diverse stress conditions. *J. Biol. Chem.*, **283**, 7064–7073.
51. Griffin, E., Re, A., Hamel, N., Fu, C., Bush, H., McCaffrey, T. and Asch, A.S. (2001) A link between diabetes and atherosclerosis: glucose regulates expression of CD36 at the level of translation. *Nat. Med.*, **7**, 840–846.
52. Grant, C.M. and Hinnebusch, A.G. (1994) Effect of sequence context at stop codons on efficiency of reinitiation in GCN4 translational control. *Mol. Cell Biol.*, **14**, 606–618.
53. 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.
54. Janzen, D.M., Frolova, L. and Geballe, A.P. (2002) Inhibition of translation termination mediated by an interaction of eukaryotic release factor 1 with a nascent peptidyl-tRNA. *Mol. Cell Biol.*, **22**, 8562–8570.
55. Jousse, C., Bruhat, A., Carraro, V., Urano, F., Ferrara, M., Ron, D. and Fafournoux, P. (2001) Inhibition of CHOP translation by a peptide encoded by an open reading frame localized in the chop 5'UTR. *Nucleic Acids Res.*, **29**, 4341.
56. Rabadan-Diehl, C., Martinez, A., Volpi, S., Subburaju, S. and Aguilera, G. (2007) Inhibition of vasopressin V1b receptor translation by upstream open reading frames in the 5'-untranslated region. *J. Neuroendocrinol.*, **19**, 309–319.
57. Crowe, M.L., Wang, X.Q. and Rothnagel, J.A. (2006) Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides. *BMC Genomics*, **7**, 16.
58. Medenbach, J., Seiler, M. and Hentze, M.W. (2011) Translational control via protein-regulated upstream open reading frames. *Cell*, **145**, 902–913.
59. Munzarova, V., Panek, J., Gunisova, S., Danyi, I., Szamecz, B. and Valasek, L.S. (2011) Translation reinitiation relies on the interaction between eIF3a/TIF32 and progressively folded cis-acting mRNA elements preceding short uORFs. *PLoS Genet.*, **7**, e1002137.
60. Roy, B., Vaughn, J.N., Kim, B.H., Zhou, F., Gilchrist, M.A. and Von Arnim, A.G. (2010) The h subunit of eIF3 promotes reinitiation competence during translation of mRNAs harboring upstream open reading frames. *RNA*, **16**, 748–761.
61. Xue, S. and Barna, M. (2012) Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.*, **13**, 355–369.
62. Wiestner, A., Schlemper, R.J., van der Maas, A.P. and Skoda, R.C. (1998) An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. *Nat. Genet.*, **18**, 49–52.
63. Lukowski, S.W., Bombieri, C. and Trezise, A.E. (2011) Disrupted posttranscriptional regulation of the cystic fibrosis transmembrane conductance regulator (CFTR) by a 5'UTR mutation is associated with a cfr-related disease. *Hum. Mutat.*, **32**, E2266–E2282.
64. Zhou, W. and Song, W. (2006) Leaky scanning and reinitiation regulate BACE1 gene expression. *Mol. Cell Biol.*, **26**, 3353–3364.
65. Spevak, C.C., Park, E.H., Geballe, A.P., Pelletier, J. and Sachs, M.S. (2006) her-2 upstream open reading frame effects on the use of downstream initiation codons. *Biochem. Biophys. Res. Commun.*, **350**, 834–841.
66. Baek, I.C., Kim, J.K., Cho, K.H., Cha, D.S., Cho, J.W., Park, J.K., Song, C.W. and Yoon, S.K. (2009) A novel mutation in Hr causes abnormal hair follicle morphogenesis in hairpoor mouse, an animal model for Marie Unna Hereditary Hypotrichosis. *Mamm. Genome*, **20**, 350–358.

67. 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.
68. Oyama, M., Kozuka-Hata, H., Suzuki, Y., Semba, K., Yamamoto, T. and Sugano, S. (2007) Diversity of translation start sites may define increased complexity of the human short ORFeome. *Mol. Cell. Proteomics*, **6**, 1000–1006.
69. 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.