ANNUAL REVIEWS

Annual Review of Biochemistry

Eukaryotic Microproteins

Nadiya Jaunbocus,[1,2,*] Valerie Ebenki,[1,2,*] Haomiao Su,[1,2] and Sarah A. Slavoff[1,2,3]

[1]Department of Chemistry, Yale University, New Haven, Connecticut, USA; email: sarah.slavoff@yale.edu

[2]Institute for Biomolecular Design and Discovery, Yale University, West Haven, Connecticut, USA

[3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, USA

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

ews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

Annu. Rev. Biochem. 2025. 94:1–28

First published as a Review in Advance on April 17, 2025

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

https://doi.org/10.1146/annurev-biochem-080124-012840

Copyright © 2025 by the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information.

*These authors contributed equally to this article

OPEN ACCESS CC BY

Keywords

microprotein, small open reading frame, smORF

Abstract

Microproteins are polypeptides of 100–150 amino acids or fewer that have not been annotated by genome annotation consortia, given their small size and other noncanonical properties. Translated microproteins are now known to number in the thousands in the human genome, to function in critical cellular and physiological processes, and to be dysregulated or mutated in diseases including neurodegeneration and cancer. Knowledge about microproteins has rapidly accumulated since the advent of ribosome profiling enabled their global discovery 15 years ago. In this review, we summarize what is known about eukaryotic microprotein discovery, the sequences and expression mechanisms of small open reading frames, and microprotein functions from yeast to human.

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

Contents

INTRODUCTION

The Human Genome Project (HGP) was a collaborative effort launched in the late twentieth century to provide a comprehensive blueprint of the human genome—and, by extension, predicted proteome—and elucidate the molecular basis of human biology and disease (1). A major outcome of the HGP was the prediction of approximately 20,000 human protein coding genes based on numerous parametric assumptions, namely, protein size [>100 amino acids (aa)], an AUG start codon, monocistronic transcripts, and no overlap in open reading frames (ORFs) (2). These assumptions aimed to rigorously exclude abundant, yet nonexpressed, random background ORFs, which tend to be short on average (3). However, these rules were refractory to detecting bona fide small ORFs (smORFs) encoding microproteins below the imposed size limit, as well as other classes of noncanonical genes.

A series of key discoveries led to the recognition of the ubiquity and importance of smORFs. In a seminal study, Marilyn Kozak (4) cloned hundreds of human cDNAs, revealing the presence of initiation codons in the 5′ untranslated region (UTR) and suggesting that translation could occur outside of protein coding regions. Subsequent work revealed that these upstream ORFs (uORFs) downregulate translation of downstream cistrons via multiple mechanisms (5). Later, Oyama et al. (6, 7) provided proteomic evidence for the existence of a handful of smORF-derived microproteins in human cell lines, indicating that they could stably persist in cells at the protein level. These studies provided early evidence for microprotein translation beyond annotated protein coding

sequences (CDSs). Around the same time, computational and genetic evidence for expression of smORFs (8, 9), combined with individual examples of functional smORF-encoded microproteins arising from presumptively noncoding genomic regions, was identified in multiple organisms. Key examples included the *tal/pri*-encoded peptides in insects and humanin in human (9–11). These discoveries set the stage for discoveries of functional microproteins, which we discuss throughout this article. Prokaryotic smORFs and microproteins have been reviewed elsewhere (12, 13), so we focus on eukaryotic microproteins in this article. Although the precise number of smORF-encoded microproteins in eukaryotic genomes remains debated, there is now substantial evidence that thousands of smORFs are translated in eukaryotic cells and that some play critical roles in biological processes and human disease (14).

MICROPROTEIN DEFINITION AND NOMENCLATURE

As discussed above, sporadic discoveries of expressed smORFs occurred before the recognition of their ubiquity. Many of these studies—summarized in **Supplemental Table 1**—introduced independent terminology for smORFs and their protein products (2, 3, 13, 15–39). The various terms introduced over the past ~15 years have complicated efforts to systematize the study of smORFs (14, 23). Nonetheless, we posit that these terms retain value in connoting the unique features of the novel genes or gene classes they were originally used to describe, which vary in size and properties, and in clearly distinguishing these recently discovered genes from previously known, canonical protein CDSs of comparable sizes. It thus may be valid to continue to use these various terms in their relevant contexts. It is advisable that future reports clearly define chosen terminology for smORFs and simultaneously reference other commonly utilized terms for smORFs and their products to facilitate keyword searches. In this review, we define microproteins as smORF-encoded sequences of <150 aa that were not annotated at the time of their discovery.

Supplemental Material >

METHODS AND RESOURCES FOR DISCOVERING EUKARYOTIC MICROPROTEINS

Experimental methods have been developed to discriminate translated smORFs from background noise. In this section, we discuss the development of and current advances in methods for candidate smORF discovery in eukaryotes.

Ribosome Profiling

Ribosome profiling, also known as ribosome sequencing (Ribo-seq) (40, 41), has emerged as the leading method for identification of translated smORFs on a global scale. The broad adoption of Ribo-seq can be attributed to its ability to detect thousands of translated ORFs outside annotated CDSs (20, 42–47), as well as quantitate changes in translation under specific conditions and in disease (21, 48, 49).

Ribo-seq refers to the nucleotide-precision identification of footprints of elongating ribosomes arrested with translation elongation inhibitors such as cycloheximide. Arrested ribosomes are then enriched in complex with RNA, followed by nuclease digestion of unprotected RNA, isolation of ribosome-protected footprints from the resulting purified monosomes, and deep sequencing. The resulting data must be further processed to infer translated ORFs. Lastly, rigorous data analysis, analysis of experimental replicates, and high-resolution datasets are essential for determining smORF translation using Ribo-seq, because their short lengths and lower abundance lead to a lower signal-to-noise ratio in smORF-mapped reads (42).

Although Ribo-seq is powerful for identifying novel coding regions, elongation inhibitors like cycloheximide are not well-suited to deconvolution of some translation initiation sites, especially

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

for ORFs with multiple start sites or overlapping reading frames (50). However, translation initiation sequencing (TI-seq) offers profiling of the footprints of initiating eukaryotic ribosomes using specific inhibitors (51–54). This allows deconvolution of smORF translation initiation from larger main ORFs in multicistronic messenger RNAs (mRNAs) and is especially important for detecting nested and out-of-frame smORFs filtered out by scoring algorithms relying on parameters like three-nucleotide periodicity of P-site occupancies of ribosomes translating a CDS (which is violated by overlapping, out-of-frame ORFs). TI-seq can be combined with Ribo-seq to identify translated smORFs with higher confidence (55).

Low study-to-study concordance is often observed in translated smORF identifications (14, 56); this is in part due to the low signal-to-noise ratio for smORF ribosome footprints as a result of their translation by monosomes (57, 58), as well as variability in smORF identification by different algorithms even when analyzing the same input dataset (56). Some studies recommend analysis of replicate data (56, 59) and intersecting the results of multiple types of data analysis software (56) as best practices. A broad effort to curate translated human smORFs following rigorous guidelines has led to the first publicly available, high-confidence Ribo-seq translatome resource (14).

Mass Spectrometry

Mass spectrometry–based proteomics pipelines tailored for direct detection of smORF translation products have been devised (15, 22, 60, 61). These methods require specialized sample preparation, mass spectrometry, and data analysis methods for sensitive detection of unannotated proteins and microproteins. These approaches have been recently reviewed in detail (18, 62–64), so we provide a brief overview with a focus on smORF-specific innovations and applications.

A necessary advance was development of methods to enrich microproteins from whole proteomes for sensitive detection. Given their short lengths, when digested with trypsin, microproteins typically generate only one, or few, unique tryptic peptides. As a result, microprotein proteomics is a particularly demanding application requiring that the detected tryptic peptide be well-separated from other ions (that is, avoidance of complex spectra or ion suppression). It is therefore critical to enrich microproteins and remove larger proteins, which generate numerous tryptic peptides per protein molecule, from the sample (15, 65–71). Deeper coverage of microproteins is further afforded via application of several sample preparation protocols in parallel (66, 67, 70, 71), namely two-dimensional liquid chromatography–mass spectrometry with offline fractionation (18), subcellular fractionation (72), and chemical labeling (73, 74). In sum, sample preparation is a key consideration for detection of microproteins in whole-proteome extracts. Supporting this idea, high-quality evidence for microprotein expression has been obtained from peptidomic analysis of peptides displayed on major histocompatibility complex class I (MHC I) and, in human, human leukocyte antigen class I (HLA I), which present peptides derived from proteolysis of intracellular proteins (75, 76). The sequence constraints for HLA I binding to peptide epitopes, as well as the instrument parameters required for sequencing nontryptic HLA I–associated peptides, may afford effective normalization of microproteins to canonical proteins, enhancing their detection.

Historically, microprotein sequences were absent from the proteome databases utilized for protein identification by peptide-spectral matching (PSM). Specialized databases must therefore be constructed to identify microproteins. Most databases for eukaryotic smORF proteomics have been constructed from conceptual translation of transcriptomes or predicted smORFeomes—which contain the annotated and/or smORF translatomes but also contain many entries that are not really expressed (18). Importantly, this database expansion generates false positive

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

identifications, and thus peptide-spectral matches cannot be considered conclusive evidence for microprotein expression, especially when only one proteotypic peptide is identified. The intersection of proteomics and ribosome profiling has advanced smORF discovery, particularly in the creation of such translatome databases as the above, which ideally would contain no extraneous entries that can lead to false positive identifications and thus proving optimal for microprotein identification via PSM (75).

Overall, both Ribo-seq and mass spectrometry–based smORF discovery platforms can identify only candidate smORFs and microproteins, since both methods are indirect and rely on statistical scoring. smORF validation requires molecular and genetic approaches, such as development of specific antibodies, smORF silencing or genetic deletion, and/or epitope tag knock-in at the genomic smORF locus. Nonetheless, proteomics has demonstrated utility not only in providing supporting evidence for smORF translation but also for probing the chemical properties, reactivity, localization, interactions, and posttranslational modifications of microproteins that are challenging to predict with bioinformatic tools (64, 77, 78).

On the Discordance Between Proteomics and Ribo-seq

Proteomic microprotein discovery methods, when operated in shotgun mode and on whole-cell protein extracts, typically detect one to two orders of magnitude fewer smORFs than ribosome profiling—or less (79)—and it is useful to consider the reasons for this discordance. First, shotgun proteomics, even with optimized protocols, is relatively insensitive to detection of membrane proteins, given their hydrophobic nature that leads to production of few tryptic fragments (72), as well as basic proteins, which are overdigested by trypsin. Both of these motifs can be enriched in smORFs, which are encoded in A/T-rich intergenic regions that tend to result in codons for hydrophobic and aromatic amino acids (80, 81), as well as alternative reading frames, which tend to be enriched in disorder-promoting and charged amino acids (82). Second, due to their short lengths, microproteins tend to be detected by only a single proteotypic tryptic peptide, or only a few (15); not only does this decrease the statistical confidence to identify microproteins via PSM, but it also increases the likelihood of coelution with canonical peptides, which decreases spectral quality for microprotein identification. This problem is exacerbated for low-abundance proteins, which are also refractory to detection with shotgun proteomics due to stochastic selection of parent ions for fragmentation and sequencing (72). Finally, specialized or targeted methods are typically required to detect secreted microproteins (21, 83). As a result, proteomics may have a ceiling to its sensitivity for smORF-encoded microprotein detection. Thus, ribosome profiling offers increased sensitivity for identifying some classes—perhaps the most abundant classes—of translated microproteins.

It is also possible that some translated smORFs do not contribute to the stable cellular proteome. For example, as discussed above, peptides derived from smORFs are displayed on HLA I, suggesting that they are involved in self- versus nonself-discrimination by the adaptive immune system (75, 84–86). Along the same lines, translation products arising from A/T-rich genomic regions tend to generate hydrophobic C-terminal epitopes that may be triaged for proteasomal degradation by the GET complex (81). Furthermore, although yeast de novo genes, some of which are smORFs, exhibit low endogenous expression levels, they can exhibit cellular phenotypes and can increase yeast cell fitness when overexpressed (59, 80). Taken together, these studies suggest that poorly expressed and/or unstable microproteins may not contribute to the stable, cellular proteome, yet may still be biologically significant, either contributing to the immunopeptidome or having adaptive potential on evolutionary timescales. In this sense, the smORF translatome may be distinct from the canonical proteome in stability and functionality

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

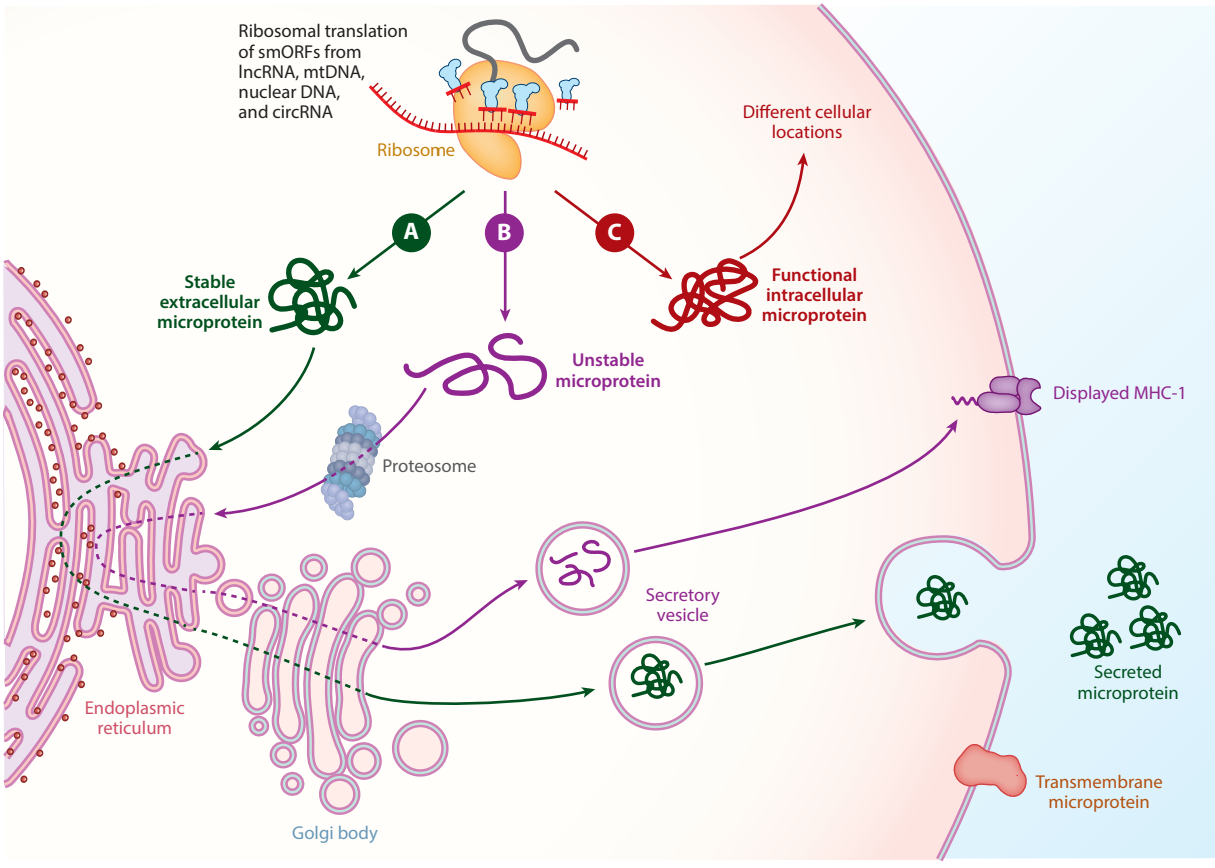

Figure 1

Microprotein translation and trafficking. Ribosomal translation of smORFs from various sources produces microproteins with fates falling in three broad categories. A: Microproteins with extracellular functions can be secreted via canonical or noncanonical pathways. B: Some smORF-encoded proteins are unstable and thus degraded by the proteasome. Some degradation products can be displayed on the cell surface as part of the MHC I complex. C: Stable intracellular microproteins are trafficked to various organelles. Abbreviations: circRNA, circular RNA; lncRNA, long noncoding RNA; MHC I, major histocompatibility complex class I; mtDNA, mitochondrial DNA; smORF, small open reading frame.

(59); this property may speculatively contribute to differences in smORF detection between ribosome profiling and shotgun proteomics, where the former detects all translated smORFs and the latter detects only the most stable cellular microproteins capable of producing unique tryptic peptides. **Figure 1** depicts the fates of stable versus unstable microproteins.

Genetic Screens

Genetic screens have been applied to query smORF functionality at scale. A clustered regularly interspaced short palindromic repeat (CRISPR) screen in human induced pluripotent stem cells and K562 cells targeting ~2,000 smORFs, as well as protein N-terminal extensions, found that ~500 noncanonical ORFs exhibit cell growth phenotypes when knocked out (20), demonstrating for the first time that human smORFs are broadly functional. A subsequent CRISPR screen in cancer cell lines demonstrated viability defects for ~10% of targeted smORFs and noncanonical ORFs; it further employed guide RNA tiling to assign phenotypes to translation of the novel

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

ORFs and not to cryptic DNA or RNA regulatory elements (16). Finally, overexpression screens have been employed in yeast to demonstrate that some smORFs, even when they are species-specific, can improve fitness when overexpressed (80). Genome-wide screens have thus provided clear evidence that smORFs are broadly functional in cells and will continue to find utility in species-, cell-, tissue-, and disease-specific contexts.

Computation and Conservation

Although computational methods have made outsize contributions to prokaryotic smORF identification (87), they have also demonstrated utility for eukaryotic smORF prediction despite the greater complexity of eukaryotic genomes. For example, early studies utilized computational prediction to identify hundreds to thousands of smORFs in mouse (9) and yeast (8). Furthermore, computational methods like PhyloCSF have found application in identifying highly conserved, and thus likely functional, smORFs (70, 88).

smORF CLASSES

Unannotated smORFs have been identified in varied genomic loci (**Figures 2** and **3**) and categorized accordingly, with implications for the evolution, expression, and function of the encoded microproteins. Indeed, several years ago, an insightful review discussed the classes of smORFs in the context of the functional relevance of their genomic locations (89); herein, we update this framework.

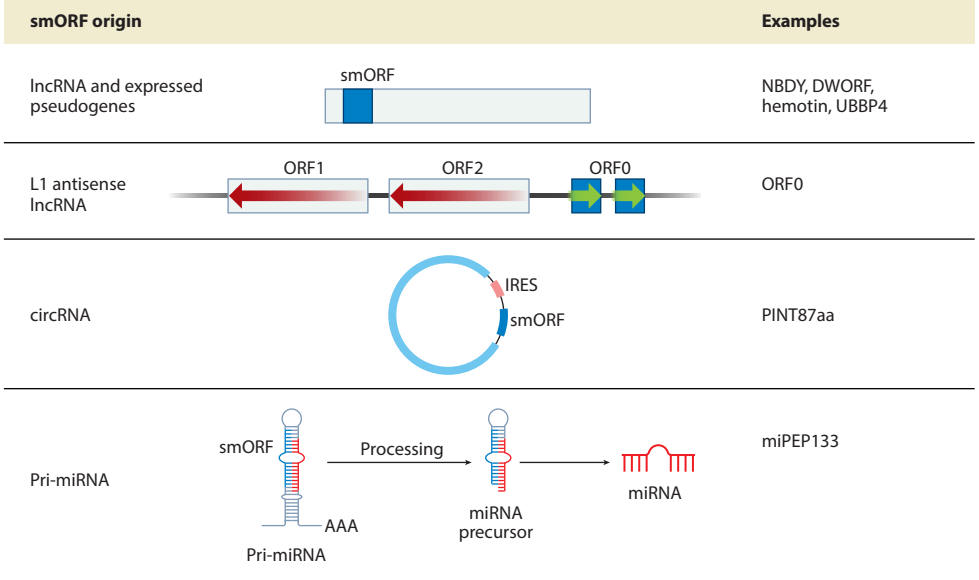

Figure 2

smORFs within noncoding transcripts. Microproteins can be translated from multiple classes of noncoding transcripts. Specifically, many lncRNAs and pseudogenes are now known to produce biologically functional peptides. Antisense transcription of LINE-1 retrotransposons can also result in a translatable smORF. In circRNAs, smORFs under control of an IRES can undergo translation. Pri-miRNA transcripts provide yet another source for microprotein synthesis prior to Drosha processing of pri-miRNAs into miRNA precursors and finally into mature miRNAs. Abbreviations: circRNA, circular RNA; IRES, internal ribosome entry site; L1, LINE-1; lncRNA, long noncoding RNA; miRNA, microRNA; pri-miRNA, primary miRNA; smORF, small open reading frame.

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

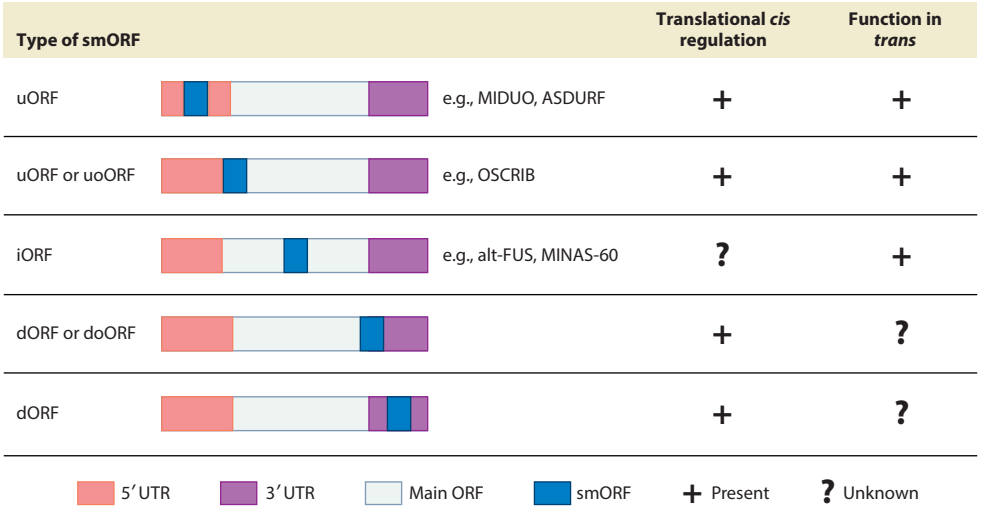

Figure 3

smORFs within mRNA. Unannotated ORFs encoding microproteins have been detected in canonical genes. They can be further categorized according to their specific location vis-à-vis the main ORF. uORFs are found in the 5′ UTR, whereas uoORFs overlap the canonical ORF in an alternative reading frame. Similarly, dORFs are located in the 3′ UTR, and doORFs lie at the intersection of the 3′ UTR and main ORF. Nested within the major coding sequence are iORFs, which are out-of-frame. ORFs in the UTRs are known to regulate translation of the canonical ORF, and the production of functional microproteins has been documented for some uORFs, uoORFs, and iORFs. Abbreviations: dORF, downstream ORF; doORF, downstream overlapping ORF; iORF, internal ORF; mRNA, messenger RNA; ORF, open reading frame; smORF, small ORF; uORF, upstream ORF; uoORF, upstream overlapping ORF; UTR, untranslated region.

Noncoding RNA smORFs

Perhaps best studied are smORFs assigned to noncoding genomic regions, such as long noncoding RNAs (lncRNAs), as well as other classes of ncRNA including microRNA precursors, promoter-derived lncRNAs, transposons, endogenous retroviruses, and transcribed pseudogenes (**Figure 2**). Some estimates suggest as many as 40% of lncRNAs may be translated (45) [and others, few to none (90)]; at the same time, dozens of cytoplasmic, polyadenylated lncRNAs can be found in complex with translating ribosomes in some cell types (91), and even nonpolyadenylated lncRNAs such as MALAT1 can be translated by cytoplasmic ribosomes under specific conditions (92). Despite the ongoing refinement of catalogs of coding lncRNAs, many microproteins encoded in lncRNAs have yielded to functional characterization in human and mouse because of their tractability to genetic perturbation. In general, because polyadenylated lncRNAs are, in a sense, mRNA-like, the microproteins they produce tend to function like small, canonical proteins and peptides. For example, the *NBDY* gene encoding a 7-kDa microprotein regulator of mRNA decapping and P-body formation was originally annotated as lncRNA *LOC550643* (38, 77, 93).

Additionally, microRNA precursor-encoded peptides (miPEPs) have been reported to encode plant microproteins (94). Several circular RNAs (circRNAs) have also been proposed to be translated (95), which would require internal translation initiation. A translated microprotein assigned to *orf0*, a smORF transcribed in the opposite direction from the promoter for the LINE-1 retrotransposon, has also been predicted (96). Another class of smORFs encoded within antisense transcripts, such as PIGB opposite strand (PIGBOS), which localizes to endoplasmic reticulum (ER)-mitochondria contact sites, has been described (97). Importantly, in order to be translated,

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

lncRNAs must localize to the cytoplasm and associate with polysomes; in fact, acquisition of nuclear export has been proposed as a critical step in the evolution of de novo genes from previously noncoding, nuclear transcripts (98).

smORFs Within mRNA

Translated smORFs can also arise within mRNA and are categorized on the basis of their location with respect to the annotated CDS (**Figure 3**). smORFs can regulate translation of the CDS in *cis* and sometimes also produce a functional microprotein. Translation of smORFs within mRNA can thus result in complex biological outcomes. Because they produce more than one protein product, these genes are dual-coding (33, 99).

First, we consider uORFs, smORFs that initiate at start codons within the 5′ UTR and terminate prior to the start site of the downstream CDS. smORFs can also initiate at an upstream, out-of-frame start codon, extending into an alternative reading frame that partially overlaps the CDS; these are sometimes termed upstream overlapping ORFs (uoORFs). In most cases, the uoORF terminates at a stop codon in its unique reading frame within the overlapping region and is thus shorter than the co-encoded protein; in some cases, a uoORF can continue past the stop codon of the CDS and terminate downstream. Subsequent to Kozak's (4) identification of uORFs in hundreds of human transcript leaders, it was revealed (5, 100) that uORFs and uoORFs generally downregulate translation of downstream cistrons via multiple mechanisms, including competitive translation, mRNA destabilization via nonsense-mediated decay, and interaction of the translated uORF or uoORF peptide product with the ribosome. Specific examples, such as the translational regulation of *GCN4* by uORFs in yeast (101) and translational repression of mammalian *ATF4* by a uoORF (102), demonstrate that uORF versus CDS translation can be rebalanced in response to environmental conditions or stress. Although uORF- and uoORF-mediated translational regulation is independent of the sequence of the translated product, rare uORFs and uoORFs are conserved at the amino acid level, suggesting that the microprotein products of at least a subset of these smORFs may be stable and functional in *trans*. One example of a conserved, functional uORF is ASDURF (103). The percentage of uORFs and uoORFs that encode functional microproteins in addition to regulating downstream CDS translation is an important, currently unanswered, question.

Of emerging interest are nested or internal ORFs (iORFs) representing frameshifted sequences fully contained within an annotated protein CDS. Importantly, iORFs are encoded in alternative reading frames relative to the CDS [similar to uoORFs and doORFs (downstream overlapping ORFs), the latter of which are described below], and thus the amino acid sequence of the encoded microprotein is completely different from the annotated protein it overlaps. If the iORF was in the same reading frame as the CDS, it would represent an internal initiation site and would terminate at the same stop codon as the CDS. Although ribosome profiling has enabled discovery of internal, in-frame start codons with biological significance (104), they are not further considered in this article.

Out-of-frame iORFs have been detected with proteomics, where they can represent up to 30% of identified noncanonical ORFs (15, 22) but may be undercounted by both shotgun proteomics and ribosome profiling. iORFs are also challenging to characterize using genetic tools: That iORFs completely overlap a canonical protein CDS in genomic space means genetic tools to perturb, knock down, or knock out iORFs cannot a priori separate the function of a putative iORF from the CDS it overlaps. Only a handful of iORF-encoded microproteins, such as MINAS-60 and alt-FUS (74, 105), have been studied at the molecular level to date. This class of smORFs may therefore merit increased attention in the future, but improved tools are needed for their identification and characterization.

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

Downstream ORFs (dORFs) are entirely contained within the 3′ UTR; doORFs that initiate within the CDS in an alternative reading frame and extend into the 3′ UTR have also been reported, although both dORFs and doORFs are typically detected in lesser numbers than uORFs/uoORFs (7, 15). Interestingly, quantitative ribosome profiling suggests that dORF translation is positively correlated with CDS translation, suggesting that dORFs could play a regulatory role (106)—or, alternatively, that ribosomes in the 3′ UTR occasionally remain associated with highly translated transcripts and require "rescue" (107). Accordingly, few, if any, dORF- or doORF-encoded microproteins have been characterized to date, and whether the roles of these classes of smORFs go beyond *cis* translational regulation remains to be addressed.

Mitochondrial DNA smORFs

Finally, a class of microproteins including humanin, the short humanin-like peptides (SHLPs), SHMOOSE, and MTALTND4 has been identified that maps to the mitochondrial genome (**Figure 4**) (108, 109). Interestingly, several of the smORFs encoding these microproteins overlap the noncoding mitochondrial ribosomal RNAs and mitochondrial transfer RNA genes, and their expression mechanisms are currently undefined. Surprisingly, mitochondrial DNA–encoded microproteins not only participate in energy metabolism but also regulate diverse aspects of physiology, including inflammation and neurodegeneration, and several are secreted, again by an undefined mechanism (108, 109). The molecular rationale for encoding secreted microproteins in the mitochondrial genome is currently unknown.

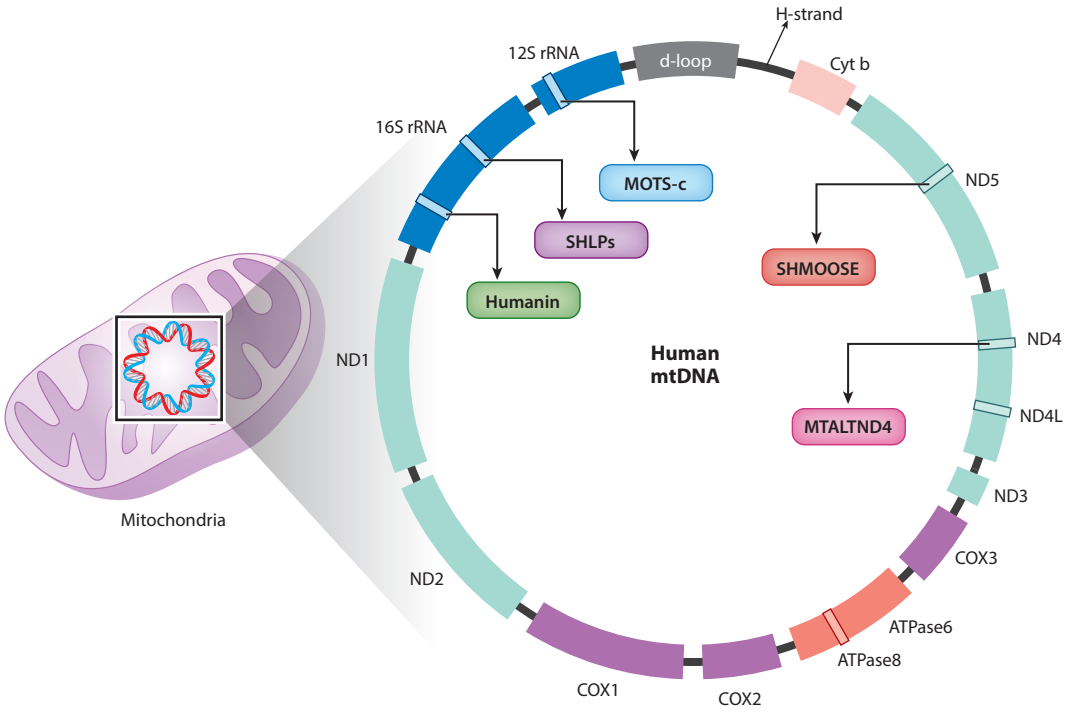

Figure 4

Location of mitochondrial-derived peptides in the human mitochondrial genome. Humanin and SHLPs (SHLP1–SHLP6) are found in *16S rRNA*, MOTS-c in *12S rRNA*, SHMOOSE overlapping with the protein coding gene *ND5*, and MTALTND4 within the protein coding gene *ND4*. Abbreviations: mtDNA, mitochondrial DNA; rRNA, ribosomal RNA; SHLP, short humanin-like peptide.

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

FUNCTIONS OF EUKARYOTIC MICROPROTEINS

Having discussed the history, nomenclature, discovery methods, and loci encoding eukaryotic microproteins, we now consider the functions of microproteins in various eukaryotic organisms, from yeast to human. We discuss individual examples of broader significance, as well as the impacts different model organisms have had on our understanding of smORF-encoded microproteins more broadly. We discuss individual microproteins of biological significance in the context of the organism in which they were discovered, but many are evolutionarily conserved. We focus in this section on microproteins identified in mammals; for discussion of microproteins characterized in other eukaryotes, see the **Supplemental Text**.

Supplemental Material >

MAMMALIAN MICROPROTEINS

As the biological relevance of mammalian microproteins in physiology and disease becomes clearer, it is now possible to define their roles in the context of biological processes from immunity to metabolism. We also consider shared mechanisms of action of mammalian microproteins such as allosteric control and regulation of macromolecular complexes (**Figure 5**). Of course, there are important microprotein studies we are unable to include, and we refer the reader to recent reviews that offer additional examples (110–112). In addition, we do not consider noncanonical proteins above our defined size limit of 150 aa.

Metabolism

Several translated smORFs, such as humanin, have emerged as important players in metabolic homeostasis (10, 113) (**Table 1**). As a mitochondrial-derived peptide (MDP)—a microprotein encoded in the mitochondrial DNA—humanin confers cytoprotection against Alzheimer's disease. The molecular mechanism of humanin is not fully understood, and multiple targets have been proposed (114), including G protein–coupled formylpeptide receptor-like-1 (115), insulin-like growth factor–binding protein 3 (116), and the trimeric complex of ciliary neurotrophic factor receptor alpha (CNTFR), IL-27 receptor subunit WSX-1, and glycoprotein 130 (113). The physiological effects of humanin are thought to extend beyond neuroprotection to other aging-related diseases, including type 2 diabetes mellitus via activation of signaling pathways such as MEK (MAPK/ERK kinase)/ERK (extracellular signal-regulated kinase), PI3K (phosphatidylinositol 3-kinase)/AKT (protein kinase B), and JAK (Janus Kinase)/STAT (Signal Transducer and Activator of Transcription) (117). Humanin has been implicated in improving insulin sensitivity, particularly in the context of polycystic ovary syndrome, and increasing health span (118–120). A class of MDPs closely related in sequence to humanin is composed of six SHLPs (121). The most well-studied of these is SHLP2, which has been proposed to be cytoprotective in age-related macular degeneration and Parkinson's disease, as well as to regulate energy homeostasis and insulin sensitivity, much like humanin (121–123). MOTS-c is yet another MDP with functions in improving insulin resistance and glucose metabolism, specifically in skeletal and heart muscle tissue (124). Finally, the MDP known as MOXI or mitoregulin is required for very long chain fatty acid oxidation (29, 125, 126). The MDPs therefore broadly play roles in metabolism, insulin resistance, and aging, not only suggesting new therapeutic avenues for neurodegeneration and diabetes but also intriguingly linking their mitochondrial DNA origin to energy homeostasis.

Microproteins encoded in both the mitochondrial and nuclear genomes can regulate mitochondrial bioenergetics directly and indirectly. For example, MTLAND4 is a secreted MDP that represses mitochondrial respiration (127). In contrast, mutations in the MDP SHMOOSE, like humanin, are associated with Alzheimer's disease, and the SHMOOSE microprotein localizes to mitochondria and promotes aerobic respiration (128). In an example with strong mechanistic

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

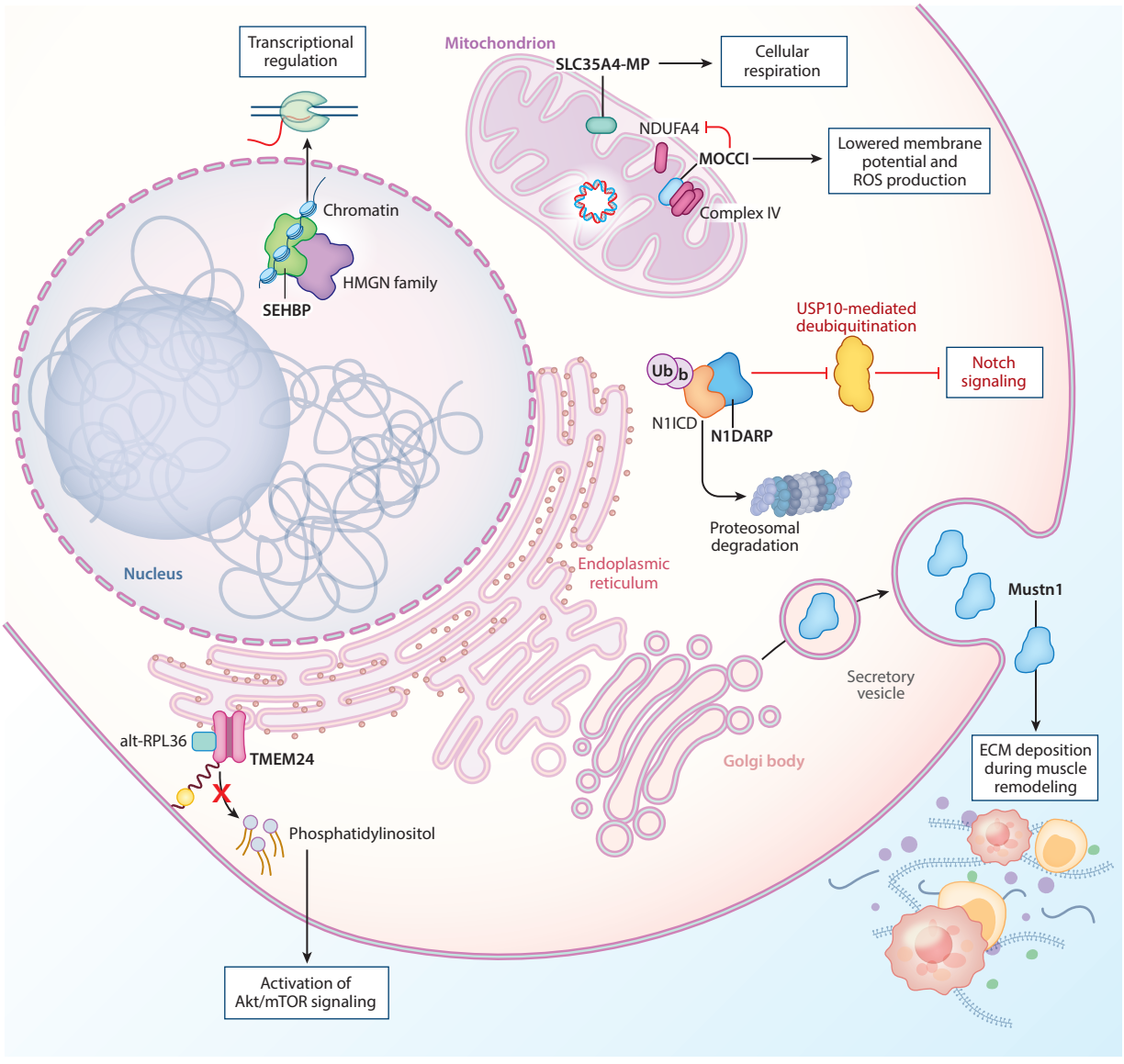

Figure 5

Functional microproteins in eukaryotic cells. Shown are selected examples of microproteins that function in different cellular compartments and processes. Abbreviations: ECM, extracellular matrix; ROS, reactive oxygen species; Ub, ubiquitin.

support, the BRAWNIN microprotein acts in concert with the canonical small proteins UQCC3, UQCC4, and SMIM4 to promote assembly and function of respiratory chain complex III (129), as does mitolamban (130). Interestingly, uORF-encoded microproteins have also been associated with mitochondrial energetics: The inner membrane–localized SLC35A4-MP promotes mitochondrial respiration by an unknown mechanism (131); the glioblastoma suppressor MP31 encoded upstream of PTEN inhibits lactate conversion to pyruvate (132); and the MIEF1

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

Table 1 Microproteins involved in metabolism

Microprotein	Function	Type of smORF	Reference(s)
Humanin	Provides neuroprotection as well as improved insulin sensitivity and glucose consumption	Mitochondrial DNA	10, 113–120
SHLP2	Cytoprotective insulin sensitizer that promotes mitochondrial biogenesis	Mitochondrial DNA	121–123
MOTS-c	Enhances glucose uptake, insulin sensitivity, and fatty acid oxidation by activating the AMPK pathway	Mitochondrial DNA	124
MOXI/mitoregulin	Promotes beta oxidation, lipid homeostasis, and respiratory chain complex I function	lncRNA	29, 125, 126
MTALTND4	Decreases mitochondrial respiration rate	Mitochondrial DNA	127
SHMOOSE	Interacts with mitofilin at the inner mitochondrial membrane and enhances aerobic respiration	Mitochondrial DNA	128
BRAWNIN	Necessary for the proper assembly of respiratory chain complex III	*C12ORF73*	129
Mitolamban	Positive regulator of respiratory chain complex III	lncRNA	130
SLC35A4-MP	Necessary for cellular respiration	uORF	131
MP31	Disrupts lactate-pyruvate conversion and lysosome function by binding lactate dehydrogenase B	uORF	132
MIEF1-MP (MIDUO)	Binds to mitoribosome to promote mitochondrial translation	uORF	133

Abbreviations: lncRNA, long noncoding RNA; ORF, open reading frame; smORF, small ORF; uORF, upstream ORF.

upstream microprotein (now annotated as MIDUO) promotes mitochondrial function, likely via interaction with the mitoribosome (133). Interestingly, mitochondrial microproteins can also have tissue-specific functions beyond metabolism, such as the murine Kastor and Polluks microproteins expressed from a bicistronic mRNA. These outer mitochondrial membrane–localized proteins are required for mitochondrial condensation during spermatogenesis (134). Overall, microproteins both derived from and localizing to the mitochondrion have wide-ranging effects on human metabolism and aging-related diseases.

Cellular Transport

Transporters and pumps compose another class of proteins that can interact with microproteins, with several examples in bacteria (135, 136) as well as in eukaryotes (**Table 2**). We consider two eukaryotic microproteins that interact with transporters and regulate movement of their substrates across membranes. First, the 14-kDa alt-RPL36 protein, which overlaps the CDS of human

Table 2 Microproteins involved in cellular transport

Microprotein	Function	Type of smORF	Reference
Alt-RPL36	Phosphorylation-dependent inhibition of phospholipid transport across ER-plasma membrane contact points by TMEM24	Overlapping ORF	78
NEMEP	Associates with GLUT1 and GLUT3 to promote cellular glucose uptake during mesendoderm differentiation	lncRNA	139

Abbreviations: ER, endoplasmic reticulum; lncRNA, long noncoding RNA; smORF, small open reading frame.

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

ribosomal protein L36 in an alternative reading frame, regulates TMEM24 (78). TMEM24 transports phospholipids across ER-plasma membrane contact sites in response to calcium signaling (137, 138). The lipid substrates of TMEM24, including phosphatidylinositol among others, once in the plasma membrane, function in processes such as insulin secretion, neuronal signaling (137, 138), and Akt activation (78). Importantly, although the molecular details remain to be fully elucidated, phosphorylated alt-RPL36 represses TMEM24 to downregulate inositide levels and Akt signaling. Alt-RPL36 repression does not occur in the absence of its phosphorylation, suggesting a mechanism by which TMEM24 activity could be dynamically regulated.

NEMEP is a recently reported transmembrane microprotein previously annotated as a lncRNA in mouse (139). Its human homolog, encoded by *TMEM155*, is an annotated small protein (SMIM43) of unknown function. During mouse embryonic development, NEMEP interacts with glucose transporters GLUT1 and GLUT3 to promote glucose uptake, a process required for mesendoderm formation.

Muscle and Heart Function

Muscle and heart contractility is a Ca^{2+} signaling–dependent process that relies on the activity of sarcoplasmic reticulum (SR)/ER Ca^{2+} ATPase (SERCA), a pump that sequesters Ca^{2+} in the sarcoplasmic reticulum to terminate contraction (**Table 3**). SERCA isoforms are expressed in different tissues and are precisely regulated in space and time by a suite of microprotein effectors, the regulins (140). Phospholamban and sarcolipin have long been known as SERCA inhibitors in heart and skeletal muscle, respectively, whose repression is alleviated by their phosphorylation. Surprisingly, lncRNAs specifically expressed in cardiac and skeletal muscle encode additional, critical SERCA regulatory microproteins, which have been termed micropeptides in the literature. The first discovered novel regulin was myoregulin, which is expressed in all skeletal muscle, where it regulates SERCA1 (26). Genetic deletion of myoregulin improves muscle performance by improving Ca^{2+} cycling. Subsequently, endoregulin and another-regulin were identified and found to inhibit and colocalize with nonmuscle SERCA isoforms SERCA2a in cardiac tissue and SERCA3 in other epithelial cells, respectively (27). Finally, in contrast to the tissue-specific inhibitory regulins, DWORF is a muscle-expressed, lncRNA-encoded micropeptide activator of SERCA, which may act by displacing phospholamban to alleviate its inhibition, although DWORF has also been

Table 3 Microproteins involved in muscle and heart function

Microprotein	Function	Type of smORF	Reference(s)
Myoregulin	Regulates muscle contraction by inhibiting SERCA1 in skeletal muscles	lncRNA	26
Endoregulin	SERCA3 inhibitor in endothelial and epithelial cells	lncRNA	27
Another-regulin	SERCA2b inhibitor in nonmuscle cells	lncRNA	27
DWORF	Activator of all SERCA isoforms, primarily targeting SERCA2a in cardiac muscles	lncRNA	30, 141, 142
SPAR	Binds to vacuolar ATPase and deactivates the amino acid–stimulated mTORC1 pathway, thereby hindering skeletal muscle regeneration	lncRNA	143
Myomixer/myomerger	Fusogen combining myoblast membranes during muscle formation	Uncharacterized protein	144
Mustn1	Modulates changes to extracellular matrix in muscles after exercise or injury	Uncharacterized protein	145

Abbreviations: ER, endoplasmic reticulum; lncRNA, long noncoding RNA; SERCA, sarcoplasmic reticulum/ER Ca^{2+} ATPase; smORF, small open reading frame.

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

reported to exhibit direct activation activity (30, 141). DWORF overexpression has been reported to alleviate dilated cardiomyopathy (141, 142). As discussed in the **Supplemental Text**, distantly related regulins have also been identified in invertebrates, suggesting that this class of micropeptides is broadly conserved.

Supplemental Material >

Microproteins have also been reported to regulate muscle development and regeneration. A key molecular cascade in muscle regeneration, mTORC1 signaling, is downregulated by the lncRNA-encoded SPAR microprotein (143). Highly expressed in skeletal muscles, SPAR interacts with vacuolar ATPase in lysosomal membranes and keeps it strongly bound to the Ragulator complex and Rags, thus inactivating mTORC1 even in the presence of amino acids. The increase in myogenic gene expression downstream of increased mTORC1 activation in SPAR-deficient mice likely explains the faster healing observed postinjury. These same myogenic genes are also required for skeletal muscle formation during development, which itself relies on a microprotein named myomerger or myomixer (144). This microprotein is embedded in the plasma membrane of myoblasts, enabling them to fuse into myofibers making up muscle tissue. Finally, Mustn1 is a microprotein involved in muscle extracellular matrix remodeling (145).

Gene Expression and Genome Maintenance

Numerous mammalian microproteins have been reported to play roles in gene expression both inside and outside the nucleus (**Table 4**). First, we consider microproteins associated with chromatin. An example is SEHBP, which was identified in association with histone H2B and dysregulates transcription globally when deleted (146). A microprotein encoded in a circRNA, LINC-PINT, has also been reported to promote transcriptional termination via interaction with PAF1c (147). Microproteins are also involved in DNA repair. In particular, isoforms of the CYREN (formerly MRI) microprotein interact with the Ku heterodimer to promote association of nonhomologous end joining (NHEJ) factors to promote DNA double-strand break repair during G1 phase of the cell cycle (148, 149). CYREN is also required to prevent aberrant fusion of telomeres by NHEJ during S and G2 (150). More recently, the lncRNA-encoded DDUP microprotein was found to

Table 4 Microproteins involved in gene expression and genome maintenance

Microprotein	Function	Type of smORF	Reference(s)
SEHBP	Histone-binding transcription factor	uORF	146
PINT87aa	Inhibits cell proliferation and tumor growth in glioblastoma by inhibiting PAF1c-mediated oncogenic gene translation	circRNA	147
CYREN-1	Induces DNA double-strand break repair via classical nonhomologous end joining	Predicted protein	150
DDUP	Promotes DNA repair through involvement in RAD51C-mediated homologous recombination and PCNA-mediated postreplication repair	lncRNA	151, 152
Alt-LAMA3	Promotes preribosomal RNA transcription	uORF	73
MINAS-60	Negative regulator of pre-60S ribosomal subunit assembly and/or cytoplasmic export	iORF	74
APPLE	Positively regulates eIF4F translation initiation complex assembly by binding PABC1 in acute monocytic leukemia cells	Antisense lncRNA	153
NBDY	Globally regulates mRNA decapping and decay	lncRNA	38, 77, 93, 154

Abbreviations: circRNA, circular RNA; iORF, internal open reading frame; lncRNA, long noncoding RNA; mRNA, messenger RNA; smORF, small ORF; uORF, upstream ORF.

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

be upregulated by DNA damage, whereupon it retains RAD18 at sites of DNA damage to promote both homology-directed and postreplication repair (151). DDUP expression also correlates with cisplatin resistance (152), suggesting that DNA repair–promoting microproteins could be therapeutically relevant.

Multiple nucleolar microproteins have been reported to regulate ribosome biogenesis as well as translation by mature ribosomes. For example, C11ORF98 interacts with nucleolar proteins nucleophosmin and nucleolin, although its function is otherwise currently unknown (39). The 148-aa, uORF-encoded alt-LAMA3 protein localizes to the nucleolus and interacts with the PeBoW complex to promote ribosomal RNA transcription (73). MINAS-60, which is encoded in an iORF that overlaps the *RBM10* gene, also localizes to the nucleolus and downregulates large ribosomal subunit export into the cytoplasm (74). There are thus multiple microproteins localized to the nucleolus in mammalian cells, and it seems reasonable to speculate that additional microproteins may localize to this membraneless organelle to regulate the complex pathway of ribosome biogenesis. Of course, the job of mature ribosomes is to translate mRNA, and the oncomicroprotein APPLE, which is encoded in an antisense lncRNA, has been reported to regulate this process, influencing translation initiation by promoting the interaction between PACB1 and eIF4G (153). Microproteins thus affect protein synthesis by regulating both ribosome biogenesis and translation.

Finally, gene expression depends on not only transcription but also RNA stability and degradation. The NBDY microprotein directly regulates mRNA decay by binding to the cytoplasmic mRNA decapping complex via the coactivator proteins DCP1A/B and EDC4 (38). Deletion of NBDY broadly dysregulates mRNA half-lives in human cells (93). A substantial fraction of mRNA decapping enzyme (DCP2) substrates are stabilized in NBDY knockout cells, including mRNAs encoding factors involved in inflammation and immunity, suggesting that NBDY is required for their decapping and subsequent degradation (93). However, a smaller subset of DCP2 substrates exhibit dramatic destabilization in cells lacking NBDY (93, 154). Phosphorylation of NBDY promotes dissociation of membraneless organelles termed P-bodies that are associated with the mRNA decapping complex (77); this phenomenon may contribute to its regulation of specific DCP2 substrates, but the mechanistic link between P-bodies and mRNA decapping remains controversial. The example of NBDY demonstrates how a microprotein can regulate protein localization and association to control mRNA stability.

Proteostasis

Proteostasis is the maintenance of proteome integrity via the balance of protein synthesis, quality control, and degradation, and microproteins are associated with this process at multiple levels (**Table 5**). First, some noncanonical translation products, particularly hydrophobic peptides expressed from noncoding genomic regions, can be targeted for proteasomal degradation by the

Table 5 Microproteins involved in proteostasis

Microprotein	Function	Type of smORF	Reference(s)
pTINCR	Induces epithelial differentiation and suppresses tumor growth in squamous cell carcinoma	lncRNA	157, 158
ASDURF	Core component of the prefoldin-like module of the PAQosome chaperone complex	uORF	49, 103
PIGBOS	Regulates the UPR at ER-mitochondria junctions	Antisense lncRNA	97

Abbreviations: ER, endoplasmic reticulum; lncRNA, long noncoding RNA; PIGBOS, PIGB opposite strand; smORF, small open reading frame; uORF, upstream ORF; UPR, unfolded protein response.

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

GET complex (81). Multiple ubiquitin-like microproteins have also been reported (155, 156) and have recently been reviewed (63). For example, the pTINCR ubiquitin-like microprotein has been described by several groups as a tumor suppressor in squamous cell carcinoma (157, 158), highlighting the disease relevance of this class of microproteins. Microproteins are thus regulated by, and also may contribute to, protein degradation.

Microproteins are involved in chaperones and the unfolded protein response (UPR) as well. Notably, the ASDURF microprotein has been identified as the previously cryptic sixth subunit of the prefoldin-like subunit of the PAQosome (103), which is required for assembly of large, complex molecular machines (159). ASDURF is upregulated in childhood medulloblastoma (49). In another example, the PIGBOS microprotein has been reported to function in ER-mitochondria junctions to regulate the UPR (97). PIGBOS is encoded in a lncRNA antisense to the *PIGB* gene. PIGBOS interacts with the CLCC1 protein, and this interaction is required for regulation of the UPR. In its absence, cells are sensitized to ER stress and apoptosis.

Immunity and Inflammation

Multiple microproteins have been implicated in the immune system, especially in regulation of inflammation and innate immunity (**Table 6**). For example, the rodent lncRNA *Aw112010* encodes a secreted (83) microprotein that is translationally upregulated in M1 polarized macrophages and is essential for innate immune responses against bacterial pathogens (21). In contrast, the Mm47 microprotein is downregulated in lipopolysaccharide-treated myeloid cells but is required for Nlrp3 inflammasome activation (160). More recently, the lncRNA-encoded, ER-localized MAVI1 microprotein was found to interact with mitochondrial MAVS to inhibit type I interferon signaling (161).

Two miPEPs are involved in innate and adaptive immune regulation. One of these is MOCCI (162), which, interestingly, is a miPEP whose transcript can be processed to generate miR-147b, which regulates RIG-I antiviral defense and downregulates the *NDUFA4* mRNA. The latter is consequential because the MOCCI microprotein replaces NDUFA4 in respiratory complex IV during inflammation to lower mitochondrial membrane potential and reduces reactive oxygen species production, providing cytoprotection and reducing interferon responses in endothelial cells. *MOCCI* thus provides multilevel regulation of mitochondrial activity, inflammatory signaling, and antiviral gene expression. Another miPEP is encoded in the lncRNA *MIR155HG*, upstream of pri–miR-155 (163). The micropeptide, termed P155, inhibits antigen presentation on dendritic cells by MHC II by disrupting chaperones. Although P155 would thus be detrimental

Table 6 Microproteins involved in immunity

Microprotein	Function	Type of smORF	Reference(s)
Aw112010	Protects against mucosal infection through the regulation of proinflammatory pathways	lncRNA	21
Mm47	Activator of Nlrp3 inflammasome	lncRNA	160
MAVI1	Downregulates innate immune response by binding MAVS and inactivating type I interferon signaling	lncRNA	161
MOCCI	Generates miR-147b, which regulates RIG-I antiviral defense and downregulates the *NDUFA4* mRNA	pri-miRNA	162
miPEP155	Inhibits antigen presentation on dendritic cells by MHC II	pri-miRNA	163

Abbreviations: lncRNA, long noncoding RNA; MHC II, major histocompatibility complex class II; mRNA, messenger RNA; pri-miRNA, primary microRNA; smORF, small open reading frame.

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

Table 7 Microproteins associated with cancer and cell cycle regulation

Microprotein	Function	Type of smORF	Reference(s)
MIAC	Inhibits tumor growth and metastasis in head and neck squamous cell carcinoma and renal cell carcinoma	lncRNA	167, 168
miPEP133	Antiproliferative and anti-invasive in nasopharyngeal carcinoma and ovarian cancer	pri-miRNA	169
YY1BM	Promotes apoptosis due to nutrient deficiency in esophageal squamous cell carcinoma by binding YY1	lncRNA	170
N1DARP	Disrupts Notch1 signaling pathways in pancreatic cancer by inducing N1ICD proteasomal degradation	lncRNA	171
CIP2A-BP	Inhibits PI3K/Akt/NFκB pathways by competitively binding CIP2A in triple negative breast cancer to prevent metastasis	lncRNA	172
SMIM26	Downregulates tumor progression and metastasis in clear cell renal cell carcinoma by inactivating AKT signaling	lncRNA	173
SMIM30	Induces G1/S transition by decreasing Ca^{2+} levels in the cytoplasm and activating SERCA	lncRNA	174, 175
TRPC5OS	Promotes breast tumor development by binding ENO1 and enhancing glucose uptake	Antisense lncRNA	176
ASAP	Enhances ATP synthase activity in mitochondrial respiration, thereby promoting colorectal cancer growth	lncRNA	177
PACMP	Regulates DNA damage response by promoting PARP1-dependent PARylation and blocking CtIP proteasomal degradation	lncRNA	178

Abbreviations: lncRNA, long noncoding RNA; mRNA, messenger RNA; pri-miRNA, primary microRNA; SERCA, sarcoplasmic reticulum/endoplasmic reticulum Ca^{2+} ATPase; smORF, small open reading frame.

to adaptive immunity, treatment with synthetic P155 was beneficial in animal models of autoinflammation (163). Overall, microproteins play important roles in immunity and in the future may inform therapeutic approaches to immunomodulation.

Cancer and Cell Cycle Regulation

Perhaps the greatest advances in our understanding of microproteins in human disease pertain to cancer (164) (**Table 7**). This topic has been extensively and recently reviewed (112, 165, 166). In this section we highlight several key additional examples of cancer-associated microproteins as well as microproteins that regulate the cell cycle.

In addition to MP31, PINT87aa, and pTINCR, numerous tumor suppressor microproteins have been reported that regulate multiple aspects of cancer cell biology. lncRNA-encoded MIAC has been reported not only to repress cell growth and migration in head and neck squamous cell carcinoma (167) but also to cause enhanced apoptosis in renal cell carcinoma (168); both phenotypes are proposed to occur through the interaction of MIAC with aquaporin 2 to inhibit actin cytoskeleton processes—a link that may require further mechanistic validation. Similarly, miPEP133—yet another disease-associated human pri-miRNA-encoded microprotein—has been

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

shown to harbor tumor suppressor activity because it increases apoptosis downstream of *TP53* transcriptional activation in nasopharyngeal carcinoma cells (169). Esophageal squamous cell carcinoma has been reported to be more invasive during nutrient deprivation by downregulating the lncRNA-derived polypeptide YYB1M (170), which blocks the YY1–androgen receptor interaction during nutrient deprivation, thereby inhibiting eEF2K signaling and promoting apoptosis. Multiple additional tumor suppressor microproteins, such as N1DARP (171), CIP2A-BP (172), and SMIM26 (173), have been reported, suggesting that anticancer microproteins may be numerous.

Like ASDURF and APPLE, a number of oncogenic microproteins have also been reported. SMIM30 originates from an lncRNA, and its oncogenic mechanisms are particularly well-studied. SMIM30 is upregulated in many cancers, where it has been reported to enhance the G1/S transition and thus proliferation (174). Interestingly, SMIM30 is anchored in the ER membrane where it decreases cytosolic Ca^{2+} (174). SMIM30 was further found to promote hepatocellular carcinoma progression by upregulating MAPK signaling via the membrane localization of SRC/YES1 (175). The TRPC5OS microprotein, which is encoded in a testis-specific antisense lncRNA, has been reported to be aberrantly expressed in breast cancer cells, where it promotes glucose uptake and cell proliferation (176). Several other recently reported lncRNA-encoded microproteins such as ASAP (177) and PACMP (178), among others, have also been reported to promote tumor growth and/or survival. Overall, microproteins can influence cancer cell signaling and metabolic pathways, conferring growth advantage or inhibition on tumor cells. These findings place smORFs at the forefront of advances in cancer diagnosis and treatment.

As discussed above, microproteins are overrepresented relative to canonical proteins in the MHC I immunopeptidome and have been directly detected in the immunopeptidomes of melanoma (75) and hepatocellular carcinoma (76). Although these HLA-displayed microprotein fragments may not directly affect cancer cell biology, they are of outsize importance to immunotherapy and have been proposed to inform the design of tumor-specific anticancer peptide vaccines (75, 76).

CONCLUSION AND FUTURE OUTLOOK

Throughout this article, we have highlighted the ways in which technical innovations in microprotein discovery have illuminated new functional biomolecules involved in signaling and regulation in eukaryotic cells. However, major questions remain about how many translated microproteins are indeed functional. Indeed, the defective ribosomal products (DRiPs) hypothesis posits that a substantial fraction of newly synthesized proteins, perhaps up to 30%, are quickly degraded and presented on MHC I; as discussed above, triage of some noncoding microproteins by GET suggests that they may represent such ephemeral DRiPs (179). We must now elucidate which smORFs give rise to stable, functional cellular gene products and which function primarily in immune self-recognition. Additionally, it is currently unclear to what extent unstable microproteins can be transiently stabilized to regulate cellular processes or to attain stability and expression optimization over evolutionary timescales, as has been suggested for yeast de novo proteins (59). Addressing these questions will be critical to understanding the roles of smORF translation in cells, immunity, and de novo protein evolution. Building more on the concept of smORFs as evolutionarily transient genes, it is of current interest to examine how the evolutionary plasticity of smORFs plays a role in speciation and adaptation (59).

Microproteins have also revealed surprises about the sequences and structures of eukaryotic genes. Human genes have generally been assumed to encode only a single CDS without overlaps, a rule informed by the scanning model of eukaryotic translation initiation (180). The observation

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

that human mRNAs can dually encode a canonical protein and a microprotein is therefore surprising, and determining the mechanisms by which these smORFs can be translated is critical. Just as conceptually significant is the molecular rationale for dual coding, given that eukaryotic genomes are not apparently under pressure to colocalize related genes, as in bacterial operons, or minimize size via overlaps, as in viral genomes.

Current efforts to reanalyze and standardize microprotein detection across datasets have now yielded a reference human translatome (14). What, then, does the future of microprotein discovery and technology development look like? There will always be more microproteins to identify that are expressed in cell, developmental, and other stimulus-dependent conditions, and that work is likely best accomplished with ribosome profiling, which affords the greatest sensitivity for identifying large numbers of translated microproteins, followed by proteomic, genetic and molecular validation. In the same vein, identification of microproteins as disease biomarkers is an important ongoing effort that can be accomplished with quantitative ribosome profiling or proteomics. At the same time, questions about microprotein functions can be addressed with advanced proteomic and interactomic methods, including enrichment and identification of microprotein posttranslational modifications, localization, and binding partners.

Finally, although CRISPR screens suggest that not all translated microproteins are essential for cell viability or proliferation, many could have regulatory or modulatory functions with relevance in fine-tuning cellular processes, in responding to environmental cues, or when overexpressed. High-throughput assays coupled with novel bioassays could reveal these hypothetical functions at scale. These studies have the potential to provide insights into cellular pathways and to inform novel therapeutic and diagnostic strategies for disease.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was funded in part by the National Institutes of Health (R01 GM155404), The Mark Foundation for Cancer Research Emerging Leader Award, Paul G. Allen Frontiers Group Distinguished Investigator Award, and Sloan Research Fellowship (FG-2022-18417) to S.A.S. V.E. was supported in part by the Emerging Scholars Initiative Post-Baccalaureate Research Program and American Chemical Society Bridge Program of the Yale Provost's Office and the Yale Graduate School of Arts and Sciences. N.J. was supported in part by a Roberts Fellowship from the Yale Department of Chemistry.

LITERATURE CITED

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921
2. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420(6915):563–73
3. Basrai MA, Hieter P, Boeke JD. 1997. Small open reading frames: beautiful needles in the haystack. *Genome Res*. 7(8):768–71
4. Kozak M. 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*. 15(20):8125–48
5. Barbosa C, Peixeiro I, Romão L. 2013. Gene expression regulation by upstream open reading frames and human disease. *PLOS Genet*. 9(8):e1003529

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

6. Oyama M, Itagaki C, Hata H, Suzuki Y, Izumi T, et al. 2004. Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs. *Genome Res*. 14(10b):2048–52
7. Oyama M, Kozuka-Hata H, Suzuki Y, Semba K, Yamamoto T, Sugano S. 2007. Diversity of translation start sites may define increased complexity of the human short ORFeome. *Mol. Cell. Proteom*. 6(6):1000–6
8. Kastenmayer JP, Ni L, Chu A, Kitchen LE, Au W-C, et al. 2006. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res*. 16(3):365–73
9. Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. 2007. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLOS Biol*. 5(5):e106
10. Hashimoto Y, Ito Y, Niikura T, Shao Z, Hata M, et al. 2001. Mechanisms of neuroprotection by a novel rescue factor humanin from Swedish mutant amyloid precursor protein. *Biochem. Biophys. Res. Commun.* 283(2):460–68
11. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y. 2007. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.* 9(6):660–65
12. Hemm MR, Weaver J, Storz G. 2020. *Escherichia coli* small proteome. *EcoSal Plus* 9(1). **https://doi.org/10.1128/ecosalplus.esp-0031-2019**
13. Storz G, Wolf YI, Ramamurthi KS. 2014. Small proteins can no longer be ignored. *Annu. Rev. Biochem.* 83:753–77
14. Mudge JM, Ruiz-Orera J, Prensner JR, Brunet MA, Calvet F, et al. 2022. Standardized annotation of translated open reading frames. *Nat. Biotechnol.* 40(7):994–99
15. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, et al. 2013. Peptidomic discovery of short open reading frame–encoded peptides in human cells. *Nat. Chem. Biol.* 9(1):59–64
16. Prensner JR, Enache OM, Luria V, Krug K, Clauser KR, et al. 2021. Noncanonical open reading frames encode functional proteins essential for cancer cell survival. *Nat. Biotechnol.* 39(6):697–704
17. Khitun A, Ness TJ, Slavoff SA. 2019. Small open reading frames and cellular stress responses. *Mol. Omics* 15(2):108–16
18. Khitun A, Slavoff SA. 2019. Proteomic detection and validation of translated small open reading frames. *Curr. Protoc. Chem. Biol.* 11(4):e77
19. Olexiouk V, Crappe J, Verbruggen S, Verhegen K, Martens L, Menschaert G. 2016. sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res*. 44(D1):D324–29
20. Chen J, Brunner AD, Cogan JZ, Nunez JK, Fields AP, et al. 2020. Pervasive functional translation of noncanonical human open reading frames. *Science* 367(6482):1140–46
21. Jackson R, Kroehling L, Khitun A, Bailis W, Jarret A, et al. 2018. The translation of non-canonical open reading frames controls mucosal immunity. *Nature* 564(7736):434–38
22. Vanderperre B, Lucier JF, Bissonnette C, Motard J, Tremblay G, et al. 2013. Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLOS ONE* 8(8):e70698
23. Orr MW, Mao Y, Storz G, Qian SB. 2020. Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res*. 48(3):1029–42
24. Pueyo JI, Couso JP. 2011. Tarsal-less peptides control Notch signalling through the Shavenbaby transcription factor. *Dev. Biol.* 355(2):183–93
25. Magny EG, Pueyo JI, Pearl FMG, Cespedes MA, Niven JE, et al. 2013. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* 341(6150):1116–20
26. Anderson DM, Anderson KM, Chang C-L, Makarewich CA, Nelson BR, et al. 2015. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 160(4):595–606
27. Anderson DM, Makarewich CA, Anderson KM, Shelton JM, Bezprozvannaya S, et al. 2016. Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. *Sci. Signal.* 9(457):ra119
28. Makarewich CA, Olson EN. 2017. Mining for micropeptides. *Trends Cell Biol*. 27(9):685–96
29. Makarewich CA, Baskin KK, Munir AZ, Bezprozvannaya S, Sharma G, et al. 2018. MOXI is a mitochondrial micropeptide that enhances fatty acid β-oxidation. *Cell Rep*. 23(13):3701–9
30. Makarewich CA, Munir AZ, Schiattarella GG, Bezprozvannaya S, Raguimova ON, et al. 2018. The DWORF micropeptide enhances contractility and prevents heart failure in a mouse model of dilated cardiomyopathy. *eLife* 7:e38319

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

31. Cardon T, Salzet M, Franck J, Fournier I. 2019. Nuclei of HeLa cells interactomes unravel a network of ghost proteins involved in proteins translation. *Biochim. Biophys. Acta Gen. Subj.* 1863(10):1458–70
32. Cardon T, Fournier I, Salzet M. 2021. Shedding light on the ghost proteome. *Trends Biochem. Sci.* 46(3):239–50
33. Brunet MA, Leblanc S, Roucou X. 2020. Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. *Exp. Cell Res.* 393(1):112057
34. Gagnon M, Savard M, Jacques JF, Bkaily G, Geha S, et al. 2021. Potentiation of B2 receptor signaling by AltB2R, a newly identified alternative protein encoded in the human bradykinin B2 receptor gene. *J. Biol. Chem.* 296:100329
35. Staudt A, Wenkel S. 2011. Regulation of protein function by 'microProteins.' *EMBO Rep.* 12(1):35–42
36. Bhati KK, Kruusvee V, Straub D, Chandran AKN, Jung K-H, Wenkel S. 2020. Global analysis of cereal microproteins suggests diverse roles in crop development and environmental adaptation. *G3: Genes, Genomes, Genetics* 10(10):3709–17
37. Eguen T, Straub D, Graeff M, Wenkel S. 2015. MicroProteins: small size – big impact. *Trends Plant Sci.* 20(8):477–82
38. D'Lima NG, Ma J, Winkler L, Chu Q, Loh KH, et al. 2017. A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* 13(2):174–80
39. Chu Q, Rathore A, Diedrich JK, Donaldson CJ, Yates JR 3rd, Saghatelian A. 2017. Identification of microprotein–protein interactions via APEX tagging. *Biochemistry* 56(26):3299–306
40. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324(5924):218–23
41. Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147(4):789–802
42. Martinez TF, Chu Q, Donaldson C, Tan D, Shokhirev MN, Saghatelian A. 2020. Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* 16(4):458–68
43. van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, et al. 2019. The translational landscape of the human heart. *Cell* 178(1):242–60.e29
44. Erhard F, Halenius A, Zimmermann C, L'Hernault A, Kowalewski DJ, et al. 2018. Improved Ribo-seq enables identification of cryptic translation events. *Nat. Methods* 15(5):363–66
45. Ji Z, Song R, Regev A, Struhl K. 2015. Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* 4:e08890
46. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, et al. 2016. Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods* 13(2):165–70
47. Bazzini AA, Johnstone TG, Christiano R, MacKowiak SD, Obermayer B, et al. 2014. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* 33(9):981–93
48. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, et al. 2017. Translation from unconventional 5′ start sites drives tumour initiation. *Nature* 541(7638):494–99
49. Hofman DA, Ruiz-Orera J, Yannuzzi I, Murugesan R, Brown A, et al. 2024. Translation of non-canonical open reading frames as a cancer cell survival mechanism in childhood medulloblastoma. *Mol. Cell* 84(2):261–76.e18
50. Lee S, Liu B, Lee S, Huang S-X, Shen B, Qian S-B. 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *PNAS* 109(37):E2424–32
51. Schneider-Poetsch T, Ju J, Eyler DE, Dang Y, Bhat S, et al. 2010. Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat. Chem. Biol.* 6(3):209–17
52. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7(8):1534–50
53. Gao X, Wan J, Liu B, Ma M, Shen B, Qian S-B. 2015. Quantitative profiling of initiating ribosomes in vivo. *Nat. Methods* 12(2):147–53
54. Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, et al. 2012. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* 22(11):2208–18

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

55. Zhang P, He D, Xu Y, Hou J, Pan BF, et al. 2017. Genome-wide identification and differential analysis of translational initiation. *Nat. Commun.* 8(1):1749
56. Tong G, Hah N, Martinez TF. 2024. Comparison of software packages for detecting unannotated translated small open reading frames by Ribo-seq. *Brief. Bioinform.* 25(4):bbae268
57. Heyer EE, Moore MJ. 2016. Redefining the translational status of 80S monosomes. *Cell* 164(4):757–69
58. Biever A, Glock C, Tushev G, Ciirdaeva E, Dalmay T, et al. 2020. Monosomes actively translate synaptic mRNAs in neuronal processes. *Science* 367(6477):eaay4991
59. Wacholder A, Parikh SB, Coelho NC, Acar O, Houghton C, et al. 2023. A vast evolutionarily transient translatome contributes to phenotype and fitness. *Cell Syst.* 14(5):363–81.e8
60. Menschaert G, Van Criekinge W, Notelaers T, Koch A, Crappé J, et al. 2013. Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. *Mol. Cell. Proteom.* 12(7):1780–90
61. Koch A, Gawron D, Steyaert S, Ndah E, Crappé J, 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(23–24):2688–98
62. Zhang Z, Li Y, Yuan W, Wang Z, Wan C. 2022. Proteomics-driven identification of short open reading frame-encoded peptides. *Proteomics* 22(15–16):2100312
63. Mohsen JJ, Martel AA, Slavoff SA. 2023. Microproteins—discovery, structure, and function. *Proteomics* 23(23–24):2100211
64. Chen Y, Cao X, Loh KH, Slavoff SA. 2023. Chemical labeling and proteomics for characterization of unannotated small and alternative open reading frame-encoded polypeptides. *Biochem. Soc. Trans.* 51(3):1071–82
65. Cassidy L, Kaulich PT, Tholey A. 2019. Depletion of high-molecular-mass proteins for the identification of small proteins and short open reading frame encoded peptides in cellular proteomes. *J. Proteome Res.* 18(4):1725–34
66. Cassidy L, Prasse D, Linke D, Schmitz RA, Tholey A. 2016. Combination of bottom-up 2D-LC-MS and semi-top-down GelFree-LC-MS enhances coverage of proteome and low molecular weight short open reading frame encoded peptides of the archaeon *Methanosarcina mazei*. *J. Proteome Res.* 15(10):3773–83
67. Kaulich PT, Cassidy L, Weidenbach K, Schmitz RA, Tholey A. 2020. Complementarity of different SDS-PAGE gel staining methods for the identification of short open reading frame-encoded peptides. *Proteomics* 20(19–20):2000084
68. Kaulich PT, Cassidy L, Bartel J, Schmitz RA, Tholey A. 2021. Multi-protease approach for the improved identification and molecular characterization of small proteins and short open reading frame-encoded peptides. *J. Proteome Res.* 20(5):2895–903
69. Cassidy L, Helbig AO, Kaulich PT, Weidenbach K, Schmitz RA, Tholey A. 2021. Multidimensional separation schemes enhance the identification and molecular characterization of low molecular weight proteomes and short open reading frame-encoded peptides in top-down proteomics. *J. Proteom.* 230:103988
70. Ma J, Diedrich JK, Jungreis I, Donaldson C, Vaughan J, et al. 2016. Improved identification and analysis of small open reading frame encoded polypeptides. *Anal. Chem.* 88(7):3967–75
71. Ma J, Ward CC, Jungreis I, Slavoff SA, Schwaid AG, et al. 2014. Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *J. Proteome Res.* 13(3):1757–65
72. Yuan P, D'Lima NG, Slavoff SA. 2018. Comparative membrane proteomics reveals a nonannotated *E. coli* heat shock protein. *Biochemistry* 57(1):56–60
73. Na Z, Dai X, Zheng S-J, Bryant CJ, Loh KH, et al. 2022. Mapping subcellular localizations of unannotated microproteins and alternative proteins with MicroID. *Mol. Cell* 82(15):2900–11.e7
74. Cao X, Khitun A, Harold CM, Bryant CJ, Zheng SJ, et al. 2022. Nascent alt-protein chemoproteomics reveals a pre-60S assembly checkpoint inhibitor. *Nat. Chem. Biol.* 18(6):643–51
75. Ouspenskaia T, Law T, Clauser KR, Klaeger S, Sarkizova S, et al. 2022. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.* 40(2):209–17

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

76. Camarena ME, Theunissen P, Ruiz M, Ruiz-Orera J, Calvo-Serra B, et al. 2024. Microproteins encoded by noncanonical ORFs are a major source of tumor-specific antigens in a liver cancer patient meta-cohort. *Sci. Adv.* 10(28):eadn3628
77. Na Z, Luo Y, Cui DS, Khitun A, Smelyansky S, et al. 2021. Phosphorylation of a human microprotein promotes dissociation of biomolecular condensates. *J. Am. Chem. Soc.* 143(32):12675–87
78. Cao X, Khitun A, Luo Y, Na Z, Phoodokmai T, et al. 2021. Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM24. *Nat. Commun.* 12(1):508
79. Bogaert A, Fijalkowska D, Staes A, Van de Steene T, Demol H, Gevaert K. 2022. Limited evidence for protein products of noncoding transcripts in the HEK293T cellular cytosol. *Mol. Cell. Proteom.* 21(8):100264
80. Vakirlis N, Acar O, Hsu B, Castilho Coelho N, Van Oss SB, et al. 2020. De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences. *Nat. Commun.* 11(1):781
81. Kesner JS, Chen Z, Shi P, Aparicio AO, Murphy MR, et al. 2023. Noncoding translation mitigation. *Nature* 617(7960):395–402
82. Kovacs E, Tompa P, Liliom K, Kalmar L. 2010. Dual coding in alternative reading frames correlates with intrinsic protein disorder. *PNAS* 107(12):5429–34
83. Martinez TF, Lyons-Abbott S, Bookout AL, De Souza EV, Donaldson C, et al. 2023. Profiling mouse brown and white adipocytes to identify metabolically relevant small ORFs and functional microproteins. *Cell Metab.* 35(1):166–83.e11
84. Chen L, Zhang Y, Yang Y, Yang Y, Li H, et al. 2021. An integrated approach for discovering noncanonical MHC-I peptides encoded by small open reading frames. *J. Am. Soc. Mass. Spectrom.* 32(9):2346–57
85. Laumont CM, Daouda T, Laverdure JP, Bonneil E, Caron-Lizotte O, et al. 2016. Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nat. Commun.* 7:10238
86. Ruiz Cuevas MV, Hardy M-P, Hollý J, Bonneil É, Durette C, et al. 2021. Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Rep.* 34(10):108815
87. Durrant MG, Bhatt AS. 2021. Automated prediction and annotation of small open reading frames in microbial genomes. *Cell Host Microbe* 29(1):121–31.e4
88. Lin MF, Jungreis I, Kellis M. 2011. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* 27(13):i275–82
89. Couso JP, Patraquim P. 2017. Classification and function of small open reading frames. *Nat. Rev. Mol. Cell Biol.* 18(9):575–89
90. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. 2013. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154(1):240–51
91. Douka K, Birds I, Wang D, Kosteletos A, Clayton S, et al. 2021. Cytoplasmic long noncoding RNAs are differentially regulated and translated during human neuronal differentiation. *RNA* 27(9):1082–101
92. Xiao W, Halabi R, Lin C-H, Nazim M, Yeom K-H, Black DL. 2024. The lncRNA Malat1 is trafficked to the cytoplasm as a localized mRNA encoding a small peptide in neurons. *Genes Dev.* 38:294–307
93. Na Z, Luo Y, Schofield JA, Smelyansky S, Khitun A, et al. 2020. The NBDY microprotein regulates cellular RNA decapping. *Biochemistry* 59(42):4131–42
94. Lauressergues D, Ormancey M, Guillotin B, San Clemente H, Camborde L, et al. 2022. Characterization of plant microRNA-encoded peptides (miPEPs) reveals molecular mechanisms from the translation to activity and specificity. *Cell Rep.* 38(6):110339
95. Zheng X, Chen L, Zhou Y, Wang Q, Zheng Z, et al. 2019. A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling. *Mol. Cancer* 18(1):47
96. Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, et al. 2015. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* 163(3):583–93
97. Chu Q, Martinez TF, Novak SW, Donaldson CJ, Tan D, et al. 2019. Regulation of the ER stress response by a mitochondrial microprotein. *Nat. Commun.* 10(1):4883
98. An NA, Zhang J, Mo F, Luan X, Tian L, et al. 2023. De novo genes with an lncRNA origin encode unique human brain developmental functionality. *Nat. Ecol. Evol.* 7(2):264–78

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

99. Brunet MA, Levesque SA, Hunting DJ, Cohen AA, Roucou X. 2018. Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. *Genome Res.* 28(5):609–24
100. Calvo SE, Pagliarini DJ, Mootha VK. 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *PNAS* 106(18):7507–12
101. Hinnebusch AG. 2005. Translational regulation of *GCN4* and the general amino acid control of yeast. *Annu. Rev. Microbiol.* 59:407–50
102. Vattem KM, Wek RC. 2004. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *PNAS* 101(31):11269–74
103. Cloutier P, Poitras C, Faubert D, Bouchard A, Blanchette M, et al. 2020. Upstream ORF-encoded ASDURF is a novel prefoldin-like subunit of the PAQosome. *J. Proteome Res.* 19(1):18–27
104. Brubaker SW, Gauthier AE, Mills EW, Ingolia NT, Kagan JC. 2014. A bicistronic *MAVS* transcript highlights a class of truncated variants in antiviral immunity. *Cell* 156(4):800–11
105. Brunet MA, Jacques JF, Nassari S, Tyzack GE, McGoldrick P, et al. 2021. The *FUS* gene is dual-coding with both proteins contributing to *FUS*-mediated toxicity. *EMBO Rep.* 22(1):e50640
106. Wu Q, Wright M, Gogol MM, Bradford WD, Zhang N, Bazzini AA. 2020. Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J.* 39(17):e104763
107. Guydosh NR, Green R. 2014. Dom34 rescues ribosomes in 3′ untranslated regions. *Cell* 156(5):950–62
108. Kumagai H, Miller B, Kim SJ, Leelaprachakul N, Kikuchi N, et al. 2023. Novel insights into mitochondrial DNA: mitochondrial microproteins and mtDNA variants modulate athletic performance and age-related diseases. *Genes* 14(2):286
109. Miller B, Kim SJ, Kumagai H, Yen K, Cohen P. 2022. Mitochondria-derived peptides in aging and healthspan. *J. Clin. Investig.* 132(9):e158449
110. Wright BW, Yi Z, Weissman JS, Chen J. 2022. The dark proteome: translation from noncanonical open reading frames. *Trends Cell Biol.* 32(3):243–58
111. Posner Z, Yannuzzi I, Prensner JR. 2023. Shining a light on the dark proteome: non-canonical open reading frames and their encoded miniproteins as a new frontier in cancer biology. *Protein Sci.* 32(8):e4708
112. Hassel KR, Brito-Estrada O, Makarewich CA. 2023. Microproteins: overlooked regulators of physiology and disease. *iScience* 26(6):106781
113. Hashimoto Y, Kurita M, Aiso S, Nishimoto I, Matsuoka M. 2009. Humanin inhibits neuronal cell death by interacting with a cytokine receptor complex or complexes involving CNTF receptor α/WSX-1/gp130. *Mol. Biol. Cell* 20(12):2864–73
114. Lee C, Yen K, Cohen P. 2013. Humanin: a harbinger of mitochondrial-derived peptides? *Trends Endocrinol. Metab.* 24(5):222–28
115. Ying G, Iribarren P, Zhou Y, Gong W, Zhang N, et al. 2004. Humanin, a newly identified neuroprotective factor, uses the G protein-coupled formylpeptide receptor-like-1 as a functional receptor. *J. Immunol.* 172(11):7078–85
116. Ikonen M, Liu B, Hashimoto Y, Ma L, Lee K-W, et al. 2003. Interaction between the Alzheimer's survival peptide humanin and insulin-like growth factor-binding protein 3 regulates cell survival and apoptosis. *PNAS* 100(22):13042–47
117. Kim SJ, Guerrero N, Wassef G, Xiao J, Mehta HH, et al. 2016. The mitochondrial-derived peptide humanin activates the ERK1/2, AKT, and STAT3 signaling pathways and has age-dependent signaling differences in the hippocampus. *Oncotarget* 7(30):46899–912
118. Muzumdar RH, Huffman DM, Atzmon G, Buettner C, Cobb LJ, et al. 2009. Humanin: a novel central regulator of peripheral insulin action. *PLOS ONE* 4(7):e6334
119. Wang Y, Zeng Z, Zhao S, Tang L, Yan J, et al. 2021. Humanin alleviates insulin resistance in polycystic ovary syndrome: a human and rat model-based study. *Endocrinology* 162(8):bqab056
120. Yen K, Mehta HH, Kim SJ, Lue YH, Hoang J, et al. 2020. The mitochondrial derived peptide humanin is a regulator of lifespan and healthspan. *Aging* 12(12):11185–99

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

121. Cobb LJ, Lee C, Xiao J, Yen K, Wong RG, et al. 2016. Naturally occurring mitochondrial-derived peptides are age-dependent regulators of apoptosis, insulin sensitivity, and inflammatory markers. *Aging* 8(4):796–809
122. Nashine S, Cohen P, Nesburn AB, Kuppermann BD, Kenney MC. 2018. Characterizing the protective effects of SHLP2, a mitochondrial-derived peptide, in macular degeneration. *Sci. Rep.* 8(1):15175
123. Kim SK, Tran LT, NamKoong C, Choi HJ, Chun HJ, et al. 2023. Mitochondria-derived peptide SHLP2 regulates energy homeostasis through the activation of hypothalamic neurons. *Nat. Commun.* 14(1):4321
124. Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, et al. 2015. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab.* 21(3):443–54
125. Stein CS, Jadiya P, Zhang X, McLendon JM, Abouassaly GM, et al. 2018. Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency. *Cell Rep.* 23(13):3710–20.e8
126. Zhang S, Guo Y, Fidelito G, Robinson DRL, Liang C, et al. 2023. LINC00116-encoded microprotein mitoregulin regulates fatty acid metabolism at the mitochondrial outer membrane. *iScience* 26(9):107558
127. Kienzle L, Bettinazzi S, Choquette T, Brunet M, Khorami HH, et al. 2023. A small protein coded within the mitochondrial canonical gene *nd4* regulates mitochondrial bioenergetics. *BMC Biol.* 21(1):111
128. Miller B, Kim S-J, Mehta HH, Cao K, Kumagai H, et al. 2023. Mitochondrial DNA variation in Alzheimer's disease reveals a unique microprotein called SHMOOSE. *Mol. Psychiatry* 28(4):1813–26
129. Zhang S, Reljic B, Liang C, Kerouanton B, Francisco JC, et al. 2020. Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly. *Nat. Commun.* 11(1):1312
130. Makarewich CA, Munir AZ, Bezprozvannaya S, Gibson AM, Young Kim S, et al. 2022. The cardiac-enriched microprotein mitolamban regulates mitochondrial respiratory complex assembly and function in mice. *PNAS* 119(6):e2120476119
131. Rocha AL, Pai V, Perkins G, Chang T, Ma J, et al. 2024. An inner mitochondrial membrane microprotein from the *SLC35A4* upstream ORF regulates cellular metabolism. *J. Mol. Biol.* 436(10):168559
132. Huang N, Li F, Zhang M, Zhou H, Chen Z, et al. 2021. An upstream open reading frame in phosphatase and tensin homolog encodes a circuit breaker of lactate metabolism. *Cell Metab.* 33(1):128–44.e9
133. Rathore A, Chu Q, Tan D, Martinez TF, Donaldson CJ, et al. 2018. MIEF1 microprotein regulates mitochondrial translation. *Biochemistry* 57(38):5564–75
134. Mise S, Matsumoto A, Shimada K, Hosaka T, Takahashi M, et al. 2022. Kastor and Polluks polypeptides encoded by a single gene locus cooperatively regulate VDAC and spermatogenesis. *Nat. Commun.* 13(1):1071
135. Du D, Neuberger A, Orr MW, Newman CE, Hsu P-C, et al. 2020. Interactions of a bacterial RND transporter with a transmembrane small protein in a lipid environment. *Structure* 28(6):625–34.e6
136. Hobbs EC, Yin X, Paul BJ, Astarita JL, Storz G. 2012. Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. *PNAS* 109(41):16696–701
137. Lees JA, Messa M, Sun EW, Wheeler H, Torta F, et al. 2017. Lipid transport by TMEM24 at ER-plasma membrane contacts regulates pulsatile insulin secretion. *Science* 355(6326):eaah6171
138. Sun EW, Guillen-Samander A, Bian X, Wu Y, Cai Y, et al. 2019. Lipid transporter TMEM24/C2CD2L is a Ca^{2+}-regulated component of ER-plasma membrane contacts in mammalian neurons. *PNAS* 116(12):5775–84
139. Fu H, Wang T, Kong X, Yan K, Yang Y, et al. 2022. A Nodal enhanced micropeptide NEMEP regulates glucose uptake during mesendoderm differentiation of embryonic stem cells. *Nat. Commun.* 13(1):3984
140. Rathod N, Bak JJ, Primeau JO, Fisher ME, Espinoza-Fonseca LM, et al. 2021. Nothing regular about the regulins: distinct functional properties of SERCA transmembrane peptide regulatory subunits. *Int. J. Mol. Sci.* 22(16):8891
141. Reddy UV, Weber DK, Wang S, Larsen EK, Gopinath T, et al. 2022. A kink in DWORF helical structure controls the activation of the sarcoplasmic reticulum Ca^{2+}-ATPase. *Structure* 30(3):360–70.e6
142. Makarewich CA, Bezprozvannaya S, Gibson AM, Bassel-Duby R, Olson EN. 2020. Gene therapy with the DWORF micropeptide attenuates cardiomyopathy in mice. *Circ. Res.* 127(10):1340–42
143. Matsumoto A, Pasut A, Matsumoto M, Yamashita R, Fung J, et al. 2017. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* 541(7636):228–32

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

144. Bi P, Ramirez-Martinez A, Li H, Cannavino J, McAnally JR, et al. 2017. Control of muscle formation by the fusogenic micropeptide myomixer. *Science* 356(6335):323–27
145. Ducommun S, Jannig PR, Cervenka I, Murgia M, Mittenbühler MJ, et al. 2024. Mustn1 is a smooth muscle cell-secreted microprotein that modulates skeletal muscle extracellular matrix composition. *Mol. Metab.* 82:101912
146. Koh M, Ahmad I, Ko Y, Zhang Y, Martinez TF, et al. 2021. A short ORF-encoded transcriptional regulator. *PNAS* 118(4):e2021943118
147. Zhang M, Zhao K, Xu X, Yang Y, Yan S, et al. 2018. A peptide encoded by circular form of *LINC-PINT* suppresses oncogenic transcriptional elongation in glioblastoma. *Nat. Commun.* 9(1):4475
148. Hung P, Johnson B, Chen B-R, Byrum AK, Bredemeyer AL, et al. 2018. MRI is a DNA damage response adaptor during classical non-homologous end joining. *Mol. Cell* 71(2):332–42.e8
149. Slavoff SA, Heo J, Budnik BA, Hanakahi LA, Saghatelian A. 2014. A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. *J Biol. Chem.* 289(16):10950–57
150. Arnoult N, Correia A, Ma J, Merlo A, Garcia-Gomez S, et al. 2017. Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. *Nature* 549(7673):548–52
151. Yu R, Hu Y, Zhang S, Li X, Tang M, et al. 2022. LncRNA CTBP1-DT-encoded microprotein DDUP sustains DNA damage response signalling to trigger dual DNA repair mechanisms. *Nucleic Acids Res.* 50(14):8060–79
152. Ren L, Qing X, Wei J, Mo H, Liu Y, et al. 2023. The DDUP protein encoded by the DNA damage-induced CTBP1-DT lncRNA confers cisplatin resistance in ovarian cancer. *Cell Death Dis.* 14(8):568
153. Sun L, Wang W, Han C, Huang W, Sun Y, et al. 2021. The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. *Mol. Cell* 81(21):4493–508.e9
154. Luo Y, Schofield JA, Simon MD, Slavoff SA. 2020. Global profiling of cellular substrates of human Dcp2. *Biochemistry* 59(43):4176–88
155. Nita A, Matsumoto A, Tang R, Shiraishi C, Ichihara K, et al. 2021. A ubiquitin-like protein encoded by the "noncoding" RNA TINCR promotes keratinocyte proliferation and wound healing. *PLOS Genet.* 17(8):e1009686
156. Dubois M-L, Meller A, Samandi S, Brunelle M, Frion J, et al. 2020. *UBB* pseudogene 4 encodes functional ubiquitin variants. *Nat. Commun.* 11(1):1306
157. Morgado-Palacin L, Brown JA, Martinez TF, Garcia-Pedrero JM, Forouhar F, et al. 2023. The TINCR ubiquitin-like microprotein is a tumor suppressor in squamous cell carcinoma. *Nat. Commun.* 14(1):1328
158. Boix O, Martinez M, Vidal S, Giménez-Alejandre M, Palenzuela L, et al. 2022. pTINCR microprotein promotes epithelial differentiation and suppresses tumor growth through CDC42 SUMOylation and activation. *Nat. Commun.* 13(1):6840
159. Coulombe B, Cloutier P, Pinard M, Forget D, Poitras C, Gauthier M-S. 2020. The PAQosome, a novel molecular chaperoning machine for assembly of human protein complexes and networks. *FASEB J.* 34(S1). **https://doi.org/10.1096/fasebj.2020.34.s1.00078**
160. Bhatta A, Atianand M, Jiang Z, Crabtree J, Blin J, Fitzgerald KA. 2020. A mitochondrial micropeptide is required for activation of the Nlrp3 inflammasome. *J. Immunol.* 204(2):428–37
161. Shi T, Huang Y, Li Y, Dai X, He Y, et al. 2024. MAVI1, an endoplasmic reticulum-localized microprotein, suppresses antiviral innate immune response by targeting MAVS on mitochondrion. *Sci. Adv.* 9(35):eadg7053
162. Lee CQE, Kerouanton B, Chothani S, Zhang S, Chen Y, et al. 2021. Coding and non-coding roles of MOCCI (C15ORF48) coordinate to regulate host inflammation and immunity. *Nat. Commun.* 12(1):2130
163. Niu L, Lou F, Sun Y, Sun L, Cai X, et al. 2020. A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation. *Sci. Adv.* 6(21):eaaz2059
164. Erady C, Boxall A, Puntambekar S, Suhas Jagannathan N, Chauhan R, et al. 2021. Pan-cancer analysis of transcripts encoding novel open-reading frames (nORFs) and their potential biological functions. *npj Genom Med.* 6(1):4
165. Merino-Valverde I, Greco E, Abad M. 2020. The microproteome of cancer: from invisibility to relevance. *Exp. Cell Res.* 392(1):111997

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26

166. Zhou B, Yang H, Yang C, Bao Y, Yang S, et al. 2021. Translation of noncoding RNAs and cancer. *Cancer Lett*. 497:89–99
167. Li M, Li X, Zhang Y, Wu H, Zhou H, et al. 2020. Micropeptide MIAC inhibits HNSCC progression by interacting with aquaporin 2. *J. Am. Chem. Soc.* 142(14):6708–16
168. Li M, Liu G, Jin X, Guo H, Setrerrahmane S, et al. 2022. Micropeptide MIAC inhibits the tumor progression by interacting with AQP2 and inhibiting EREG/EGFR signaling in renal cell carcinoma. *Mol. Cancer* 21(1):181
169. Kang M, Tang B, Li J, Zhou Z, Liu K, et al. 2020. Identification of miPEP133 as a novel tumor-suppressor microprotein encoded by miR-34a pri-miRNA. *Mol. Cancer* 19(1):143
170. Wu S, Zhang L, Deng J, Guo B, Li F, et al. 2020. A novel micropeptide encoded by Y-linked LINC00278 links cigarette smoking and AR signaling in male esophageal squamous cell carcinoma. *Cancer Res*. 80(13):2790–803
171. Zhai S, Lin J, Ji Y, Zhang R, Zhang Z, et al. 2023. A microprotein N1DARP encoded by LINC00261 promotes Notch1 intracellular domain (N1ICD) degradation via disrupting USP10-N1ICD interaction to inhibit chemoresistance in Notch1-hyperactivated pancreatic cancer. *Cell Discov*. 9(1):95
172. Guo B, Wu S, Zhu X, Zhang L, Deng J, et al. 2020. Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. *EMBO J.* 39(1):e102190
173. Meng K, Lu S, Li Y, Hu L, Zhang J, et al. 2023. LINC00493-encoded microprotein SMIM26 exerts anti-metastatic activity in renal cell carcinoma. *EMBO Rep*. 24(6):e56282
174. Yang J-E, Zhong W-J, Li J-F, Lin Y-Y, Liu F-T, et al. 2023. LINC00998-encoded micropeptide SMIM30 promotes the G1/S transition of cell cycle by regulating cytosolic calcium level. *Mol. Oncol.* 17(5):901–16
175. Pang Y, Liu Z, Han H, Wang B, Li W, et al. 2020. Peptide SMIM30 promotes HCC development by inducing SRC/YES1 membrane anchoring and MAPK pathway activation. *J. Hepatol.* 73(5):1155–69
176. Cui Y, Peng J, Zheng M, Ge H, Wu X, et al. 2022. TRPC5OS induces tumorigenesis by increasing ENO1-mediated glucose uptake in breast cancer. *Transl. Oncol.* 22:101447
177. Ge Q, Jia D, Cen D, Qi Y, Shi C, et al. 2021. Micropeptide ASAP encoded by LINC00467 promotes colorectal cancer progression by directly modulating ATP synthase activity. *J. Clin. Investig.* 131(22):e152911
178. Zhang C, Zhou B, Gu F, Liu H, Wu H, et al. 2022. Micropeptide PACMP inhibition elicits synthetic lethal effects by decreasing CtIP and poly(ADP-ribosyl)ation. *Mol. Cell* 82(7):1297–312.e8
179. Yewdell JW, Nicchitta CV. 2006. The drip hypothesis decennial: support, controversy, refinement and extension. *Trends Immunol*. 27(8):368–73
180. Kozak M. 1989. The scanning model for translation: an update. *J. Cell Biol.* 108(2):229–41

Downloaded from www.annualreviews.org. Guest (guest) IP: 172.59.173.216 On: Sat, 30 May 2026 19:33:26