ISSN 0006-2979, Biochemistry (Moscow), 2025, Vol. 90, No. 11, pp. 1521-1535 © The Author(s) 2025. This article is an open access publication.
Russian Text © The Author(s), 2025, published in Biokhimiya, 2025, Vol. 90, No. 11, pp. 1621-1637.

REVIEW

Methodological Toolbox for Identifying and Studying Micropeptides: From Genome to Function

Aleksandr I. Lavrov[1,2], Nikita M. Shepelev[1,3], Olga A. Dontsova[1,3], and Maria P. Rubtsova[1,3,a]*

[1]*Lomonosov Moscow State University, Faculty of Chemistry, 119991 Moscow, Russia*

[2]*Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119992 Moscow, Russia*

[3]*Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117997 Moscow, Russia*

[a]*e-mail: mprubtsova@gmail.com*

Received July 22, 2025
Revised October 2, 2025
Accepted October 6, 2025

Abstract—Micropeptides encoded by small open reading frames (sORFs) represent a novel, actively studied class of functional molecules regulating key cellular processes. Studying micropeptides is complicated by methodological challenges, in particular, their small size, low cellular abundance, and difficulty in generating specific antibodies. The review systematizes modern approaches to the identification and functional characterization of micropeptides. The main strategies for their discovery include the use of bioinformatic algorithms, global translation analysis via ribosome profiling, direct detection using mass spectrometry-based proteomics, and phenotypic screenings. The methods for confirming the functions of micropeptides and elucidating molecular mechanisms of their action genetic knockouts, affinity tagging for visualization, and investigation of protein-protein interactions. The review discusses key challenges and future prospects in the field, emphasizing the importance of an integrated multi-omics approach for the comprehensive micropeptidome mapping.

DOI: 10.1134/S0006297925602242

Keywords: micropeptides, small open reading frames, long non-coding RNAs, mass spectrometry, Ribo-Seq, affinity labeling, co-immunoprecipitation

INTRODUCTION

Any extended nucleotide sequence, whether natural or random, contains numerous open reading frames (ORFs), as random nucleotide combinations form various start and stop codons [1]. Historically, large genome annotation consortia have focused exclusively on identifying protein-coding ORFs and ignored small ORFs (sORFs) encoding peptides shorter than 100 amino acid residues [2]. For a long time, sORFs had been considered an untranslated genomic "noise," incapable of encoding stable and functional peptides.

This conception has changed dramatically with the emergence of high-throughput sequencing technologies, particularly, ribosome profiling (Ribo-Seq), which enables global mapping of translated RNA regions at a single-nucleotide resolution [3]. Combined with the proteomic data [4, 5], these studies have provided compelling evidence for the translation of thousands of sORFs in various organisms, including humans [6, 7]. Furthermore, functional studies have shown that many sORF translation products – micropeptides – are involved in the regulation of fundamental processes, such as signaling, metabolism, homeostasis, muscle activity, DNA repair, and immune response [8-11]. Moreover, micropeptides have been found to influence the development

* To whom correspondence should be addressed.

of various pathologies, including cancer. In other words, micropeptides represent the "dark matter" of the proteome and possess a significant regulatory potential [12].

Due to some properties of micropeptides, including their low cellular abundance, potential instability, and difficulty in generating specific antibodies, the search for these compounds and their characterization are associated with methodological challenges, which necessitated adaptation of existing methods and development of new approaches for the reliable identification, confirmation, and functional analysis of micropeptides. The goal of this review is to provide a comprehensive and systematic analysis of modern methodological arsenal used in the studies of micropeptides and to discuss the prospects for new developments in this field.

MICROPEPTIDES

sORFs are short nucleotide sequences (no longer than 100 codons from the start codon to the stop codon) in DNA or RNA. A significant number of sORFs are located within protein-coding genes, in particular, in the 5′-untranslated regions (5′-UTRs) of mRNA (the so-called upstream ORFs that often play a regulatory role in the translation of the main coding sequence), 3′-untranslated regions (3′-UTRs) of mRNA (downstream ORFs), or in the coding sequences in alternative reading frames. Additionally, many sORFs have been identified in transcripts initially classified as non-coding RNAs, including long non-coding RNAs (lncRNAs), primary microRNA transcripts, and circular RNAs [13].

sORFs can serve as templates for the synthesis of short proteins, called micropeptides or microproteins. Over two decades, numerous micropeptides involved in embryogenesis [8, 9], metabolism [10, 11], and DNA repair [14] have been described. Some micropeptides are known to promote carcinogenesis, while others act as tumor suppressors [15, 16]. Micropeptides with the neuroprotective properties that inhibited the development of the Alzheimer's disease have also been identified [17].

For example, physiologically active micropeptide DWORF functions as an activator of the calcium ATPase SERCA responsible for transporting Ca^{2+} ions from the cytoplasm into the sarcoplasmic reticulum (a process necessary for muscle relaxation). Overexpression of DWORF in the cardiac muscle increases SERCA activity, improves myocardial contractility, and affects calcium homeostasis [18, 19]. Conversely, the micropeptides myoregulin (MLN), phospholamban (PLN), and sarcolipin (SLN) inhibit SERCA and suppress its activity [20, 21].

METHODS FOR MICROPEPTIDE DISCOVERY

The search for and identification of functional micropeptides is a multi-step process that begins with a large-scale screening of genome and transcriptome for potential candidates. Historically, three main approaches have been developed: bioinformatic sequence analysis, global translation mapping by ribosome profiling, and direct peptide detection using mass spectrometry-based proteomics. In the last decade, functional phenotypic screening has been added to this list, allowing direct identification of micropeptides involved in specific cellular processes (Fig. 1).

BIOINFORMATIC APPROACHES

The bioinformatic search for sORFs is a non-trivial task. The early methods of genome analysis used to predict coding ORFs have set the minimal protein product length as 100 amino acid residues. This resulted in the loss of information about functional peptides shorter than this threshold, despite known existence of such molecules [22].

Firstly, the bioinformatic search for sORFs based on the evolutionary sequence conservation across different species [23]. An example of this approach is the use of the Ka/Ks metric, which reflects the ratio of non-synonymous to synonymous substitutions in codons in the nucleotide sequence alignment for different species and indicates whether the sequence is under selective pressure [24]. More advanced methods have been developed later, e.g., PhyloCSF, which uses codon substitution frequencies in both coding and non-coding genomic regions [25].

To improve the reliability of predictions, other features indicating the evolutionary conservation of a sequence should be considered as well, such as the absence of frameshift-causing insertions or deletions, a decrease in the sequence conservatism at the edges of coding regions, and others [26]. However, for short sequences, the statistical significance of such analysis is low [27]. Moreover, many sORFs are species-specific or arise *de novo* in the genome, so a functional peptide might not exhibit the sequence conservatism and will be missed in analysis [6, 13].

The identification of sORFs is also possible through the analysis of known features of coding sequences, such as their codon composition, GC content, and others. One of the early algorithms involved identification of coding sequences by comparing the codon frequencies in coding and non-coding genomic regions [22, 28]. Another approach used analysis of six mathematical metrics of DNA sequence to assess the coding potential of sORFs [29]. Machine learning-based

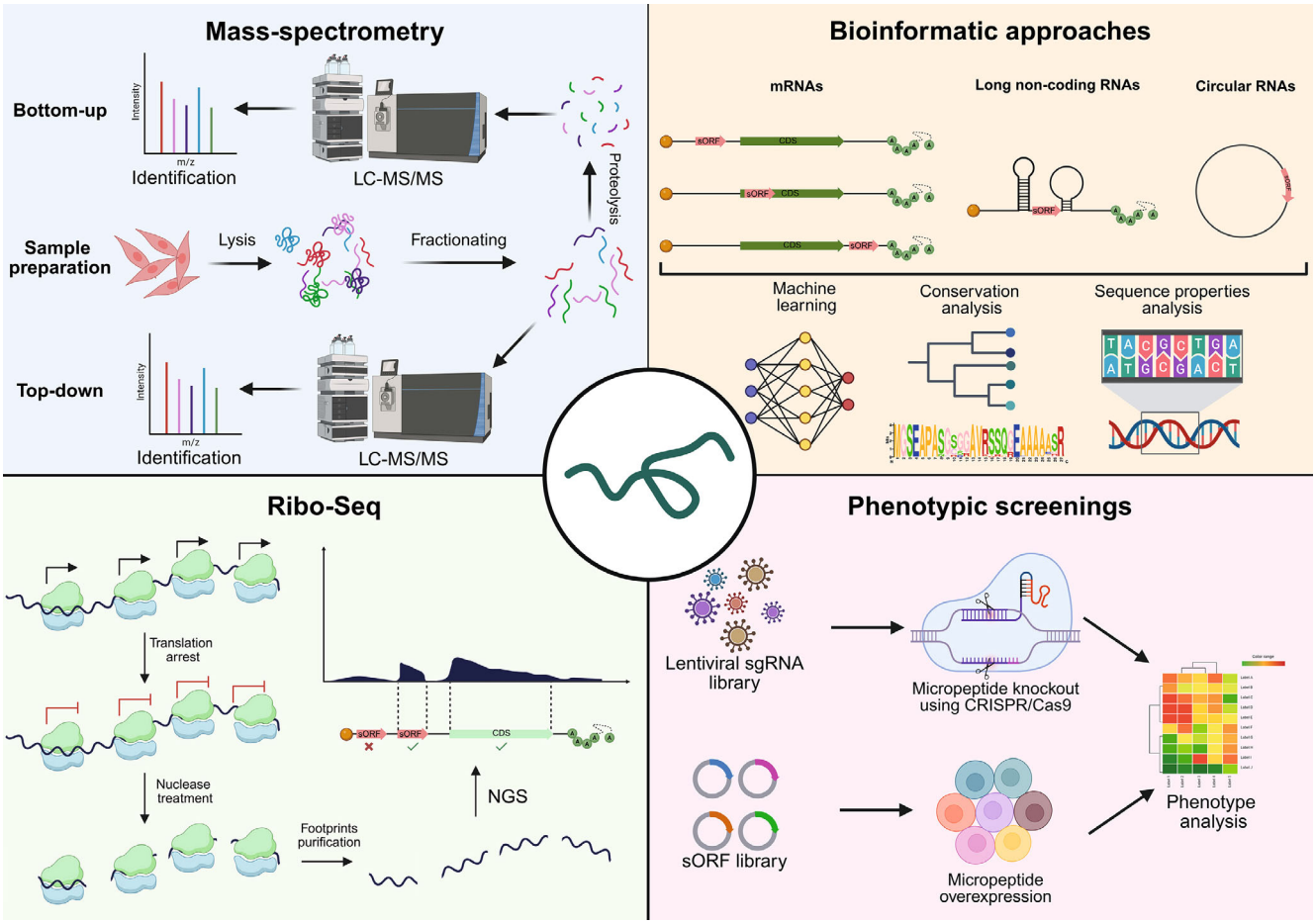

Fig. 1. Main approaches to the large-scale micropeptide discovery. Bioinformatic approaches include analysis of nucleotide sequences to predict the coding potential of sORFs. The key methods are assessment of evolutionary conservation across species, evaluation of sequence properties (e.g., codon composition), and application of machine learning algorithms to integrate various features. Ribo-Seq (ribosome profiling) is an experimental method that maps all actively translated regions of the transcriptome. It is based on ribosome stalling, followed by enzymatic cleavage of mRNA and isolation and high-throughput sequencing of ribosome-protected fragments (footprints) in order to detect sORF translation *in vivo*. Mass spectrometry enables direct detection of sORF translation products and can be employed as either bottom-up strategy, in which proteins are first digested by proteases into peptides before analysis, or top-down strategy that implies analysis of intact micropeptides. Both strategies require careful sample preparation, including cell lysis, fractionation for the enrichment with low-molecular-weight proteins, and analysis by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Phenotypic screenings aim to identify functional micropeptides. The loss-of-function screenings use guide RNA libraries to knock out sORFs using CRISPR/Cas9, followed by the analysis of changes in the cell phenotype (e.g., cell proliferation rate). The gain-of-function screenings use sORF libraries to identify micropeptides whose overexpression induces a specific cellular response.

analysis, which has become a common research tool in recent years, allows identification of complex, non-linear patterns in large datasets, as well as integration of various features of analyzed sequences, including their evolutionary conservatism, predicted structure, and other calculated metrics, to build highly accurate predictive models [30, 31].

Most modern studies on the identification of coding sORFs use the same approaches for the analysis of high-throughput sequencing data – ribosome profiling (Ribo-Seq) and RNA sequencing (RNA-Seq), which significantly increases the reliability of obtained results [32, 33].

RIBOSOME PROFILING

High-throughput ribosome profiling was first proposed in 2009 [3]. In the classic version of the method, cells are treated with cycloheximide, which binds to the E-site of the 60S ribosomal subunit and inhibits translocation by stopping it at the translation elongation stage [34]. The cells are then lysed and treated with nucleases to cleave RNA, while regions protected by stalled ribosomes are preserved. These fragments, which are approximately 30 nucleotides long, are called ribosome footprints. They are purified and sequenced using high-throughput methods.

The resulting sequences are mapped onto the transcriptome assembled based on the RNA-Seq data to obtain a ribosome profile, i.e., the positions of ribosomes on mRNAs at the moment of translation arrest with a single-nucleotide resolution [3]. The main advantage of the approach is its ability to reflect the translation levels of specific mRNAs at a given time, thus allowing to study rapid changes in gene expression in response to external factors or during specific cellular processes [35, 36].

Despite its broad capabilities, RNA-Seq has several limitations. In particular, the experimental results depend significantly on the sample quality. The most common problem is co-isolation of large ribonucleoprotein complexes and non-coding RNAs along with 80S ribosomes, leading to false signals [37]. Also, the percentage of footprint reads from the obtained library is limited due to the contamination with ribosomal RNA. Another fundamental problem is distortion of ribosome profiles because of the action of certain antibiotics [38].

Other difficulties are associated with the analysis of sORFs. Because of the sORF length, it is sometimes impossible to accurately determine the translation initiation site if several potential start codons are located nearby [36]. Another problem arises when sORFs overlap with the main ORF or are located within it [38].

To determine the exact position of start codons, researchers use translation initiation inhibitors. These compounds do not disrupt elongation or translation termination, but stalls ribosomes mostly in the translation initiation regions, so ribosome profiles provide accurate information about the position of ribosome [20]. The first antibiotic used to map translation initiation sites was harringtonine [39] that binds to the A-site of free 60S ribosomal subunit. After formation of the 80S ribosome during translation initiation, harringtonine blocks the transfer of methionine from the initiator tRNA to the aminoacyl-tRNA located in the A-site, leading to ribosome stalling at the translation initiation site [40]. Importantly, harringtonine does not bind to the 60S subunit in the content of the 80S ribosome and does not affect translation elongation and termination; hence, the obtained footprints show positions of translation initiation sites [39, 40]. However, when using harringtonine, some identified footprints were found for the region located downstream of the start codon, which, in certain cases, failed to provide a sufficiently narrow peak for the initiation site mapping. A more accurate identification of start codons can be achieved by using the GTI-Seq (Global Translation Initiation Sequencing) method, which employs lactimidomycin to stall translation initiation. Lactimidomycin blocks the empty E-site of the 80S ribosome and causes ribosome to stall immediately after translation initiation [41]. In a subsequent modification of the method, QTI-Seq (Quantitative Translation Initiation Sequencing), lysed cells are treated with lactimidomycin for a short period of time, followed by the addition of puromycin to causes the dissociation of elongating ribosomes. This approach allows to achieve the maximum coverage of translation initiation sites, while reducing the noise from elongating ribosomes and artifacts associated with prolonged incubation of cells with lactimidomycin [42].

Although ribosome profiling can identify translated regions, it cannot establish whether the identified sORFs are coding, regulatory, or non-functional [43]. An indirect evidence of coding potential can be obtained through bioinformatic analysis of candidate sequences [44, 45]. The existence of an encoded peptide in the cell can be proven by mass spectrometry methods [46, 47].

MASS SPECTROMETRY

Mass spectrometry-based proteomic analysis allows direct detection of sORF translation products. As in classical proteomics, either peptide fragments after digestion with proteases (bottom-up approach) or intact proteins (top-down approach) can be analyzed. In both cases, identification of micropeptides faces two main obstacles. First, it requires the use of special sample preparation procedures to enrich the sample with micropeptides because of their small size, low stability, and low cellular abundance. Second, identification of previously unannotated micropeptides requires the use of special reference databases, as databases based on Ensembl, Ref-Seq, or UniProt do not contain most potential micropeptides.

To prevent the degradation of micropeptides, cell lysates are heated to 95°C to inactivate proteases [5]. In some cases, this procedure is followed by precipitation of large proteins with trichloroacetic acid [48]. The resulting samples are fractionated by polyacrylamide gel electrophoresis (PAGE) [49, 50], size-exclusion chromatography [48], or reverse-phase chromatography [51] to enriched them with low-molecular-weight proteins. More advanced methods of preliminary fractionation have been developed, such as GELFrEE (Gel-Eluted Liquid Fraction Entrapment Electrophoresis, i.e., separation on a column containing polyacrylamide gel) [52, 53], ERLIC (Electrostatic Repulsion Hydrophilic Interaction Chromatography, i.e., chromatography based on electrostatic repulsion and hydrophilic interaction) [54], and others. Depending on the selected approach, the proteins are then either subjected to proteolysis (most often, with trypsin) for the bottom-up analysis or directly analyzed by liquid chromatography combined with tandem mass spectrometry (LC-MS/MS).

The identification of micropeptides is based on matching experimental mass spectra with the theoretical ones. However, construction of databases poses a problem, as *in silico* translation of all reading frames in the genome or transcriptome covers almost all potential micropeptides, but the size of the resulting database is tens of times larger than that of standard databases, leading to a high computational complexity and high probability of false identifications [55]. Therefore, limited databases based on RNA-seq and Ribo-Seq data are used [5, 56], which, however, does not exclude the appearance of false-positive results. In order to confirm an identified micropeptide its isotopically labeled synthetic analog, which has an identical mass spectrum that is shifted relative to the spectrum of the validated peptide, is added in the reaction [57, 58].

Micropeptides can be studied using the top-down proteomics approach, which is particularly useful for analyzing various proteoforms of micropeptides formed by alternative splicing [59] and post-translational modifications [57]. For example, identification of micropeptides presented by the major histocompatibility complex (MHC-I) allows to detect intact peptides. A database of translated sequences obtained by the ribosome profiling of cancer cells, was used to identify thousands of previously unannotated peptides by mass spectrometry [60].

PHENOTYPIC SCREENINGS

In addition to the analysis of sequences and translation data, functional sORFs can be discovered through phenotypic screening, which can be divided into two approaches: loss-of-function screening and gain-of-function screening.

CRISPR/Cas9 system has become the main tool for the loss-of-function screening. In these experiments, single-guide RNAs (sgRNAs) target Cas9 nuclease to specific sORF regions due to a 20-nucleotide guide sequence in the sgRNA and the presence of the protospacer adjacent motif (PAM) sequence immediately downstream of the targeted DNA region. Double-strand breaks caused by Cas9 are repaired mostly through the non-homologous end joining (NHEJ), leading to the insertions or deletions of nucleotides that can cause frameshifts and appearance of premature stop codons, thus preventing the biosynthesis of a functional micropeptide [61].

Genome-wide and targeted screenings use the sgRNA libraries targeting a large number of sORFs. The libraries are typically delivered to the cells by lentiviral vectors, allowing inactivation of thousands of genomic loci in a single experiment [62]. Cells with specific phenotypes resulting from the sORF knockout are selected from the general population, and the cassettes containing the sgRNA sequences are sequenced to identify the sORFs whose knockout caused changes in the cell phenotype. Averaging the results across multiple sgRNAs for a specific sORF allows to account for the potential off-target effects of CRISPR/Cas9.

The phenotypes observed after CRISPR-mediated inactivation of sORFs can range from changes in cell growth, viability, morphology, signaling pathways, and drug resistance to interactions with other molecules. A specific phenotype used in the screening is chosen in each particular experiment [62].

So far, CRISPR/Cas9-mediated phenotype-based screenings have not yet been commonly accepted as a method for identification of functional micropeptides. However, several functional micropeptides have been discovered in proliferation screenings of libraries based on Ribo-Seq data [63-65].

An alternative approach to the phenotype analysis involves exogenous expression of constructs encoding peptides from a pre-compiled library and cloned into plasmid or lentiviral vectors [66]. The advantage of this method is that it allows to analyze any sequence without requiring the presence of specific motifs (as in the case of CRISPR/Cas9). However, this approach has the same drawbacks as the commonly used overexpression of proteins for studying their functions.

CONFIRMATION OF EXISTENCE AND CHARACTERIZATION OF MICROPEPTIDES

Although the methods for investigating micropeptides are similar to those used for conventional proteins, the overall approaches to studying these classes of molecules differ. Unlike “classical” proteins, most micropeptides do not contain potential structural domains, have a small size, and are present in cells at low concentrations, which requires adjustment of standard approaches. Moreover, if a micropeptide is encoded in an lncRNA, it may be necessary to distinguish between the functional roles of the transcript and the peptide encoded within it. The studies of micropeptides include three main stages: candidate selection, confirmation of its existence, and determination of its function.

There are two main approaches for the search of potential candidates. The first is the analysis of data obtained by high-throughput methods (see the first part of the review). Typically, these methods provide only indirect evidence of the micropeptide existence, so they have to be supplemented by mass spectrometry and Ribo-Seq data [46]. A completely different approach is a manual search for sORFs and annotation of the transcripts of interest. For example, the micropeptide MIEF1-mp was found during the

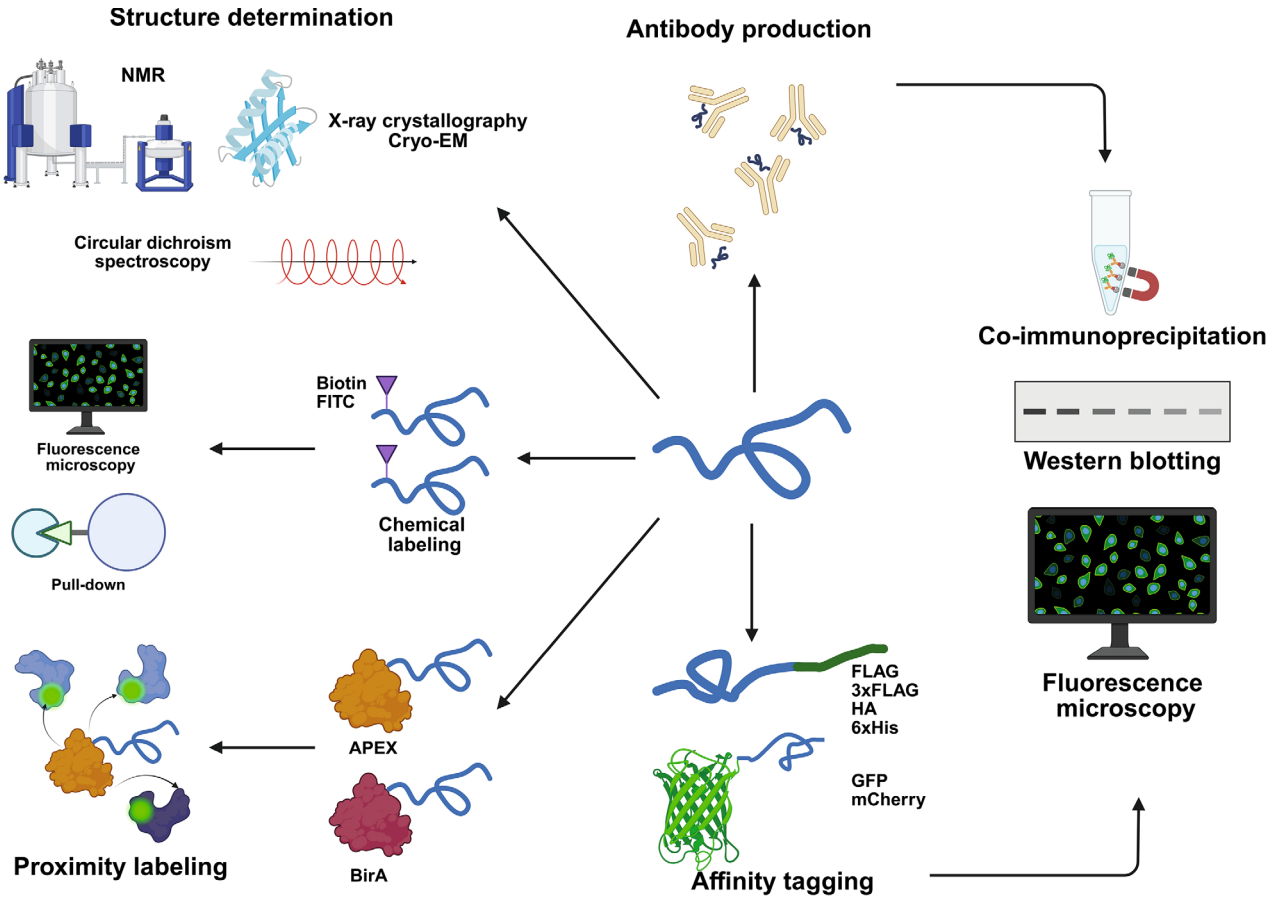

Fig. 2. Main strategies for studying micropeptides. The diagram shows key methods used to confirm the existence and determine the structure, location, function, and molecular partners of a selected micropeptide candidate. The structure of the micropeptide is determined by analysis of its secondary and tertiary structures. Thus, circular dichroism spectroscopy is used to assess elements of the secondary structure (α-helices, β-sheets). Nuclear magnetic resonance (NMR) allows to determine the three-dimensional structure of small peptides in solution, while X-ray crystallography and cryo-electron microscopy are more often used to study the complexes of micropeptides with their partner proteins. Affinity labeling is a central approach in the studies of micropeptides. Fusing a micropeptide with short affinity tags (FLAG, HA, 6xHis) or fluorescent proteins (GFP, mCherry) allows its visualization in the cells by fluorescence microscopy, detection by Western blotting, and isolation of protein complexes by co-immunoprecipitation (Co-IP). An alternative to affinity labeling is generation of specific antibodies to native micropeptides. The studies of micropeptide interactions includes several approaches. Chemical labeling of synthetic micropeptides (e.g., with biotin) allows to search for the interacting proteins *in vitro* in pull-down experiments, whereas weak or transient interactions *in vivo* can be investigated by proximity labeling methods using BirA ligase (BioID) or ascorbate peroxidase (APEX), which covalently label neighboring proteins.

analysis of the 5′-UTR sequence of the *MIEF1* gene mRNA that encodes a described protein product [67]. Another example of this approach is the discovery of the cancer-associated peptide PACMP in the analysis of differential expression of lncRNAs in breast cancer cells [68].

The second stage is confirmation of the micropeptide presence in the cells, as the protein product of sORF translation may be unstable. In some cases, it might be necessary to perform RACE (Rapid Amplification of cDNA Ends) to confirm the sequence of the target transcript in the cell [69]. Direct evidence of the micropeptide biosynthesis can be obtained by mass spectrometry. For example, the presence of the APPLE peptide was confirmed by mass spectrometry after immunoprecipitation of the endogenously labeled peptide [46]. However, in most cases, the confirmation of translation of the studied sORF and biosynthesis of a stable protein product can be achieved in a simpler experiment by using exogenous expression of the micropeptide fused with an affinity tag or a fluorescent protein [70, 71].

The next step after confirmation of the micropeptide biosynthesis is elucidation of its function. The methods used to study micropeptides are almost the same as those used for conventional proteins and include affinity labeling, Western blotting, immunocytochemistry (ICC), proximity labeling, etc. (Fig. 2), although most of these classical methods have to be modified due to the small size of micropeptides.

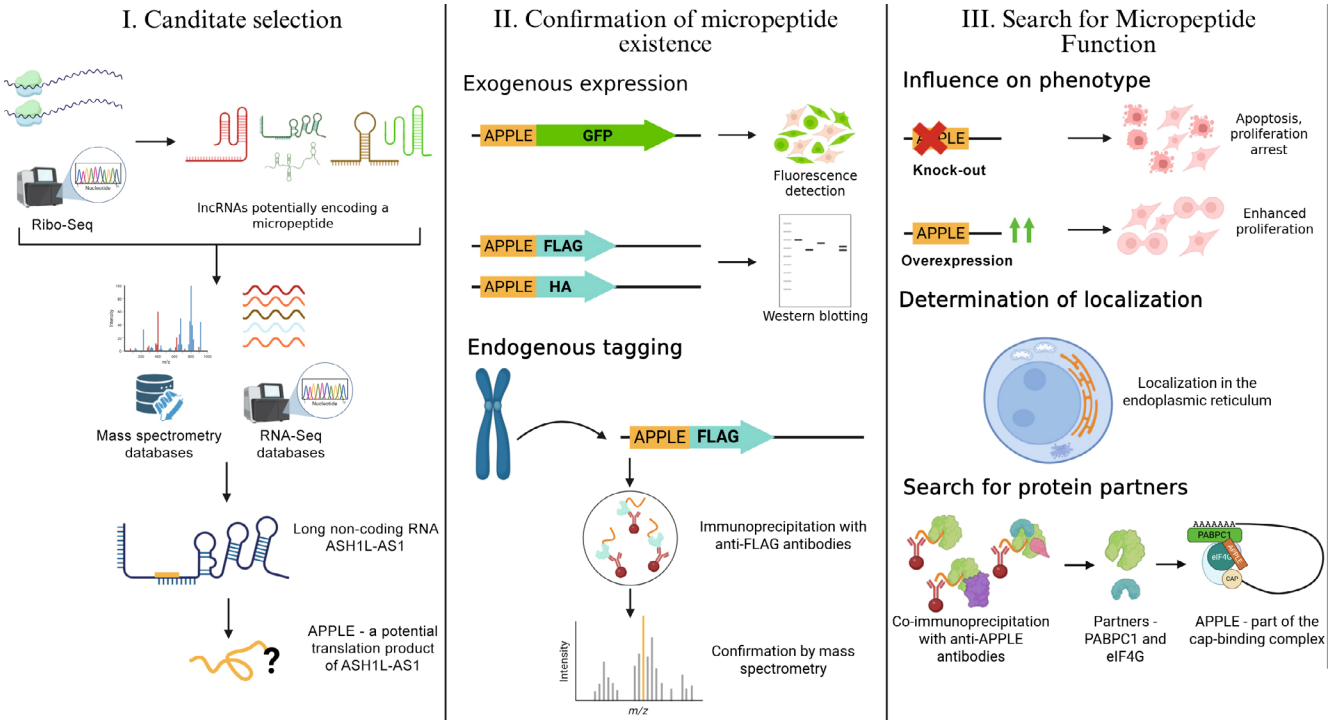

Fig. 3. The scheme of studies of the oncogenic micropeptide APPLE as an example of a comprehensive approach to micropeptide characterization. The diagram illustrates the three main stages of the study. I. Candidate selection: integrative analysis of high-throughput Ribo-Seq, mass spectrometry, and RNA-Seq data led to the identification of the lncRNA ASH1L-AS1 as a potential source of the micropeptide. II. Confirmation of the micropeptide existence: APPLE biosynthesis in cells was confirmed through two independent methods. (1) Exogenous expression: the coding sequence of APPLE was fused with various affinity and fluorescent tags (GFP, FLAG, HA), and the expression products were detected by fluorescence microscopy and Western blotting. (2) Endogenous tagging: the FLAG tag was inserted into the native gene locus using the CRISPR/Cas9 genome editing system, which enabled isolation of the endogenous peptide via immunoprecipitation. III. Elucidation of micropeptide function by functional analysis. (1) Studying the effect of peptide expression on the cell phenotype. Inactivation of APPLE expression led to cell apoptosis and proliferation arrest, while its overexpression promoted cell proliferation, indicating the oncogenic role of APPLE. (2) Determination of APPLE intracellular location: immunofluorescence analysis revealed that APPLE localizes to the endoplasmic reticulum. (3) Search for the APPLE protein partners. Co-IP followed by mass spectrometry identified PABPC1 and eIF4G as micropeptide interacting partners. Subsequent studies confirmed that APPLE is a part of the cap-binding complex and enhances translation initiation, thus promoting the development of malignancies.

The transcript of the lncRNA ASH1L-AS1 as a potential source of the functional APPLE micropeptide, which promotes development of hematopoietic malignancies, was first identified by integration of Ribo-Seq, mass spectrometry, and RNA-Seq data (Fig. 3). The biosynthesis of APPLE was confirmed by multiple methods, including generation of specific antibodies against the micropeptide and endogenous tagging with the CRISPR/Cas9 system, followed by detection by Western blotting, immunoprecipitation, and mass spectrometry analysis.

The functional role of APPLE in oncogenesis was established using knockdown and overexpression experiments, which demonstrated its effects on proliferation, differentiation, and apoptosis of acute myeloid leukemia cells, both *in vitro* and in *in vivo* (in a mouse model). A critical step in this study was the phenotypic complementation experiment on the background of the APPLE inactivation, which confirmed that the observed effects were specifically linked to the micropeptide rather than lncRNA ASH1L-AS1 [46].

The information on the methods used to characterize micropeptides is summarized in Online Resource 1.

KNOCKOUT AND KNOCKDOWN

To confirm the biological function of micropeptides, researchers can generate the cells in which the corresponding encoding sequences are knocked-out with genome editing systems (CRISPR/Cas9 or CRISPR-Cas12a). Analysis of phenotypic changes in these cells allows for assessing the role of investigated peptide. However, in some cases, selecting guide RNAs may be impossible due to the requirement for a PAM motif. The commonly used phenotypic markers are cell proliferation rate, changes in the cell cycle, and others [72]. An additional confirmation of the biological

significance of a micropeptide is the phenotype restoration through its exogenous expression in the cells where this micropeptide had been the knocked out [68].

A much less labor-intensive method for suppressing expression of a target micropeptide is RNA interference (RNAi)-mediated knockdown. Thus, small interfering RNAs were used in the studies of micropeptides translated from lncRNAs [46, 73]. However, in the case of sORFs encoded by mRNAs, RNAi-mediated knockdown will lead to the degradation of the entire mRNA molecule, and consequently, the observed effects may be associated with a decrease in the expression of the main translation product of this mRNA. In such cases, 2′-O-methyl antisense RNA oligonucleotides complementary to the region around the sORF start codon can be used to suppress the biosynthesis of the micropeptide but not of the main protein product [74].

AFFINITY TAGGING OF MICROPEPTIDES

Another important method in studying micropeptides is introduction of affinity tags. The small size of micropeptides imposes restrictions on the choice of the tag, as for many of them, the size of the introduced sequence is comparable to the size of the peptide itself. Moreover, the charge of the tag also plays a significant role. A case of incorrect choice the tag can disrupt the native localization and structure of the micropeptide or its interaction with partner proteins.

The most common tags used with micropeptides are small protein sequences, such as HA [75], FLAG [11, 46], 6xHis [70]. In some studies, GFP or another fluorescent protein are used. Fluorescent proteins can be easily detected by microscopy and Western blotting [10, 70], but there is a risk that the functions of the tagged micropeptide in the cells can be disrupted.

Endogenous tagging by CRISPR/Cas9 is also used to confirm the existence of micropeptides in the cells. Insertion of affinity tag-encoding sequences directly into the genomic locus allows for the detection of micropeptides expressed from the native promoters, which excludes artifacts associated with the overexpression of micropeptides and is a reliable method for validating their biosynthesis.

The methods of affinity tagging of micropeptides are discussed in detail in a recent review [76].

WESTERN BLOTTING

Western blotting is a standard method for detection and semi-quantitative determination of proteins. However, the small size of micropeptides limits the number of highly antigenic suitable epitopes, making it difficult to generate antibodies against micropeptides. Nevertheless, for some peptides (over 50 amino acids in length), the antibodies were successfully obtained and used for Western blotting [70-78]. In most cases, micropeptides are labeled with various affinity tags (see above) for further detection [10, 70, 79]. It is important to note that the content of native or endogenously tagged micropeptides in the cells can be very low, making their detection extremely challenging.

Another limitation is the resolving power of polyacrylamide gels used for the separation of proteins from cell lysates. To increase the resolution of low-molecular-weight proteins, researchers use the Tris-tricine buffer system [80]. For example, it was employed for the detection of the FLAG-tagged HOXB-AS3 micropeptide (7 kDa) and native form of myoregulin (10 kDa) [71, 81]. However, small proteins weakly bind to polyvinylidene fluoride and nitrocellulose membranes used for the transfer and can be easily lost during multiple membrane washes. In such cases, alternative techniques can be applied, such as fixation on the membrane by crosslinking with blocking proteins using formaldehyde or glutaraldehyde [82].

DETERMINATION OF MICROPEPTIDE INTRACELLULAR LOCATION

The location of a micropeptide can help in elucidating its protein partners and biological function. A micropeptide located in the nucleus is likely to interact with nuclear proteins, such as transcription factors or chromatin-organizing proteins.

The most common method for determining the location of micropeptides is immunocytochemistry, which relies on the use of antibodies with a high specificity for the target antigen. The procedure typically begins with cell fixation to preserve their morphology and prevent degradation of cellular components. The cells are then permeabilized to allow antibody access into the cell. Generally, primary antibodies used in the analysis are unlabeled and visualized using secondary antibodies conjugated with a fluorophore.

Similar to Western blotting, this method requires antibodies against the target micropeptide, which can be an obstacle. For example, antibodies against the native micropeptide MP31 were used to determine its mitochondrial location [83]. More often, immunocytochemical staining is performed for the exogenously expressed tagged micropeptide [45, 72, 79].

An alternative, widely used approach is the fusion of micropeptides with fluorescent proteins (GFP, mCherry, etc.), which allows their direct visualization in the cells without the use of antibodies. However, it may disrupt the properties of the micropeptide

because of the large size of the tag. To overcome this limitation, split fluorescent proteins can be used, when fluorescent protein, such as GFP, is divided into two non-functional fragments – large fragment GFP1-10 and small GFP11, consisting of 16 amino acids and used to tag the peptide of interest. The complementation of these fragments restores the tertiary structure of the protein and its fluorescence [84, 85]. This system was used to determine the location of the PIGBOS micropeptide, when the peptide tagged with three GFP11 repeats was co-expressed with GFP1-10 [77].

Chemical fluorescent tags have the least impact on the native localization of micropeptides. However, such experiments can be technically difficult due to the need for the chemical synthesis of the micropeptide for its chemical labeling with a fluorescent ligand (e.g., FITC). This labeling method was used to determine the location of the MP155 micropeptide [78].

The MicroID method is another tagging-based method for identification of new micropeptides with specific localization. This technique (a modification of the BioID method, see below) uses biotin ligases targeted to specific cellular compartments. After isolation of covalent complexes, those containing ligase crosslinked with low-molecular-weight proteins are selected by fractionation and the bound micropeptides are identified by mass spectrometry [86].

IDENTIFICATION OF PROTEIN PARTNERS

Identification of protein partners of micropeptides is critically important for understanding the biological mechanisms involving these micropeptides. The most widely used approach for studying protein–protein interactions *in vivo* is Co-IP, in which cells expressing a micropeptide labeled with an affinity tag (most commonly, FLAG or HA) are lysed under mild conditions that preserve existing protein complexes. Antibodies specific to the affinity tag are added to the cell lysate to bind the tagged micropeptide along with all proteins that interact with it. The resulting immune complexes are isolated from the solution, usually with agarose or magnetic beads coated with antibody-binding proteins A or G. After thorough washing to remove non-specifically bound proteins, the captured complexes are eluted and identified, most often by immunoprecipitation mass spectrometry (IP-MS) or Western blotting. This method was used to demonstrate the interaction of the CYREN micropeptide with Ku70/80 and other proteins involved in DNA repair [14]. The main requirement for Co-IP is a sufficient stability of protein complexes to withstand the lysis and washing procedures. However, the binding of a micropeptide to its protein partners is often not strong enough.

The pull-down is a method similar to Co-IP, except that the "bait" is typically a purified, tagged micropeptide immobilized on a solid phase. This can be achieved through exogenous expression (e.g., with the GST or 6xHis tag) or chemical synthesis, often with the addition of the biotin tag. The immobilized bait is incubated with the cell lysate or a solution of purified proteins. Proteins that bind to the bait are retained on the carrier. The carrier is washed, and the proteins are then eluted and analyzed. For example, a pull-down experiment with biotinylated P155 peptide immobilized on streptavidin beads revealed the interaction between this peptide and HSC70 chaperone [78]. Unlike Co-IP, pull-down is convenient for confirming direct protein interactions *in vitro* and does not require specific antibodies. However, the results may not take into account the influence of the cellular context or the post-translational modifications necessary for the interactions *in vivo*.

To detect weaker or transient interactions *in vivo*, proximity labeling methods, such as BioID and APEX, have been developed [87, 88]. These methods are based on the expression of a chimeric protein consisting of the target micropeptide fused with a special enzyme. The enzyme is activated by the addition of specific substrates that generate short-lived reactive molecules (usually, biotin derivatives), which covalently bind to proteins in a close proximity (within a few nanometers) to the chimeric protein. The BioID method uses the mutant biotin ligase BirA*, which, in the presence of biotin and ATP, generates activated biotin-AMP that reacts with lysine residues of neighboring proteins [89]. In the APEX method, ascorbate peroxidase in the presence of biotin-phenol and a short pulse of hydrogen peroxide rapidly (within minutes) generates the biotin-phenoxyl radical that labels tyrosine residues in a close proximity to the enzyme. After the labeling reaction, the cells are lysed, biotinylated proteins are isolated using streptavidin and identified by mass spectrometry [89]. The APEX labeling method was used to search for the protein partners of the mitochondrial peptide MIEF1 [79]. APEX labeling can also provide information on the subcellular localization of micropeptides and their functional environment, as demonstrated for C11orf98 in the nucleolus [90].

DETERMINATION OF MICROPEPTIDE STRUCTURE

X-ray crystallography is rarely used for determining the three-dimensional structure of micropeptides, as it requires crystallization of proteins or peptides into an ordered three-dimensional lattice to obtain a diffraction pattern suitable for the

structure resolution. Because of relatively small size, micropeptides often do not form stable and well-ordered crystals suitable for obtaining high-resolution diffraction patterns. Also, most micropeptides do not have their own function but act as modulators of protein partners and structural units of protein complexes, so the crystal structure of the peptide itself is often uninformative. However, X-ray crystallography can be used to determine the structure of peptide complexes with protein partners, for example, the structure of phospholamban complex with SERCA [91].

The presence of a stable secondary structure can indicate the micropeptide stability in the cells and its biological function. A relatively simple method for determining the secondary structure of proteins is circular dichroism (CD) spectroscopy, which is based on the differential absorption of circularly polarized light by chiral molecules. Proteins, which are composed of chiral amino acids, are optically active and exhibit characteristic CD spectra. While this method does not directly determine the tertiary structure, it is sensitive to the presence of certain secondary structure elements in the protein and can be used to elucidate the types of secondary structures present in a micropeptide [75]. This approach was used to describe the interaction of the CYREN peptide with the Ku70/Ku80 protein complex [14].

A more informative method for resolving the tertiary structure of peptides is nuclear magnetic resonance (NMR). Most micropeptides have small molecular weights (up to 11 kDa) and, therefore, are good subjects for the structure determination by NMR [92]. This method was used for the characterization of the DWORF structure and revealed a proline-induced bend necessary for the activation of SERCA [19].

CONCLUSION

The studies of micropeptides encoded by sORFs expand our understanding of the genome coding potential and complexity of the proteome. Over the past 10-15 years, a large body of evidence has been accumulated regarding the functional role of micropeptides. As a result, sORFs are no longer considered as non-functional sequences, as it has become clear that they provide a new level in the regulation of cellular processes. This paradigm shift has become possible due to the prior development of molecular biology methods. Ribosome profiling has revealed the scale of sORF translation, while improvement in sample preparation for mass spectrometry allowed direct detection of their peptide products, and CRISPR/Cas9-based genome editing provided a powerful tool for the functional studies.

Despite a significant progress, detection of micropeptides is still associated with certain challenges. It is important to reliably differentiate sORFs that encode micropeptides from those translated but non-functional or those that act as translational regulatory elements, which may require a deeper investigation of their evolutionary origin and analysis by molecular biology methods. The confirmation of the existence of micropeptides and characterization of their properties remain a non-trivial task due to small size of these molecules and limitations it imposes on the use of classical molecular biology methods. Current strategies used for micropeptide discovery may need certain improvements. For example, the detection and quantification of short-lived micropeptides require further increase in the sensitivity of mass spectrometry methods, optimization of sample preparation protocols, and generation of sufficiently comprehensive but non-redundant reference databases.

Nevertheless, a substantial amount of data on micropeptides has already been accumulated. Future research will likely involve more comprehensive integration of genomics, transcriptomics, and proteomics data. This multi-omics approach will open new horizons in studying the role of micropeptides and their potential applications in disease diagnostics and therapy.

Abbreviations

Co-IP	co-immunoprecipitation
GFP	green fluorescent protein
Ribo-Seq	ribosome profiling
RNA-seq	RNA sequencing
sgRNA	single-guide RNA
sORF	small open reading frame

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1134/S0006297925602242.

Contributions

A.I.L. wrote the text of the article; O.A.D., M.P.R., and N.M.Sh. edited the manuscript.

Funding

This work was supported by the Russian Science Foundation (project no. 23-14-00058) and as part of the State Assignment to the Lomonosov Moscow State University (registration no. 121031300037-7).

Ethics approval and consent to participate

This work does not contain any studies involving human and animal subjects.

Conflict of interest

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. Basrai, M. A., Hieter, P., and Boeke, J. D. (1997) Small open reading frames: beautiful needles in the haystack, *Genome Res.*, **7**, 768-771, https://doi.org/10.1101/gr.7.8.768.
2. Harrison, P. M. (2002) A question of size: the eukaryotic proteome and the problems in defining it, *Nucleic Acids Res.*, **30**, 1083-1090, https://doi.org/10.1093/nar/30.5.1083.
3. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., and Weissman, J. S. (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling, *Science*, **324**, 218-223, https://doi.org/10.1126/science.1168978.
4. Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B. J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., Xenarios, I., Zhang, L., Trono, D., Harari, A., Ohler, U., Coukos, G., and Bassani-Sternberg, M. (2020) Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes, *Nat. Commun.*, **11**, 1293, https://doi.org/10.1038/s41467-020-14968-9.
5. Slavoff, S. A., Mitchell, A. J., Schwaid, A. G., Cabili, M. N., Ma, J., Levin, J. Z., Karger, A. D., Budnik, B. A., Rinn, J. L., and Saghatelian, A. (2013) Peptidomic discovery of short open reading frame-encoded peptides in human cells, *Nat. Chem. Biol.*, **9**, 59-64, https://doi.org/10.1038/nchembio.1120.
6. Sandmann, C.-L., Schulz, J. F., Ruiz-Orera, J., Kirchner, M., Ziehm, M., Adami, E., Marczenke, M., Christ, A., Liebe, N., Greiner, J., Schoenenberger, A., Muecke, M. B., Liang, N., Moritz, R. L., Sun, Z., Deutsch, E. W., Gotthardt, M., Mudge, J. M., Prensner, J. R., Willnow, T. E., Mertins, P., Van Heesch, S., and Hubner, N. (2023) Evolutionary origins and interactomes of human, young microproteins and small peptides translated from short open reading frames, *Mol. Cell*, **83**, 994-1011.e18, https://doi.org/10.1016/j.molcel.2023.01.023.
7. Cao, X., Sun, S., and Xing, J. (2024) A massive proteogenomic screen identifies thousands of novel peptides from the human "dark" proteome, *Mol. Cell. Proteomics*, **23**, 100719, https://doi.org/10.1016/j.mcpro.2024.100719.
8. Pauli, A., Norris, M. L., Valen, E., Chew, G.-L., Gagnon, J. A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D., Tsai, S. Q., Joung, J. K., Saghatelian, A., and Schier, A. F. (2014) Toddler: an embryonic signal that promotes cell movement via apelin receptors, *Science*, **343**, 1248636, https://doi.org/10.1126/science.1248636.
9. Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., and Kageyama, Y. (2007) Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA, *Nat. Cell. Biol.*, **9**, 660-665, https://doi.org/10.1038/ncb1595.
10. Chugunova, A., Loseva, E., Mazin, P., Mitina, A., Navalayeu, T., Bilan, D., Vishnyakova, P., Marey, M., Golovina, A., Serebryakova, M., Pletnev, P., Rubtsova, M., Mair, W., Vanyushkina, A., Khaitovich, P., Belousov, V., Vysokikh, M., Sergiev, P., and Dontsova, O. (2019) LINC00116 codes for a mitochondrial peptide linking respiration and lipid metabolism, *Proc. Natl. Acad. Sci. USA*, **116**, 4940-4945, https://doi.org/10.1073/pnas.1809105116.
11. Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., Saghatelian, A., Nakayama, K. I., Clohessy, J. G., and Pandolfi, P. P. (2017) mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide, *Nature*, **541**, 228-232, https://doi.org/10.1038/nature21034.
12. Dong, X., Zhang, K., Xun, C., Chu, T., Liang, S., Zeng, Y., and Liu, Z. (2023) Small Open reading frame-encoded micro-peptides: an emerging protein world, *Int. J. Mol. Sci.*, **24**, 10562, https://doi.org/10.3390/ijms241310562.
13. Couso, J.-P., and Patraquim, P. (2017) Classification and function of small open reading frames, *Nat. Rev. Mol. Cell Biol.*, **18**, 575-589, https://doi.org/10.1038/nrm.2017.58.
14. Xie, L., Bowman, M. E., Louie, G. V., Zhang, C., Ardejani, M. S., Huang, X., Chu, Q., Donaldson, C. J., Vaughan, J. M., Shan, H., Powers, E. T., Kelly, J. W., Lyumkis, D., Noel, J. P., and Saghatelian, A. (2023) Biochemistry and protein interactions of the CYREN microprotein, *Biochemistry*, **62**, 3050-3060, https://doi.org/10.1021/acs.biochem.3c00397.
15. Zhu, K.-G., Yang, J., Zhu, Y., Zhu, Q., Pan, W., Deng, S., He, Y., Zuo, D., Wang, P., Han, Y., and Zhang, H.-Y. (2023) The microprotein encoded by exosomal lncAKR1C2 promotes gastric cancer lymph node

metastasis by regulating fatty acid metabolism, *Cell Death Dis.*, **14**, 708, https://doi.org/10.1038/s41419-023-06220-1.
16. Zheng, W., Guo, Y., Zhang, G., Bai, J., Song, Y., Song, X., Zhu, Q., Bao, X., Wu, G., and Zhang, C. (2023) Peptide encoded by lncRNA BVES-AS1 promotes cell viability, migration, and invasion in colorectal cancer cells via the SRC/mTOR signaling pathway, *PLoS One*, **18**, e0287133, https://doi.org/10.1371/journal.pone.0287133.
17. Guo, B., Zhai, D., Cabezas, E., Welsh, K., Nouraini, S., Satterthwait, A. C., and Reed, J. C. (2003) Humanin peptide suppresses apoptosis by interfering with Bax activation, *Nature*, **423**, 456-461, https://doi.org/10.1038/nature01627.
18. Makarewich, C. A., Munir, A. Z., Schiattarella, G. G., Bezprozvannaya, S., Raguimova, O. N., Cho, E. E., Vidal, A. H., Robia, S. L., Bassel-Duby, R., and Olson, E. N. (2018) The DWORF micropeptide enhances contractility and prevents heart failure in a mouse model of dilated cardiomyopathy, *Elife*, **7**, e38319, https://doi.org/10.7554/eLife.38319.
19. Reddy, U. V., Weber, D. K., Wang, S., Larsen, E. K., Gopinath, T., De Simone, A., Robia, S., and Veglia, G. (2022) A kink in DWORF helical structure controls the activation of the sarcoplasmic reticulum Ca^{2+}-ATPase, *Structure*, **30**, 360-370, https://doi.org/10.1016/j.str.2021.11.003.
20. Tonkin, J., and Rosenthal, N. (2015) One small step for muscle: a new micropeptide regulates performance, *Cell Metab.*, **21**, 515-516, https://doi.org/10.1016/j.cmet.2015.03.013.
21. Anderson, D. M., Makarewich, C. A., Anderson, K. M., Shelton, J. M., Bezprozvannaya, S., Bassel-Duby, R., and Olson, E. N. (2016) Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides, *Sci. Signal.*, **9**, aaj1460, https://doi.org/10.1126/scisignal.aaj1460.
22. Frith, M. C., Forrest, A. R., Nourbakhsh, E., Pang, K. C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Bailey, T. L., and Grimmond, S. M. (2006) The abundance of short proteins in the mammalian proteome, *PLoS Genet.*, **2**, e52, https://doi.org/10.1371/journal.pgen.0020052.
23. Kute, P. M., Soukarieh, O., Tjeldnes, H., Trégouët, D.-A., and Valen, E. (2022) Small open reading frames, how to find them and determine their function, *Front. Genet.*, **12**, 796060, https://doi.org/10.3389/fgene.2021.796060.
24. Hurst, L. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution, *Trends Genet.*, **18**, 486-487, https://doi.org/10.1016/s0168-9525(02)02722-1.
25. Lin, M. F., Jungreis, I., and Kellis, M. (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions, *Bioinformatics*, **27**, i275-i282, https://doi.org/10.1093/bioinformatics/btr209.
26. Mackowiak, S. D., Zauber, H., Bielow, C., Thiel, D., Kutz, K., Calviello, L., Mastrobuoni, G., Rajewsky, N., Kempa, S., Selbach, M., and Obermayer, B. (2015) Extensive identification and analysis of conserved small ORFs in animals, *Genome Biol.*, **16**, 179, https://doi.org/10.1186/s13059-015-0742-x.
27. Ladoukakis, E., Pereira, V., Magny, E. G., Eyre-Walker, A., and Couso, J. P. (2011) Hundreds of putatively functional small open reading frames in *Drosophila*, *Genome Biol.*, **12**, R118, https://doi.org/10.1186/gb-2011-12-11-r118.
28. Badger, J. H., and Olsen, G. J. (1999) CRITICA: coding region identification tool invoking comparative analysis, *Mol. Biol. Evol.*, **16**, 512-524, https://doi.org/10.1093/oxfordjournals.molbev.a026133.
29. Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine, *Nucleic Acids Res.*, **35**, W345-W349, https://doi.org/10.1093/nar/gkm391.
30. Chen, Z., Meng, J., Zhao, S., Yin, C., and Luan, Y. (2023) sORFPred: a method based on comprehensive features and ensemble learning to predict the sORFs in plant LncRNAs, *Interdiscip. Sci. Comput. Life Sci.*, **15**, 189-201, https://doi.org/10.1007/s12539-023-00552-4.
31. Zhao, S., Meng, J., Kang, Q., and Luan, Y. (2022) Identifying LncRNA-encoded short peptides using optimized hybrid features and ensemble learning, *Trans. Comput. Biol. and Bioinf.*, **19**, 2873-2881, https://doi.org/10.1109/TCBB.2021.3104288.
32. Raj, A., Wang, S. H., Shim, H., Harpak, A., Li, Y. I., Engelmann, B., Stephens, M., Gilad, Y., and Pritchard, J. K. (2016) Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling, *Elife*, **5**, e13328, https://doi.org/10.7554/eLife.13328.
33. Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., Vejnar, C. E., Lee, M. T., Rajewsky, N., Walther, T. C., and Giraldez, A. J. (2014) Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation, *EMBO J.*, **33**, 981-993, https://doi.org/10.1002/embj.201488411.
34. Schneider-Poetsch, T., Ju, J., Eyler, D. E., Dang, Y., Bhat, S., Merrick, W. C., Green, R., Shen, B., and Liu, J. O. (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin, *Nat. Chem. Biol.*, **6**, 209-217, https://doi.org/10.1038/nchembio.304.
35. Brar, G. A., and Weissman, J. S. (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis, *Nat. Rev. Mol. Cell. Biol.*, **16**, 651-664, https://doi.org/10.1038/nrm4069.
36. Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J. S., Jackson, S. E., Wills, M. R.,

and Weissman, J. S. (2014) Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes, *Cell Rep.*, **8**, 1365-1379, https://doi.org/10.1016/j.celrep.2014.07.045.

37. Bartholomäus, A., Del Campo, C., and Ignatova, Z. (2016) Mapping the non-standardized biases of ribosome profiling, *Biol. Chem.*, **397**, 23-35, https://doi.org/10.1515/hsz-2015-0197.
38. Martinez, T. F., Chu, Q., Donaldson, C., Tan, D., Shokhirev, M. N., and Saghatelian, A. (2020) Accurate annotation of human protein-coding small open reading frames, *Nat. Chem. Biol.*, **16**, 458-468, https://doi.org/10.1038/s41589-019-0425-0.
39. Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes, *Cell*, **147**, 789-802, https://doi.org/10.1016/j.cell.2011.10.002.
40. Fresno, M., Jiménez, A., and Vázquez, D. (1977) Inhibition of translation in eukaryotic systems by harringtonine, *Eur. J. Biochem.*, **72**, 323-330, https://doi.org/10.1111/j.1432-1033.1977.tb11256.x.
41. Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., and Qian, S.-B. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution, *Proc. Natl. Acad. Sci. USA*, **109**, E2424-E2432, https://doi.org/10.1073/pnas.1207846109.
42. Gao, X., Wan, J., Liu, B., Ma, M., Shen, B., and Qian, S.-B. (2015) Quantitative profiling of initiating ribosomes *in vivo*, *Nat. Methods*, **12**, 147-153, https://doi.org/10.1038/nmeth.3208.
43. Mudge, J. M., Ruiz-Orera, J., Prensner, J. R., Brunet, M. A., Calvet, F., Jungreis, I., Gonzalez, J. M., Magrane, M., Martinez, T. F., Schulz, J. F., Yang, Y. T., Albà, M. M., Aspden, J. L., Baranov, P. V., Bazzini, A. A., Bruford, E., Martin, M. J., Calviello, L., Carvunis, A.-R., Chen, J., Couso, J. P., Deutsch, E. W., Flicek, P., Frankish, A., Gerstein, M., Hubner, N., Ingolia, N. T., Kellis, M., Menschaert, G., Moritz, R. L., Ohler, U., Roucou, X., Saghatelian, A., Weissman, J. S., and Van Heesch, S. (2022) Standardized annotation of translated open reading frames, *Nat. Biotechnol.*, **40**, 994-999, https://doi.org/10.1038/s41587-022-01369-0.
44. Spealman, P., Naik, A. W., May, G. E., Kuersten, S., Freeberg, L., Murphy, R. F., and McManus, J. (2018) Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data, *Genome Res.*, **28**, 214-222, https://doi.org/10.1101/gr.221507.117.
45. Quaife, N. M., Chothani, S., Schulz, J. F., Lindberg, E. L., Vanezis, K., Adami, E., O'Fee, K., Greiner, J., Litviňuková, M., Van Heesch, S., Whiffin, N., Hubner, N., Schafer, S., Rackham, O., Cook, S. A., and Barton, P. J. R. (2023) LINC01013 is a determinant of fibroblast activation and encodes a novel fibroblast-activating micropeptide, *J. Cardiovasc. Trans. Res.*, **16**, 77-85, https://doi.org/10.1007/s12265-022-10288-z.
46. Sun, L., Wang, W., Han, C., Huang, W., Sun, Y., Fang, K., Zeng, Z., Yang, Q., Pan, Q., Chen, T., Luo, X., and Chen, Y. (2021) The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation, *Mol. Cell*, **81**, 4493-4508.e9, https://doi.org/10.1016/j.molcel.2021.08.033.
47. Van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J. F., Adami, E., Faber, A. B., Kirchner, M., Maatz, H., Hubner, N., et al. (2019) The translational landscape of the human heart, *Cell*, **178**, 242-260.e29, https://doi.org/10.1016/j.cell.2019.05.010.10.
48. Wang, S., Tian, L., Liu, H., Li, X., Zhang, J., Chen, X., Jia, X., Zheng, X., Wu, S., Chen, Y., Yan, J., and Wu, L. (2020) Large-scale discovery of non-conventional peptides in maize and arabidopsis through an integrated peptidogenomic pipeline, *Mol. Plant*, **13**, 1078-1093, https://doi.org/10.1016/j.molp.2020.05.012.
49. Cardon, T., Hervé, F., Delcourt, V., Roucou, X., Salzet, M., Franck, J., and Fournier, I. (2020) Optimized sample preparation workflow for improved identification of ghost proteins, *Anal. Chem.*, **92**, 1122-1129, https://doi.org/10.1021/acs.analchem.9b04188.
50. Ma, J., Ward, C. C., Jungreis, I., Slavoff, S. A., Schwaid, A. G., Neveu, J., Budnik, B. A., Kellis, M., and Saghatelian, A. (2014) Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue, *J. Proteome Res.*, **13**, 1757-1765, https://doi.org/10.1021/pr401280w.
51. Ma, J., Diedrich, J. K., Jungreis, I., Donaldson, C., Vaughan, J., Kellis, M., Yates, J. R., and Saghatelian, A. (2016) Improved identification and analysis of small open reading frame encoded polypeptides, *Anal. Chem.*, **88**, 3967-3975, https://doi.org/10.1021/acs.analchem.6b00191.
52. Cassidy, L., Helbig, A. O., Kaulich, P. T., Weidenbach, K., Schmitz, R. A., and 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. Proteomics*, **230**, 103988, https://doi.org/10.1016/j.jprot.2020.103988.
53. Tran, J. C., and Doucette, A. A. (2008) Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation, *Anal. Chem.*, **80**, 1568-1573, https://doi.org/10.1021/ac702197w.
54. Branca, R. M. M., Orre, L. M., Johansson, H. J., Granholm, V., Huss, M., Pérez-Bercoff, Å., Forshed, J., Käll, L., and Lehtiö, J. (2014) HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics, *Nat. Methods*, **11**, 59-62, https://doi.org/10.1038/nmeth.2732.
55. Nesvizhskii, A. I. (2014) Proteogenomics: concepts, applications and computational strategies, *Nat. Methods*, **11**, 1114-1125, https://doi.org/10.1038/nmeth.3144.

56. Fijalkowski, I., Peeters, M. K. R., and Van Damme, P. (2021) Small protein enrichment improves proteomics detection of sORF encoded polypeptides, *Front. Genet.*, **12**, 713400, https://doi.org/10.3389/fgene.2021.713400.
57. Cassidy, L., Kaulich, P. T., and Tholey, A. (2023) Proteoforms expand the world of microproteins and short open reading frame-encoded peptides, *iScience*, **26**, 106069, https://doi.org/10.1016/j.isci.2023.106069.
58. Khitun, A., and Slavoff, S. A. (2019) Proteomic detection and validation of translated small open reading frames, *Curr. Protoc. Chem. Biol.*, **11**, e77, https://doi.org/10.1002/cpch.7.
59. Slavoff, S. A., Heo, J., Budnik, B. A., Hanakahi, L. A., and Saghatelian, A. (2014) A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining, *J. Biol. Chem.*, **289**, 10950-10957, https://doi.org/10.1074/jbc.C113.533968.
60. Ouspenskaia, T., Law, T., Clauser, K. R., Klaeger, S., Sarkizova, S., Aguet, F., Li, B., Christian, E., Knisbacher, B. A., Le, P. M., Hartigan, C. R., Keshishian, H., Apffel, A., Oliveira, G., Zhang, W., Chen, S., Chow, Y. T., Ji, Z., Jungreis, I., Shukla, S. A., Justesen, S., Bachireddy, P., Kellis, M., Getz, G., Hacohen, N., Keskin, D. B., Carr, S. A., Wu, C. J., and Regev, A. (2022) Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer, *Nat. Biotechnol*, **40**, 209-217, https://doi.org/10.1038/s41587-021-01021-3.
61. Wang, J. Y., and Doudna, J. A. (2023) CRISPR technology: a decade of genome editing is only the beginning, *Science*, **379**, eadd8643, https://doi.org/10.1126/science.add8643.
62. Shalem, O., Sanjana, N. E., and Zhang, F. (2015) High-throughput functional genomics using CRISPR-Cas9, *Nat. Rev. Genet.*, **16**, 299-311, https://doi.org/10.1038/nrg3899.
63. Chen, J., Brunner, A.-D., Cogan, J. Z., Nuñez, J. K., Fields, A. P., Adamson, B., Itzhak, D. N., Li, J. Y., Mann, M., Leonetti, M. D., and Weissman, J. S. (2020) Pervasive functional translation of noncanonical human open reading frames, *Science*, **367**, 1140-1146, https://doi.org/10.1126/science.aay0262.
64. Prensner, J. R., Enache, O. M., Luria, V., Krug, K., Clauser, K. R., Dempster, J. M., Karger, A., Wang, L., Stumbraite, K., Wang, V. M., Botta, G., Lyons, N. J., Goodale, A., Kalani, Z., Fritchman, B., Brown, A., Alan, D., Green, T., Yang, X., Jaffe, J. D., Roth, J. A., Piccioni, F., Kirschner, M. W., Ji, Z., Root, D. E., and Golub, T. R. (2021) Noncanonical open reading frames encode functional proteins essential for cancer cell survival, *Nat. Biotechnol.*, **39**, 697-704, https://doi.org/10.1038/s41587-020-00806-2.
65. Schlesinger, D., Dirks, C., Navarro, C., Lafranchi, L., Spinner, A., Raja, G. L., Mun-Sum Tong, G., Eirich, J., Martinez, T. F., and Elsässer, S. J. (2025) A large-scale sORF screen identifies putative microproteins involved in cancer cell fitness, *iScience*, **28**, 111884, https://doi.org/10.1016/j.isci.2025.111884.
66. Lafranchi, L., Spinner, A., Hornisch, M., Schlesinger, D., Luzon, C. N., Brinkenstråhle, L., Shao, R., Piazza, I., and Elsässer, S. J. (2024) Pooled overexpression screening identifies PIPPI as a novel microprotein involved in the ER stress response, *bioRxiv*, https://doi.org/10.1101/2024.12.08.627409.
67. Delcourt, V., Brunelle, M., Roy, A. V., Jacques, J.-F., Salzet, M., Fournier, I., and Roucou, X. (2018) The protein coded by a short open reading frame, not by the annotated coding sequence, is the main gene product of the dual-coding gene MIEF1, *Mol. Cell. Proteom.*, **17**, 2402-2411, https://doi.org/10.1074/mcp.RA118.000593.
68. Zhang, C., Zhou, B., Gu, F., Liu, H., Wu, H., Yao, F., Zheng, H., Fu, H., Chong, W., Cai, S., Huang, M., Ma, X., Guo, Z., Li, T., Deng, W., Zheng, M., Ji, Q., Zhao, Y., Ma, Y., Wang, Q.-E., Tang, T.-S., and Guo, C. (2022) Micropeptide PACMP inhibition elicits synthetic lethal effects by decreasing CtIP and poly(ADP-ribosyl)ation, *Mol. Cell*, **82**, 1297-1312.e8, https://doi.org/10.1016/j.molcel.2022.01.020.
69. Zhang, H., Liao, Z., Wang, W., Liu, Y., Zhu, H., Liang, H., Zhang, B., and Chen, X. (2023) A micropeptide JunBP regulated by TGF-β promotes hepatocellular carcinoma metastasis, *Oncogene*, **42**, 113-123, https://doi.org/10.1038/s41388-022-02518-0.
70. Guo, B., Wu, S., Zhu, X., Zhang, L., Deng, J., Li, F., Wang, Y., Zhang, S., Wu, R., Lu, J., and Zhou, Y. (2020) Micropeptide CIP 2A-BP encoded by LINC 00665 inhibits triple-negative breast cancer progression, *EMBO J.*, **39**, e102190, https://doi.org/10.15252/embj.2019102190.
71. Huang, J.-Z., Chen, M., Chen, D., Gao, X.-C., Zhu, S., Huang, H., Hu, M., Zhu, H., and Yan, G.-R. (2017) A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth, *Mol. Cell*, **68**, 171-184.e6, https://doi.org/10.1016/j.molcel.2017.09.015.
72. Rocha, A. L., Pai, V., Perkins, G., Chang, T., Ma, J., De Souza, E. V., Chu, Q., Vaughan, J. M., Diedrich, J. K., Ellisman, M. H., and Saghatelian, A. (2024) An inner mitochondrial membrane microprotein from the SLC35A4 upstream ORF regulates cellular metabolism, *J. Mol. Biol.*, **436**, 168559, https://doi.org/10.1016/j.jmb.2024.168559.
73. Konina, D., Sparber, P., Viakhireva, I., Filatova, A., and Skoblov, M. (2021) Investigation of LINC00493/SMIM26 gene suggests its dual functioning at mRNA and protein level, *Int. J. Mol. Sci. Artic. J. Mol. Sci.*, **22**, 8477, https://doi.org/10.3390/ijms22168477.
74. Liang, X., Shen, W., Sun, H., Migawa, M. T., Vickers, T. A., and Crooke, S. T. (2016) Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames, *Nat. Biotechnol.*, **34**, 875-880, https://doi.org/10.1038/nbt.3589.

75. Anderson, D. M., Anderson, K. M., Chang, C.-L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., Kasaragod, P., Shelton, J. M., Liou, J., Bassel-Duby, R., and Olson, E. N. (2015) A micropeptide encoded by a putative long noncoding RNA regulates muscle performance, *Cell*, **160**, 595-606, https://doi.org/10.1016/j.cell.2015.01.009.
76. Ryskina, A. M., Kudriaeva, A. A., and Belogurov, A. A. (2024) Microproteins tracking: when size does really matter, *Rev. Adv. Chem.*, **14**, 305-319, https://doi.org/10.1134/S2634827624600324.
77. Chu, Q., Martinez, T. F., Novak, S. W., Donaldson, C. J., Tan, D., Vaughan, J. M., Chang, T., Diedrich, J. K., Andrade, L., Kim, A., Zhang, T., Manor, U., and Saghatelian, A. (2019) Regulation of the ER stress response by a mitochondrial microprotein, *Nat. Commun.*, **10**, 12816, https://doi.org/10.1038/s41467-019-12816-z.
78. Niu, L., Lou, F., Sun, Y., Sun, L., Cai, X., Liu, Z., Zhou, H., Wang, H., Wang, Z., Bai, J., Yin, Q., Zhang, J., Chen, L., Peng, D., Xu, Z., Gao, Y., Tang, S., Fan, L., and Wang, H. (2020) A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation, *Sci. Adv.*, **6**, eaaz2059, https://doi.org/10.1126/sciadv.aaz2059.
79. Rathore, A., Chu, Q., Tan, D., Martinez, T. F., Donaldson, C. J., Diedrich, J. K., Yates, J. R., and Saghatelian, A. (2018) MIEF1 microprotein regulates mitochondrial translation, *Biochemistry*, **57**, 5564-5575, https://doi.org/10.1021/acs.biochem.8b00726.
80. Schägger, H. (2006) Tricine-SDS-PAGE, *Nat. Protoc.*, **1**, 16-22, https://doi.org/10.1038/nprot.2006.4.
81. Stein, C. S., Jadiya, P., Zhang, X., McLendon, J. M., Abouassaly, G. M., Witmer, N. H., Anderson, E. J., Elrod, J. W., and Boudreau, R. L. (2018) Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency, *Cell Rep.*, **23**, 3710-3720.e8, https://doi.org/10.1016/j.celrep.2018.06.002.
82. Okita, N., Higami, Y., Fukai, F., Kobayashi, M., Mitarai, M., Sekiya, T., and Sasaki, T. (2017) Modified Western blotting for insulin and other diabetes-associated peptide hormones, *Sci. Rep.*, **7**, 6949, https://doi.org/10.1038/s41598-017-04456-4.
83. Huang, N., Li, F., Zhang, M., Zhou, H., Chen, Z., Ma, X., Yang, L., Wu, X., Zhong, J., Xiao, F., Yang, X., Zhao, K., Li, X., Xia, X., Liu, Z., Gao, S., and Zhang, N. (2021) An Upstream Open Reading Frame in Phosphatase and Tensin Homolog Encodes a Circuit Breaker of Lactate Metabolism, *Cell Metab.*, **33**, 128-144.e9, https://doi.org/10.1016/j.cmet.2020.12.008.
84. Cabantous, S., Terwilliger, T. C., and Waldo, G. S. (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein, *Nat. Biotechnol.*, **23**, 102-107, https://doi.org/10.1038/nbt1044.
85. Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L. A., Ishikawa, H., Leonetti, M. D., Marshall, W. F., Weissman, J. S., and Huang, B. (2016) Versatile protein tagging in cells with split fluorescent protein, *Mol. Cell*, **82**, 2900-2911.e7, https://doi.org/10.1038/ncomms11046.
86. Na, Z., Dai, X., Zheng, S.-J., Bryant, C. J., Loh, K. H., Su, H., Luo, Y., Buhagiar, A. F., Cao, X., Baserga, S. J., Chen, S., and Slavoff, S. A. (2022) Mapping subcellular localizations of unannotated microproteins and alternative proteins with MicroID, *Mol. Cell*, **82**, 2900-2911.e7, https://doi.org/10.1016/j.molcel.2022.06.035.
87. Nguyen, T. M. T., Kim, J., Doan, T. T., Lee, M.-W., and Lee, M. (2020) APEX proximity labeling as a versatile tool for biological research, *Biochemistry*, **59**, 260-269, https://doi.org/10.1021/acs.biochem.9b00791.
88. Roux, K. J., Kim, D. I., Burke, B., and May, D. G. (2018) BioID: A Screen for Protein-Protein Interactions, *Curr. Protoc. Protein Sci.*, **91**, 19.23.1-19.23.15, https://doi.org/10.1002/cpps.51.
89. Bosch, J. A., Chen, C., and Perrimon, N. (2021) Proximity-dependent labeling methods for proteomic profiling in living cells: an update, *Wiley Interdiscip. Rev. Dev. Biol.*, **10**, e392, https://doi.org/10.1002/wdev.392.
90. Chu, Q., Rathore, A., Diedrich, J. K., Donaldson, C. J., Yates, J. R., and Saghatelian, A. (2017) Identification of microprotein-protein interactions via APEX tagging, *Biochemistry*, **56**, 3299-3306, https://doi.org/10.1021/acs.biochem.7b00265.
91. Singh, D. R., Dalton, M. P., Cho, E. E., Pribadi, M. P., Zak, T. J., Šeflová, J., Makarewich, C. A., Olson, E. N., and Robia, S. L. (2019) Newly discovered micropeptide regulators of SERCA form oligomers but bind to the pump as monomers, *J. Mol. Biol.*, **431**, 4429-4443, https://doi.org/10.1016/j.jmb.2019.07.037.
92. Hu, Y., Cheng, K., He, L., Zhang, X., Jiang, B., Jiang, L., Li, C., Wang, G., Yang, Y., and Liu, M. (2021) NMR-based methods for protein analysis, *Anal. Chem.*, **93**, 1866-1879, https://doi.org/10.1021/acs.analchem.0c03830.