Zhang *et al. Molecular Cancer* (2025) 24:105
https://doi.org/10.1186/s12943-025-02278-x

Molecular Cancer

REVIEW Open Access

Small open reading frame-encoded microproteins in cancer: identification, biological functions and clinical significance

Tingting Zhang[1†], Zhang Li[1†], Jiao Li[1*] and Yong Peng[1*]

Abstract

The human genome harbors approximately twenty thousand protein-coding genes, and a significant portion of life science research focuses on elucidating their functions and the underlying mechanisms. Recent studies have revealed that small open reading frame (sORF), originating from non-coding RNAs or the 5′ leader sequences of messenger RNAs, can be translated into small peptides called microproteins through cap-dependent or cap-independent mechanisms. These microproteins interact with diverse molecular partners to modulate gene expression at multiple regulatory levels, thereby playing critical roles in various biological processes. Notably, sORF-encoded microproteins exhibit aberrant expression patterns in cancer and are implicated in tumor initiation and progression, expanding our understanding of cancer biology. In this review, we introduce the translational mechanisms and identification methods of microproteins, summarize their dysregulation in cancer and their biological functions in regulating gene expression, and emphasize their roles in driving hallmark events of cancer. Furthermore, we discuss their clinical significance as diagnostic and prognostic biomarkers, as well as therapeutic targets.

Introduction

The majority of human genome sequences can be transcribed into RNA molecules. However, only approximately 3% of them are capable of encoding proteins, indicating most transcripts belong to non-coding RNAs (ncRNAs) [1]. Compelling evidence demonstrate that ncRNAs function as dynamic regulators to play crucial roles in different cellular processes, and their dysregulation is associated with disease pathogenesis [2, 3]. With the advancement in bioinformatics analysis and experimental technologies, an increasing number of ncRNAs with small open reading frame (sORF) have been discovered to encode small peptides called microproteins, typically fewer than 100 amino acids [4]. Besides ncRNAs, including long non-coding RNA (lncRNA) [5], circular RNA (circRNA) [6] and primary microRNA (pri-miRNA) [7], the 5′ leader sequences of messenger RNAs (mRNAs) [8] have also been identified as the sources of microprotein.

The sORF-encoded microproteins are conserved among species and exhibit tissue- or development-specific expression patterns. Growing evidence indicate that sORF-encoded microproteins play important biological functions under physiological conditions, such as signal transduction and intercellular communication during

†Tingting Zhang and Zhang Li contributed equally to this work.

*Correspondence:
Jiao Li
lijiaospace@163.com
Yong Peng
yongpeng@scu.edu.cn
[1]Center for Molecular Oncology, Frontiers Science Center for Disease-related Molecular Network, State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, Chengdu, China

embryonic development [9, 10], regulation of autoimmune responses and control of muscle performance [11, 12]. Notably, aberrant expression of sORF-encoded microproteins has been observed in human diseases, such as cardiovascular diseases, muscular dystrophy and cancer [13–18]. In cancer progression, dysregulated microproteins may play oncogenic or tumor suppressive roles, highlighting their potential to serve as diagnostic and prognostic biomarkers and promising therapeutic targets for cancer management [19, 20].

In this review, we briefly introduce the translational mechanisms and identification methods of sORF-encoded microproteins, summarize how microproteins are dysregulated in cancer and regulate gene expression, emphasize their biological functions during tumorigenesis, and finally highlight their clinical significance.

Translational mechanisms of sORF-encoded microprotein

Cap-dependent translation

Eukaryotic mRNAs are linear RNA molecules usually with a 7-methylguanosine (m7G) cap at the 5′ end and a poly(A) tail at the 3' end. Typically, mRNAs are translated into proteins in a cap-dependent fashion. In this process, the small ribosomal subunit (40 S) first associates with the eIF4F complex, which includes the translation initiation factor eIF4E, to recognize the 5′ cap, followed by ribosome scanning along the mRNA to locate the first start codon AUG [21, 22]. Upon recognizing the start codon, the 40 S subunit recruits the large ribosomal subunit (60 S) to form the 80 S translating ribosome for polypeptide synthesis [23].

Most lncRNAs have a 5′ cap structure and a 3′ poly(A) tail, referred to as 'mRNA-like' ncRNA [24–26], indicating that the sORF-encoded microproteins from lncRNA are translated in the cap-dependent manner (Fig. 1A). Employing the RiboTaper algorithm [27], it was found that although 96% of these mRNA-like ncRNAs can interact with the translational machinery, about 1% have the potential to produce proteins [24].

Leaky scanning serves as an alternative translation mechanism similar to the traditional cap-dependent translation. In leaky scanning, the 40S ribosomal subunit recognize the start codon of upstream open reading frame (uORF) located within the 5'-untranslated region (UTR) of mRNA in a cap-dependent manner, where it recruits the 60 S ribosomal subunit and translation initiation factors to trigger the translation of the uORF into a microprotein (Fig. 1B) [28]. Alternatively, the 40 S subunit can bypass upstream start codons and continue to scan downstream until it reaches the main ORF (mORF), resulting in the initiation of mORF translation [29, 30].

The proteins from these two translation modes may have distinct localizations. For example, the mORF of *SLC35A4* mRNA encodes a 324-amino-acid protein localized to the Golgi apparatus, while the uORF of the same mRNA is translated into a 103-amino-acid microprotein localized to the mitochondrial inner membrane, known as SLC35A4-MP [31, 32]. Typically, the translation of uORF can be initiated by either AUG or non-AUG codons. For example, Na et al. showed that the translation of uORF within *LAMA3* mRNA begins at non-AUG codon AGG, producing the alt-LAMA3 protein [33].

IRES-dependent translation

Due to the lack of 5′ cap structure and 3′ poly(A) tail, circRNAs and certain lncRNAs are translated through a cap-independent mechanism. Internal ribosome entry sites (IRESs), typically located upstream of their corresponding ORF, are highly structured *cis*-acting RNA elements that enable the recruitment of ribosomes to or near start codon, thereby facilitating protein translation through a cap-independent pathway [34]. Increasing studies have reported that circRNAs and certain lncRNAs are translated via IRES-dependent translation (Fig. 1C). For example, Xiao's group reported that circPDHK1 had an ORF with an AUG start codon and an IRES sequence with ~150 nucleotides in length, and experimentally validated that circPDHK1 promoted tumor growth and metastasis in clear cell renal cell carcinoma (ccRCC) through encoding a functional peptide termed PDHK1-241aa [35]. Yu et al. proved that DNA damage facilitated the association of ribosome with the IRES region of lncRNA CTBP1-DT, thus enabling the translation into the novel DDUP protein with 186 amino acid residues [36]. Noteworthy, the activities of IRESs often require assistance from other factors known as IRES-transacting factors (ITAFs). In melanoma cells, heterogeneous nuclear ribonucleoprotein A1 (hnRNP-A1) has been reported to promote the IRES-dependent translation of the lncRNA *meloe* to produce the microprotein MELOE-1, a melanoma-specific neoantigen [37].

m^{6}A-dependent translation

N6-methyladenosine (m^{6}A) is a type of RNA modification in which the N6 position of adenosine (A) within RNA molecules is methylated [38, 39]. m^{6}A modification is dynamically regulated by methyltransferases called writers and demethylases termed as erasers, and this modification plays important roles in protein translation of sORF (Fig. 1D). For example, Zheng et al. found that overexpression of methyltransferases METTL14 and METTL16 significantly increased the m^{6}A levels of circMIB2 and promoted the expression of its encoded microprotein MIB2-134aa [40]. In contrast, knockdown of the demethylase ALKBH5 dramatically increased LINC00278 m^{6}A level and LINC00278-encoded 21-amino-acid microprotein [41].

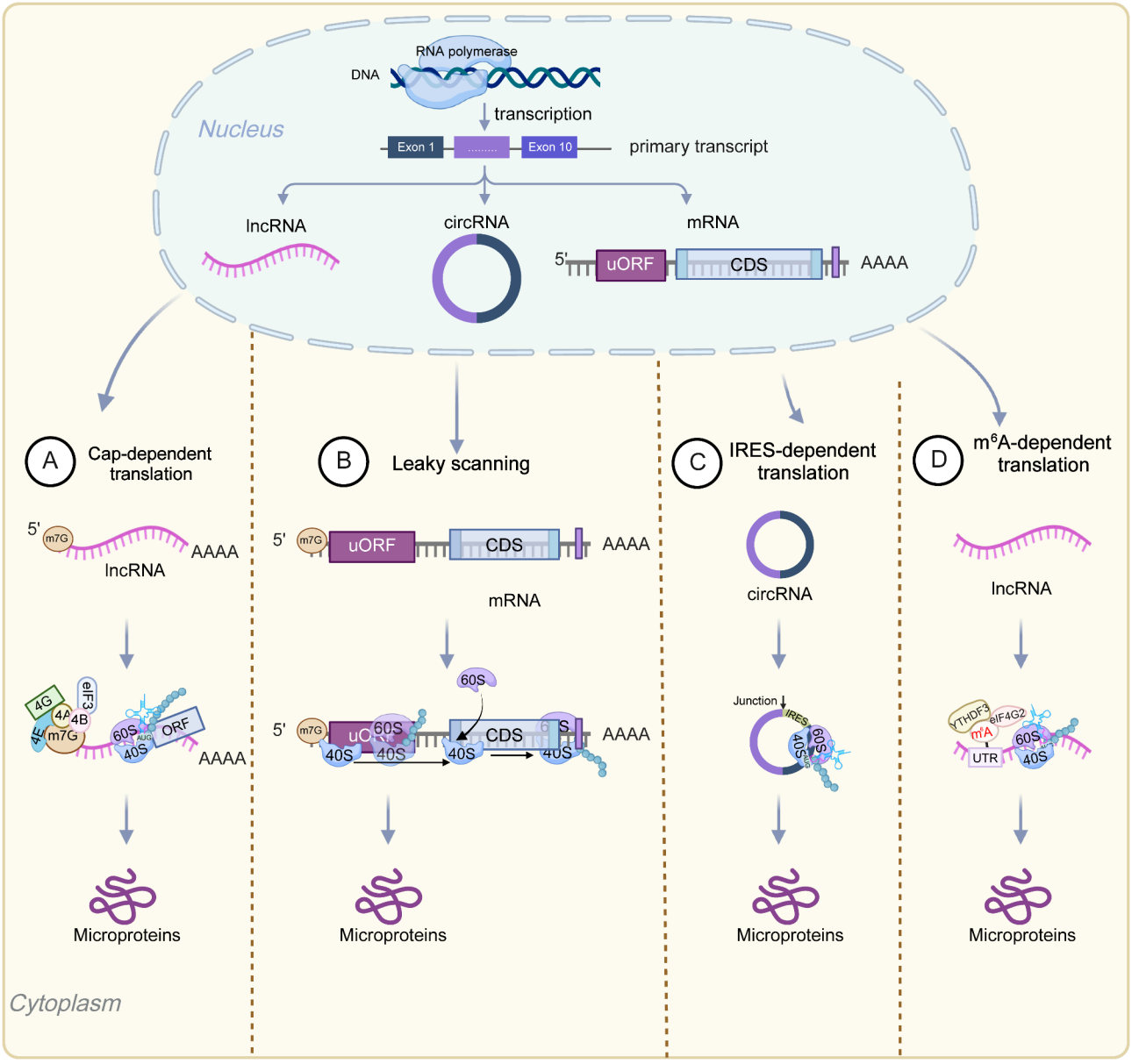

Fig. 1 Translational mechanisms of microproteins. **(A)** Some lncRNAs with a 5′ cap structure and a 3′ polyA tail are translated into microproteins in a cap-dependent manner; **(B)** The 5′UTR of mRNAs is translated by leaky scanning; **(C)** IRES-mediated transaltion; **(D)** m^{6}A modification-mediated transaltion

m^{6}A modification may be specifically recognized and bound by certain RNA-binding proteins called readers, such as IGF2BP1 and YTHDF1/2/3, to recruit translation machinery to facilitate the translation of sORFs [16, 42–44]. For example, Zeng et al. reported that YTHDF3 bound to the m^{6}A sites of circ-YAP and recruited eIF4G2 translation initiation complex, thus facilitating ribosome assembly to produce the protein YAP-220aa [44]. Besides, YTHDF2 is reported to recognize the m^{6}A modification of circMET and regulate the expression of circMET-encoded microprotein [16].

Interestingly, the lncRNA HNF4A-AS1 produces a small 51-amino acid peptide (sPEP1) using a mechanism different from either IRES- or m^{6}A-mediated translation. Song et al. revealed that miR-409-5p interacted with HNF4A-AS1 to facilitate sPEP1 translation through recruiting the translation initiation factor eIF3G [45]. However, it remains unclear whether it operates with the cap-dependent translation.

Identification methods of sORF-encoded microproteins

Bioinformatics tools for microprotein prediction

Bioinformatics tools are commonly employed to predict microproteins by analyzing genomic or transcriptomic data from databases. These tools identify microproteins

either by analyzing ORF or by coordinating analysis of m^{6}A modifications or IRES elements with the presence of ORF (Fig. 2A).

Currently, most tools predict ORFs by sequence alignments or alignment-free methods, as exemplified by ORF-FINDER, OrfPredictor, CPAT and GeneMarkS-T (Table 1). Among them, ORF-FINDER is the earliest developed tool for ORF searching in cDNA by sequence alignments [46]. Subsequently, Min et al. developed OrfPredictor to predict protein-coding regions in EST-derived sequences. OrfPredictor utilizes BLASTX to guide the identification of coding regions with a hit and predicts coding region ab initio for sequences without a hit [47]. With the advent of next-generation sequencing, the prediction of ORF in bulk transcriptomic data requires faster and more accurate analytical tools, boosting the development of machine learning-based ORF prediction techniques. Notably, Wang et al. introduced a supervised machine learning software called CPAT (Coding-Potential Assessment Tool), which utilizes a logistic regression model based on four sequence features: ORF length, ORF coverage, Fickett score and hexamer usage bias, thus enabling accurate identification of coding and noncoding transcripts from a pool of candidates [51]. Unlike CPAT, Borodovsky's team developed an unsupervised machine learning tool, GeneMarkS-T, which makes manually curated preparation of training sets unnecessary and is robust with respect to the presence of transcripts assembly errors [48]. Besides, sORF Finder and MiPepid are specifically designed to predict sORF (smaller than 300 base pairs). sORF Finder evaluates sORF coding potential based on the nucleotide composition bias between coding and non-coding sequences [52], while MiPepid, a machine-learning tool, demonstrates superior performance in evaluating the coding potential of the test set, achieving an overall accuracy of 0.96,

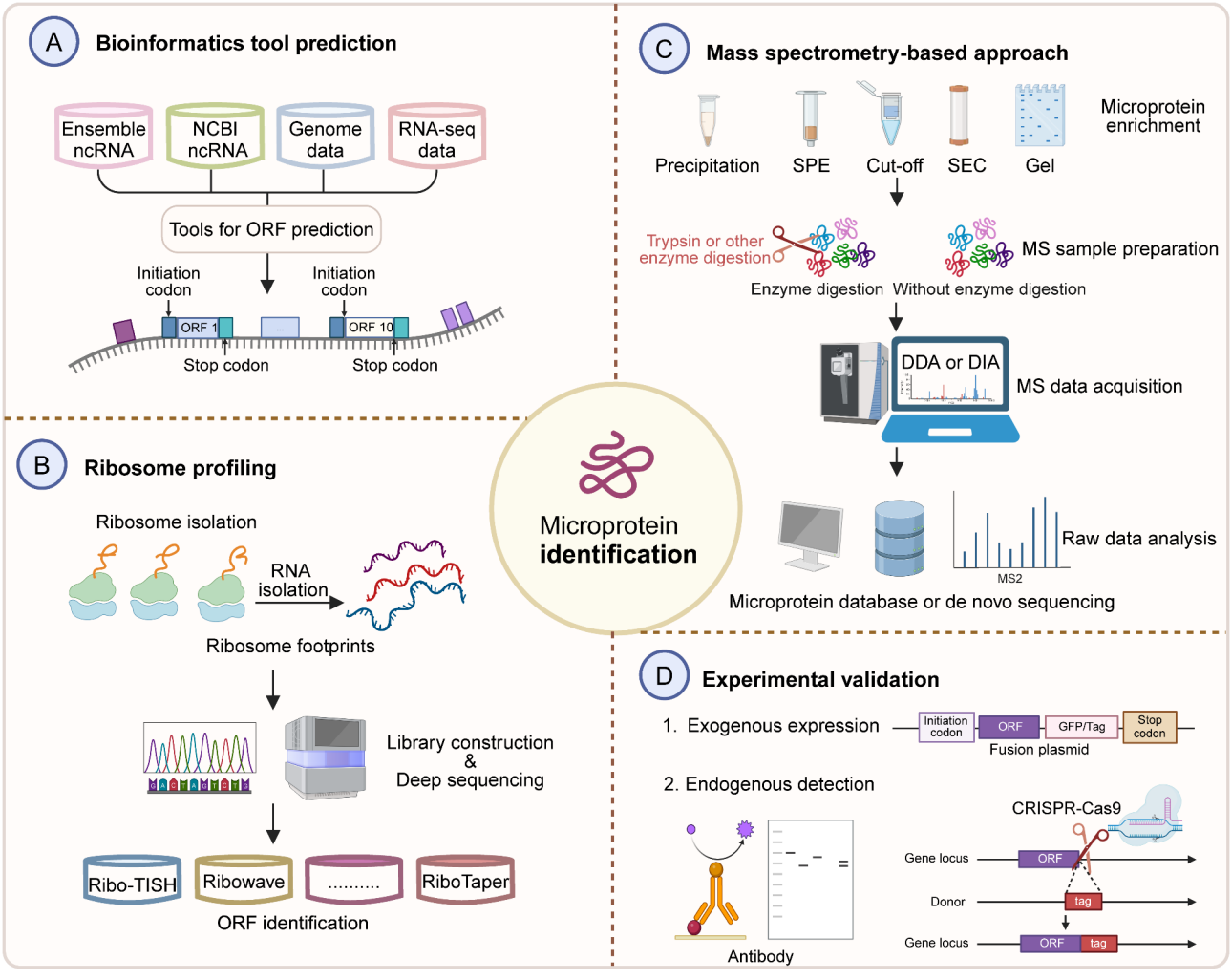

Fig. 2 Identification methods for microproteins. **(A)** Prediction of microproteins from transcriptomic and genomic data by bioinformatics tools; **(B)** Microprotein identification by ribo-seq; **(C)** Discovery of microproteins by MS; **(D)** Experimental validation of microproteins using exogenous expression and endogenous detection

Table 1 Tools for prediction of ORF, IRES and m^{6}A sites

name	advantage	disadvantage	website	ref.
ORF Finder	simple and user-friendly, suitable for beginners.	ack advanced analysis tools.	https://www.ncbi.nlm.nih.gov/orffinder/	[46]
ORF Predictor	provide detailed ORF information.	limited ability to predict non-standard start codons.	https://fungalgenome.concordia.ca/tools/OrfPredictor.html	[47]
GeneMarkS-T	specialize for transcriptome-based ORF prediction in eukaryotes and prokaryotes.	require high-quality RNA-seq data and less accurate for low-expression genes.	http://topaz.gatech.edu/GeneMark/license_download.cgi	[48]
PhyloCSF	base on evolutionary conservation, suitable for cross-species ORF prediction.	high computational resource requirements and less effective for non-conserved regions.	http://compbio.mit.edu/PhyloCSF/	[49]
PhyloCSF++	improve version of PhyloCSF with better performance and support large-scale data analysis.	require significant computational resources.	https://github.com/cpockrandt/PhyloCSFpp	[50]
CPAT	fast and accurate for coding potential assessment.	primarily suitable for known transcripts and less accurate for novel transcripts.	http://lilab.research.bcm.edu/cpat/index.php	[51]
sORF Finder	specialize for sORF prediction in multiple species.	require high-quality input data.	http://evolver.psc.riken.jp/	[52]
MiPepid	provide functional annotations.	limit scope, mainly for microbes and less effective for eukaryotes.	https://github.com/MindAI/MiPepid	[53]
uPEPperoni	specialize for uORF prediction.	limited ability to predict downstream ORFs.	http://upep-scmb.biosci.uq.edu.au.	[54]
IRESite	provide a comprehensive IRES database.	without prediction function.	http://www.iresite.org/	[55]
IRESfinder	specialize for IRES prediction in multiple species.	high computational resource requirements.	https://github.com/xiaofengsong/IRESfinder	[56]
VIPS	integrate multiple prediction tools.	less effective for complex genomes.	http://140.135.61.250/vips/	[57]
IRESPred	high accuracy for IRES prediction using machine learning and support multiple species.	require large training datasets.	http://bioinfo.net.in/IRESPred/	[58]
SRAMP	specialize for RNA modification site prediction and support multiple modification types.	mainly suitable for known modification types.	http://www.cuilab.cn/sramp/	[59]
M6AMRFS	high accuracy for m^{6}A modification prediction using multi-feature fusion.	high computational complexity.	http://server.malab.cn/M6AMRFS/	[60]
MethyRNA	provide comprehensive predictions for RNA methylation sites and their potential biological functions.	limited ability to predict complex modifications.	http://lin.uestc.edu.cn/server/methyrna	[61]
M6APred-EL	superior performance for m^{6}A prediction using ensemble learning and support multiple species.	require significant computational resources and less accurate for low-expression transcripts.	http://server.malab.cn/M6APred-EL/	[62]

compared to sORF Finder that attains an overall accuracy of 0.87 [53].

As mentioned above, IRES sites located upstream of the ORF in lncRNA and circRNA could recruit ribosomes to initiate the translation of downstream ORF. IRESite, IRESPred and IRESfinder are currently the most commonly used tools for IRES prediction. IRESite provides a collection of experimentally validated IRES sequences and related information, allowing users to query whether the RNA sequence of interest contains known IRES elements [55]. However, it cannot predict novel IRES elements for the subjected sequences. IRESPred predicts IRES elements via machine learning model, which was built based on sequence and secondary structure characteristics of UTRs and the probabilities of interactions between UTRs and small subunit ribosomal proteins [58]. However, IRESPred has some limitations, such as the use of unvalidated non-IRES sequences as negative samples, a small training dataset, and the inability to handle large-scale data. Subsequently, Song's team developed IRESfinder, which leverages 19 carefully selected k-mer features to construct a logistic regression model, achieving 80% precision and 73% accuracy in independent testing [56]. IRESfinder is also employed in CircBank database for IRES prediction within circRNA [63].

Various m^{6}A site prediction tools have been reported, including SRAMP [59], MethyRNA [61], M6AMRFS [60] and M6APred-EL [62]. These tools primarily predict m^{6}A sites based on RNA sequence features (i.e., nucleotide patterns, m^{6}A consensus motifs), RNA secondary structures and translation-related signals. These tools have been widely applied to the studies of microprotein-encoding transcripts. For instance, Duan et al. used SRAMP to show that circMAP3K4 has a high potential for m^{6}A modifications, which may be involved in circMAP3K4 translation [43].

Despite the development of these bioinformatics tools for microprotein prediction, the results may include false positives. To address this challenge, techniques such as ribosome profiling sequencing (Ribo-Seq), which

provides direct evidence of translation activity, and mass spectrometry (MS), which directly proves the presence of microproteins, can be combined with bioinformatics predictions to improve the accuracy and reliability of microprotein identification.

Ribo-Seq

Ribo-Seq is a high-throughput sequencing-based technology to study global translation activity within cells [64]. The experimental workflow includes cell lysis to release ribosomes and their bound mRNAs, followed by nuclease digestion of free mRNAs unprotected by ribosomes. Ribosomes and their protected mRNA fragments are then isolated using ultracentrifugation or affinity purification. The ribosome-protected fragments, typically 28–32 nucleotides in length, are extracted to construct a cDNA sequencing library for high-throughput sequencing. Subsequent data analysis involves mapping the short fragments to the reference genome to identify translation initiation and termination sites, translation efficiency and potential coding regions. This technique is widely used in exploring the translational potential of sORFs (Fig. 2B) [65].

To analyze Ribo-Seq results, various software tools have been developed to explore the translational potential of ORFs (Table 2). RiboTaper, RibORF and RiboCode detect actively translated regions by utilizing the triplet periodicity of ribosome-protected fragments (RPFs) [27, 67, 69]. These tools have proven instrumental in microprotein research. For example, van Heesch et al. employed bioinformatics tools including RiboTaper to identify 1,090 microproteins encoded by lncRNAs and circRNAs in human heart [80], while Jackson et al. used RibORF and RiboCode to identify 224 microproteins in mouse bone marrow-derived macrophages [81]. However, the output of these methods could be affected by low-quality or low-coverage data. To address these issues, the computational method called PRICE was developed, which employs the Expectation-Maximization (EM) algorithm to model experimental noise and accurately infer codon activity probabilities [70]. Additionally, tools like RiboToolkit and GWIPS-Viz enable online analysis of Ribo-Seq data [77, 78].

MS-based approaches

MS-based proteomics can provide direct and robust evidence to identify sORF-encoded microproteins by

Table 2 Tools for ORF prediction in ribo-seq data

name	advantage	disadvantage	website	ref.
RiboTaper	high accuracy in detecting translated ORFs.	require high-quality ribosome profiling data.	https://ohlerlab.mdc-berlin.de/software/RiboTaper_126	[27]
Ribotricer	specialize for identifying novel ORFs.	limit to specific organisms or datasets.	https://github.com/smithlabcode/ribotricer/releases/tag/v1.3.3	[66]
RiboCode	support multiple species and provide comprehensive ORF annotation.	require extensive computational resources.	https://github.com/xryanglab/RiboCode	[67]
SPECtre	focuse on detecting sORFs with high sensitivity.	produce false positives without proper filtering.	https://github.com/mills-lab/spectre	[68]
RiboORF	suitable for eukaryotic genomes.	require high-quality data and less effective for prokaryotic genomes.	https://github.com/zhejilab/RiboORF	[69]
PRICE	high precision in identifying coding sequences in ribosome profiling data.	require advanced bioinformatic skills.	https://github.com/erhard-lab/price	[70]
ORF-RATER	quantify translation efficiency and compatible with multiple ribosome profiling tools.	computationally demanding.	https://github.com/alexfields/ORF-RATER	[71]
Ribo-TISH	detect translation initiation sites (TIS) with high accuracy.	require high-depth ribosome profiling data.	https://github.com/zhpn1024/ribotish	[72]
Ribowave	analyze ribosome profiling data by wavelet transforms with high resolution.	require expertise in data analysis.	https://ribowave.ncrnalab.org/	[73]
RiboHMM	use hidden Markov models for ORF detection and robust for noisy data.	may not perform well on novel or poorly annotated genomes.	https://github.com/djf604/RiboHMM	[74]
ORFquant	detect and quantifies the translation of ORFs across different transcript isoforms.	require high-quality ribosome profiling data.	https://github.com/lcalviell/ORFquant/releases/tag/1.02	[75]
RP-BP	focuse on ribosome-protected fragment analysis with high specificity.	require advanced bioinformatic skills.	https://github.com/dieterich-lab/rp-bp	[76]
RiboToolkit	comprehensive for ribosome profiling analysis and user-friendly.	low-quality data affects analysis accuracy or leads to failure.	http://rnainformatics.org.cn/RiboToolkit/analysis.php	[77]
GWIPS-Viz	provide visualization of ribosome profiling data and integrates with public datasets.	require pre-processed data for analysis.	https://gwips.ucc.ie/cgi-bin/hgGateway	[78]
smORFer	specialize for sORF detection with high sensitivity.	may require additional validation for novel predictions.	https://github.com/AlexanderBartholomaeus/smORFer	[79]

detecting specific peptide fragments of microproteins [82]. The workflow for MS-based identification of sORF-encoded microproteins includes microprotein enrichment, MS sample preparation, MS data acquisition and raw data analysis (Fig. 2C).

Due to the low abundance of microproteins, different methods are developed to enrich microproteins from complex biological samples. Based on the properties of microproteins (e.g., size, hydrophobicity and charge), diverse enrichment methods can be employed, such as gel separation, molecular weight cutoff filters (10 or 30 kDa), precipitation with organic solvents (e.g., acetonitrile, methanol acetic acid), C8 solid-phase extraction (C8-SPE), and size-exclusion chromatography (SEC) [83–85]. The differences in enrichment methods leads to biases in microprotein identification. For instance, Zhang et al. found that C8-SPE and molecular weight cutoff methods had only a 7.58% overlap in identified microproteins, but combination of both methods successfully identified 762 novel microproteins across multiple tissues and cell lines [86]. Moreover, fractionation can effectively reduce sample complexity and improve microprotein identification [85, 87]. For example, Yang et al. used the SEC fractionation to enrich mouse microproteins, achieving a 1.4-fold increase in microproteins compared to unfractionated method [85].

Because of their small molecular weight, microproteins could be subjected to MS without enzymatic digestion into peptides. For example, Wang et al. identified 241 microproteins from Hep3B cells using undigested protein samples [88]. Despite this, enzymatic digestion is usually employed for the identification of microproteins. For enzymatic digestion, trypsin is commonly used to cleave peptide chains at the carboxyl side of lysine (K) and arginine (R) residues. However, microproteins may have few arginine and lysine residues, resulting in compromised sequence coverage detected by MS. Therefore, a multi-enzyme digestion strategy combining trypsin with other proteases, such as Glu-C, Lys-C or Asp-N, can be employed. Kaulich et al. demonstrated that digestion of multiple proteases in LC-MS/MS analysis improved the sequence coverage and the number of identified peptides for microproteins [89].

Data-Dependent Acquisition (DDA) is a commonly used MS data acquisition mode that selects precursor ions for data collection based on signal intensity, which has successfully identified thousands of microproteins in different species [86, 87, 90]. However, DDA tends to favor the detection of precursor ions with strong signals, often overlooking low-abundance molecules. Another acquisition mode, Data-Independent Acquisition (DIA), fragments ions across all mass ranges, capturing of both low- and high-abundance molecules and thereby enhancing its suitability for microprotein detection. For example, Martinez et al. successfully identified 85 microproteins in mouse adipocytes using DIA [91]. Because the spectra generated by DIA are complex to analyze [92], directDIA was thus developed to simplify the workflow and improve sensitivity and accuracy [93].

Currently, some publicly available databases have been developed for the analysis of MS raw data to identify microproteins, such as SmProt [94], sORFs.org [95] and OpenProt [96] (Table 3). Data in sORFs.org are derived from ribosome profiling, while SmProt and OpenProt

Table 3 Biological functions of lncRNA-encoded microproteins in cancer

database	resource	features	website	ref.
SmProt	Human, other species	- include 638,958 unique small proteins from 370 cell lines or tissues in 8 species. - offer comprehensive annotations of small proteins identified through ribosome profiling, literature, mass spectrometry, and other sources	http://bigdata.ibp.ac.cn/SmProt/index.html	[94]
sORFs.org	Human and other species	- include 263 354 sORFs from HCT116 (human), E14_mESC (mouse) and S2 (fruit fly). - investigate the potential of sORFs identified by ribosome profiling.	http://www.sorfs.org	[95]
OpenProt	Human and other species	- include all known proteins (RefProts), newly predicted isoforms (Isoforms), and novel proteins derived from alternative ORFs (AltProts), providing comprehensive support for the annotation of thousands of predicted ORFs. - retrieve supporting evidence for all proteins from MS, ribosome profiling, protein conservation, functional domain prediction and related proteins.	https://www.openprot.org/	[96]
NCPbook	Human and other species	- include 180,676 noncanonical peptides from 29 species of plants, animals, and microbes.	https://ncp.wiki/ncpbook/	[97]
FuncPEP v2.0	Human and other species	- include 152 functional short peptides translated from non-coding RNA, highlighting their roles in immune regulation and tumorigenesis.	https://bioinformatics.mdanderson.org/Supplements/FuncPEP/	[98]
LncPepAtlas	Human	- integrate multi-source evidence to annotate and explore the translational potential of lncRNA and their encoded peptides.	http://www.cnitbiotool.net/LncPepAtlas/	[99]
TransCirc	Human	- an interactive database integrating multi-omics evidence to predict and analyze the translational potential of circRNA and their encoded peptides.	https://www.biosino.org/transcirc/	[100]

integrate additional data sources like genomic and MS data. These databases provide detailed microprotein information, such as their sequences, genomic locations and start codons. Notably, SmProt and sORFs.org include both AUG and non-AUG initiated microproteins, while OpenProt only has AUG-initiated ones. Given the overlap between these databases is relatively low [88], combination of multiple databases may improve the identification of microproteins.

In MS data analysis, some spectra may fail to match peptides in the database. For unmatched spectra, *de novo* sequencing can be employed to identify microproteins [101], as exemplified by the pNovo software, which conducted *de novo* peptide sequencing and mapped 1,682 peptides to 2,544 sORFs randomly distributed across human chromosomes [88]. In addition, Pan et al. combined database-dependent analysis with *de novo* sequencing to successfully identify 1074 microproteins in mouse tissues [102].

Experimental validation of microproteins

Microproteins predicted or detected as above are subsequently validated using two experimental strategies: exogenous expression and endogenous detection. For exogenous expression, sORF is fused with fluorescent protein (e.g., GFP, mCherry) or epitope tag (e.g., FLAG, HA, Myc) to construct its expression plasmid, which is then introduced into cells to express the microprotein of interest [5, 103]. Specifically, epitope tag or fluorescent protein lacking start codon is cloned into the C-terminus of target sORF, and antibodies against the fluorescent protein or epitope tag are then used to detect the expression of microprotein by immunoblotting and immunofluorescence. For instance, Zhang et al. demonstrated that four lncRNA with sORF may encode microproteins by ectopic expression of microproteins with FLAG and GFP tags [85].

The endogenous microproteins can be detected through the specific antibodies developed against these microproteins. For instance, Li et al. validated the expression of microprotein MIAC using customized monoclonal antibodies [104]. However, due to the small molecular weight and low antigenicity of certain microproteins, it is challenging to develop specific and effective antibodies. In this case, CRISPR/Cas9-mediated gene editing can be employed to insert fluorescent or epitope tags into the target DNA sequence, enabling the detection of microprotein expression and localization by the corresponding tag antibodies. For example, Na et al. generated Cas9-directed knock-in HEK293T cell lines with a 3xGFP11-FLAG-HA tag appended to the 3' end of corresponding ORF and validated the endogenous expression and subcellular localization of tagged microprotein, shedding light on their potential biological functions [33].

Dysregulation of sORF-encoded microproteins in cancer

Accumulating evidence has demonstrated that the expression of sORF-encoded microproteins is widely dysregulated in a variety of malignancies. These alternations are often associated with disruptions in the regulatory mechanisms governing RNA abundance, translation efficiency, or protein stability, as outlined in Tables 4 and 5.

Dysregulation of specific transcription factors can lead to abnormal expression of microprotein-coding transcripts. For example, the transcription factor GATA3 has been shown to suppress the transcription of LINC00887 by binding to two responsive elements within its promoter. In ccRCC, reduced GATA3 expression results in the upregulation of LINC00887 and its encoded microprotein, ACLY-BP [106]. Conversely, in hepatocellular carcinoma (HCC), the TGF-β-activated transcription factor SMAD3 promoted the transcription of LINC02551, leading to a marked upregulation of its encoded microprotein, JunBP [122].

Unlike linear RNAs, circRNAs are generated by the back-splicing of primary transcripts, a process regulated by *cis*-acting elements (e.g., Alu repeats) in the flanking introns and diverse *trans*-acting factors. Dysregulation of these factors probably contributes to the aberrant biogenesis of circRNAs encoding microproteins. For instance, the DExH-box helicase 9 (DHX9) specifically binds to Alu repeats, thereby inhibiting circularization by disrupting the pairing of these repeats [184]. In intrahepatic cholangiocarcinoma (ICC), DHX9 expression is decreased upon IL-6 stimulation, leading to elevated levels of the circRNA cGGNBP2 and its encoded microprotein, cGGNBP2-184aa [154]. In addition to helicase like DHX9, splicing factors also exert indispensable roles in governing circRNA biogenesis. Song et al. demonstrated the splicing factor SLU7, which was significantly downregulated in triple-negative breast cancer (TNBC), bound to Alu elements within primary transcripts of circCAPG to inhibit its circularization, ultimately resulting in reduced expression of its encoded microprotein, CAPG-171aa [151]. However, how these *trans*-acting factors selectively regulate the biogenesis of certain microprotein-coding circRNAs remains poorly understood and warrant further investigation.

Cellular RNAs undergo extensive structural and chemical modifications, many of which are essential for their biogenesis and function regulation. Consequently, dysregulation of RNA modifications in cancer leads to aberrant expression of transcripts encoding microproteins. METTL14 plays a pivotal role in rewiring RNA behavior by introducing m^6A modifications on target transcripts. In HCC, METTL14-mediated m^6A modification on circSTX6 is found to suppress its expression, leading to the significant downregulation of the

Table 4 Representative databases for microprotein research

microprotein	lncRNA	size (aa)	cancer types	expression [a]	biological function	ref.
AC115619-22aa	AC115619	22	HCC	↓	impair m^{6}A methyltransferase complex assembly by interacting with WTAP.	[105]
ACLY-BP	LINC00887	91	ccRCC	↑	promote lipid metabolism and cancer cell proliferation by maintaining ACLY acetylation.	[106]
AF127577.4-ORF	AF127577.4	29	GBM	↓	reduce METTL3 stability by disrupting METTL3-ERK2 interaction.	[107]
APPLE	ASH1L-AS1	90	AML	↑	enhance mRNA circularization and eIF4F initiation complex.	[108]
ASAP	LINC00467	94	CRC	↑	regulate ATP synthase activity.	[109]
ASRPS	LINC00908	60	TNBC	↓	suppress tumor progression by reducing VEGF levels.	[110]
ATMLP	AFAP1-AS1	90	NSCLC	↑	disrupt mitochondrial homeostasis and suppress autophagosome formation.	[111]
BVES-AS1-201-50aa	BVES-AS1	55	CRC	↑	promote cell viability, migration and invasion via activating SRC/mTOR signaling pathway.	[112]
C20orf204-189AA	LINC00176	189	HCC	↑	promote cell proliferation by interacting with nucleolin.	[113]
CASIMO1	*CASIMO1*	83	TNBC	↑	promote cell proliferation by modulating lipid droplet formation.	[114]
CIP2A-BP	LINC00665	52	TNBC	↓	suppress the PI3K/AKT/NFkB pathway.	[115]
DDUP	CTBP1-DT	186	OC	↓	promote cisplatin resistance.	[36]
ENSEP3	NEAT1	9	ESCC	↓	suppress cancer growth by reducing RAFHSP90β complex stability.	[116]
FORCP	LINC00675	79	CRC	↓	induce apoptosis upon ER stress.	[117]
HCP5-132aa	HCP5	132	TNBC	↑	inhibit ferroptosis pathways by regulating GPX4 expression and lipid ROS levels.	[118]
HCP5-132aa	HCP5	132	GC	↑	inhibit ferroptosis via increasing *SLC7A11* and *G6PD* mRNA stability.	[119]
HDSP	HOXA10-HOXA9	112	GC	↑	drive gastric cancer progression through the MECOM-SPINK1-EGFR signaling axis.	[120]
HOXB-AS3	HOXB-AS3	53	CRC	↓	suppress glucose metabolism reprogramming.	[121]
JunBP	LINC02551	174	HCC	↑	promote cancer metastasis by increasing c-Jun phosphorylation and SMAD3 expression.	[122]
KRASIM	NCBP2-AS2	99	HCC	↓	inhibit oncogenic signaling via decreasing KRAS protein level.	[123]
LINC00511-133aa	LINC00511	133	TNBC	↑	promote cell invasion and stemness through the Wnt/β-catenin pathway.	[124]
LINC00665_18aa	LINC00665	18	OS	↓	inhibit cell proliferation and migration via the regulation of CREB1/RPS6KA3 interaction.	[125]
Linc013026-68AA	Linc013026	68	HCC	↑	promote cell proliferation.	[126]
MBOP	LINC01234	85	CRC	↑	activate MEK1/ERK/MMP2/MMP9 pathway.	[127]
MIAC	RP11-469H8.6	51	HNSCC	↓	inhibit SEPT2/ITGB4 signaling.	[104]
MIAC	RP11-469H8.6	51	CcRCC	↓	inhibit cell growth and metastasis via reducing EREG/EGFR expression.	[128]
MRP	LncRNA LY6E-DT	153	BC	↑	stabilize *EGFR* mRNA via interacting with HNRNPC.	[129]
NBASP	FAM201A	155	NB	↓	inhibite FABP5-mediated MAPK pathway activation.	[130]
NOBODY	LINC01420	71	-	-	regulate P-body quantity and mRNA decay.	[131, 132]
N1DARP	LINC00261	41	Pancreatic cancer	↑	promote N1ICD degradation.	[133]
PACMP	CTD-2256P15.2	44	TNBC	↓	maintain CtIP protein stability and promote PARP1-dependent poly(ADP-ribosyl)ation.	[134]
pep-AKR1C2	lncAKR1C2	163	GC	↑	promote FAO and ATP production.	[135]
pep-AP	Lnc-AP	37	CRC	↓	enhance oxaliplatin sensitivity by inducing ROS accumulation and apoptosis.	[136]
Pep-KDM4A-AS1	LncRNA KDM4A-AS1	61	ESCC	↓	regulate fatty acid metabolism and redox processes.	[137]
PRDM16-DT	LINC00982	148	CRC	↓	inhibit CRC metastasis and oxaliplatin resistance by regulating CHEK2 splicing.	[138]
pTINCR	*TINCR*	87	Various cancers	↓	increase CDC42 SUMOylation and trigger a pro-differentiation cascade.	[139]
RBRP	LINC00266-1	71	CRC	↑	stabilize *MYC* mRNA by binding to m^{6}A reader protein IGF2BP1.	[140]
SMIM26	LINC00493	26	ccRCC	↓	regulate mitochondrial function and inhibit AKT signaling.	[141]

Table 4 (continued)

microprotein	lncRNA	size (aa)	cancer types	expression [a]	biological function	ref.
SMIM30	LINC00998	59	HCC	↑	induce SRC/YES1 membrane anchoring and MAPK pathway activation.	[5]
SP0495	KIAA0495-ORF2	201	Various cancers	↓	bind to phosphoinositide to inhibit AKT phosphorylation and its downstream signaling.	[142]
sPEP1	HNF4A-AS1	51	NB	↑	inhibit SMAD4 transcription via interaction with eEF1A1.	[45]
SRSP	lncRNA LOC90024	130	CRC	↑	promote oncogenic L-Sp4 expression while suppressing non-oncogenic S-Sp4 level.	[143]
TINCR	*TINCR*	120	HNSCC	↓	suppress tumorigenesis.	[144]
TP53LCO2	AC010501.1	109	-	↓	inhibit cell proliferation.	[145]
XBP1SBM	lncRNA MLLT4-AS1	21	TNBC	↑	activate VEGF transcription.	[146]
YY1BM	LINC00278	21	ESCC	↓	enhance eEF2 activity and induce apoptosis.	[41]
-	LINC02381	23	GBM	-	regulate ferroptosis by SLC2A10.	[147]

a Symbols: ↑, upregulation; ↓, downregulation

circSTX6-encoded microprotein, circSTX6-144aa. This observation is further supported by the negative correlation between METTL14 expression and the levels of circSTX6 and circSTX6-144aa in HCC tissues [167]. In contrast, in breast cancer (BC), METTL14 is shown to upregulate the expression of lncRNA LY6E-DT and its encoded microprotein MRP. Notably, the knockdown of IGF2BP1, a well-characterized m^{6}A "reader" protein, reversed METTL14-induced upregulation of LY6E-DT, suggesting that METTL14-mediated m^{6}A modifications promote LY6E-DT expression in an IGF2BP1-dependent manner [129]. These findings highlight the dual roles of METTL14 in regulating target RNA abundance, likely mediated by the recruitment of distinct m^{6}A reader proteins to the modified transcripts. Furthermore, it would be of interest to investigate how other types of modifications contribute to the regulation of microprotein-coding RNAs.

During translation, eukaryotic initiation factor 3 (eIF3), a multiprotein complex composed of 13 distinct subunits (eIF3a-m), plays a critical role in both cap-dependent and cap-independent translation initiation. Notably, eIF3J has been shown to exert an inhibitory effect on the translation of a subset of circRNAs by impeding the binding of eIF3a and eIF3b to these circRNAs [185]. However, in HER2-positive BC, the transcriptional repression of eIF3J by NRF2 results in the enhanced translation of circ-β-TrCP peptide that confers trastuzumab resistance [183].

Certain E3 ubiquitin ligases facilitate the transfer of ubiquitin from ubiquitin carrier proteins to target microproteins, thereby promoting their degradation and reducing their abundance. For example, the microprotein circMAP3K4-455aa, encoded by circMAP3K4, is highly expressed in HCC. Its stability is regulated by the E3 ubiquitin ligase MIB1, which shortens its half-life through ubiquitination [43].

Biological function of sORF-encoded microproteins in cancer

Regulating gene transcription

Emerging studies have shown that certain sORF-encoded microproteins can regulate gene expression through directly interacting with transcription factors or their associated binding partners (Fig. 3A). For example, Xiang et al. reported that the microprotein PINT87aa, encoded by LINC-PINT, bound to the DNA-binding domain of the transcription factor FOXM1, effectively inhibiting its transcription activity in HCC cells [177]. Similarly, the microprotein CORO1C-47aa, encoded by hsa_circ_0000437, is shown to interact directly with the PAS-B domain of ARNT to prevent ARNT from binding to the transcription factor TACC3, ultimately suppressing VEGF transcription in endometrium tumor [171].

Regulating RNA splicing

During cancer-associated transcriptome reprogramming, sORF-encoded microproteins have been shown to precisely regulate alternative splicing by interacting with key splicing factors, including members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family, and the serine/arginine-rich splicing factor (SRSF) family (Fig. 3B). For instance, the microprotein PRDM16-DT, encoded by LINC00982, competitively interacts with hnRNP A2B1 to prevent the binding of hnRNP A2B1 to exon 9 of *CHEK2* transcript, leading to the formation of the long isoform (L-CHEK2) while simultaneously suppressing the production of the short CHEK2 variant [138]. Similarly, another study demonstrated that the microprotein SRSP, encoded by LOC90024, enhanced the binding of SRSF3 to exon 3 of the Sp4 transcript. This enhanced interaction facilitates the selective inclusion of exon 3 to produce the oncogenic long isoform L-Sp4 protein while inhibiting the expression of the non-oncogenic short isoform S-Sp4 [143]. These findings highlight the ability

Table 5 Biological functions of circRNA-encoded microproteins in cancer

microprotein	circRNA	size (aa)	cancer type	expression [a]	biological function	ref.
463aa	circHECTD1	463	GBM	↓	suppress the vasculogenic mimicry formation by promoting ubiquitination and degradation of NR2F1.	[148]
AKT3-174aa	circAKT3	174	GBM	↓	reduce AKT phosphorylation.	[149]
AXIN1-295aa	circAXIN1	295	GC	↑	promote Wnt pathway via interfering β-catenin degradation.	[150]
CAPG-171aa	circCAPG	171	TNBC	↑	activate MEKK2-MEK1/2-ERK1/2 pathway.	[151]
c-E-Cad	circ-E-Cad	254	GC	↑	activate TGF-β/Smad/PI3K/AKT pathway.	[152]
c-E-Cad	circ-E-Cad	254	GBM	↑	activate EGFR-STAT3 pathway.	[153]
cGGNBP2-184aa	cGGNBP2	184	ICC	↑	promote IL-6/STAT3 pathway.	[154]
c-IGF1R	cIGF1R	66	NSCLC	↓	inhibit Parkin-mediated mitochondrial autophagy and promote drug-tolerant persister cell apoptosis	[155]
circATG4B-222aa	circATG4B	222	CRC	↓	promote oxaliplatin resistance by promoting autophagy.	[156]
circCCDC7-180aa	circCCDC7 (15,16,17,18,19)	180	PC	↓	reduce cell migration, invasion, and viability by up-regulating FLRT3.	[157]
circDDX17-63aa	circDDX17	63	GC	↓	suppress cell proliferation and migration.	[158]
circFNDC3B-218aa	circFNDC3B	218	CRC	↓	suppress EMT by inhibiting Snail expression.	[159]
circFOXP1-231aa	circFOXP1	231	ICC	↓	promote ferroptosis and suppress ICC recurrence.	[160]
circINSIG1-121	circINSIG1	121	CRC	↑	promote the ubiquitin-dependent degradation of INSIG1.	[161]
circLgr4-peptide	circLgr4	-	CRC	↑	promote Wnt/β-catenin signaling pathway.	[162]
circMAP3K4-455aa	circMAP3K4	455	HCC	↑	prevent apoptosis via inhibiting AIF cleavage and nuclear distribution.	[43]
circMRCKα-227aa	circMRCKα	227	HCC	↑	promote glycolysis.	[163]
circMRPS35-168aa	circMRPS35	168	HCC	↑	enhance cisplatin resistance by inhibiting the cisplatin-induced caspase-3 cleavage.	[164]
circPPP1R12A-73aa	circPPP1R12A	73	CRC	↑	activate Hippo-YAP pathway.	[165]
circSRCAP-75aa	circSRCAP	75	PC	↓	stabilize AR-V7 and driving enzalutamide resistance.	[166]
circSTX6-144aa	circSTX6	144	HCC	↑	promote tumor growth and metastasis.	[167]
circUBE4B-173aa	circUBE4B	173	ESCC	↑	activate MAPK/ERK pathway.	[168]
circZKSaa	circZKSCAN1	206	HCC	↓	promote mTOR degradation and enhance sensitivity to sorafenib treatment	[169]
circβ-catenin-370aa	circβ-catenin	370	NSCLC	↑	stabilize β-catenin and activate the Wnt pathway	[170]
CORO1C-47aa	hsa_circ_0000437	47	EC	↓	reduce VEGF expression and angiogenesis.	[171]
EIF6-224aa	circ-EIF6	224	TNBC	↑	Stabilize MYH9 and activate Wnt/β-catenin pathway.	[172]
FBXW7-185aa	circ-FBXW7	185	GBM	↓	promote c-Myc degradation.	[173]
HEATR5B-881aa	circHEATR5B	881	GBM	↓	suppress aerobic glycolysis by degrading JMJD5.	[174]
MAPK1-109aa	circMAPK1	109	GC	↓	inhibit MAPK pathway.	[175]
MET404	circMET	404	GBM	↑	interact with the MET-β subunit to activate MET receptor.	[16]
PDE5A-500aa	circPDE5A	500	ESCC	↓	reduce PI3K/AKT pathway via promoting PIK3IP1 degradation.	[176]
PDHK1-241aa	circPDHK1	241	ccRCC	↑	Interact with PPP1CA to inhibit AKT dephosphorylation.	[35]
PINT87aa	circ LINC-PINT	87	GBM	↓	interact with PAF1c to inhibit transcriptional elongation of multiple oncogenes.	[6]
PINT87aa	circ LINC-PINT	87	HCC	↓	induce senescence and cell cycle arrest by inhibiting PHB2 transcription.	[177]
SHPRH-146aa	circ-SHPRH	146	GBM	↓	inhibit proliferating cell nuclear antigen activity.	[178]
SHPRH-146aa	circ-SHPRH	146	NB	↓	promote P21 to inhibit CDKs.	[179]
SMO-193aa	circ-SMO	193	GBM	↑	enhance SMO cholesterol modification and relieve its inhibition by patched transmembrane receptors.	[180]
SP3-461aa	circSP3	461	ccRCC	↑	stabilize MYH9 and activate PI3K-Akt pathway	[181]
TRIM1-269aa	circTRIM1	269	BC	↑	activate CaM-dependent MARCKS translocation and PI3K/AKT/mTOR pathway.	[182]
β-TrCP-343aa	circ-β-TrCP	343	BC	↑	promote NRF2-mediated antioxidant pathway.	[183]

a Symbols: ↑, upregulation; ↓, downregulation

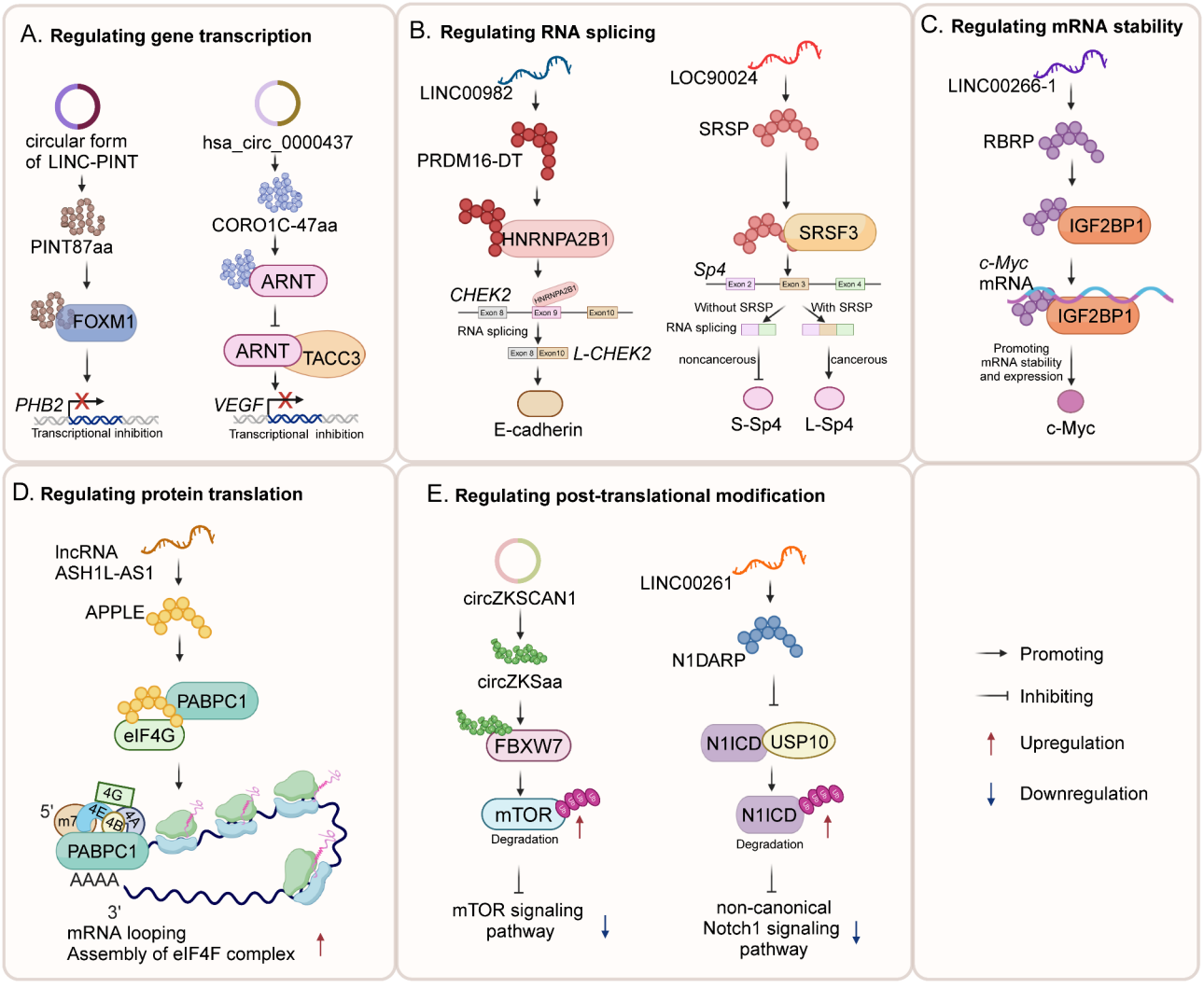

Fig. 3 Examples of the biological functions of microproteins. **(A)** Microproteins bind to transcription factors to regulate gene transcription; **(B)** Microproteins interact with splicing factors to regulate RNA splicing; **(C)** Microproteins interact with RNA-binding proteins to regulate the stability of target mRNAs; **(D)** Microproteins recruit translational factors to regulate protein translation; **(E)** Microproteins influence the stability of target proteins

of microproteins to modulate splicing patterns, thereby contributing to cancer progression.

Regulating mRNA stability

Certain RBPs, such as IGF2BP1 and HNRNPC, can specifically bind to the 3'-untranslated regions (3'-UTRs) of target mRNAs to regulate RNA stability [186–188]. Emerging evidence indicates that certain sORFs-encoded microproteins interact with these RBPs to influence the stability and fate of target mRNAs (Fig. 3C). For example, the microprotein RBRP, encoded by LINC00266-1, binds to IGF2BP1 and enhances its ability to recognize m^{6}A modifications on *c-Myc* mRNA, leading to the increased stability and translation efficiency of *c-Myc* mRNA during tumorigenesis [140]. Similarly, the microprotein MRP, derived from the lncRNA LY6E-DT, interacts with HNRNPC to strengthen the binding of HNRNPC to epidermal growth factor receptor (*EGFR*) mRNA, enhancing the stability of *EGFR* mRNA and the expression of EGFR protein in BC [129].

Regulating protein translation

sORF-encoded microproteins can function as scaffold to modulate the assembly and function of translation-related complexes, therefore controlling specific translational programs in cancer (Fig. 3D). For example, the microprotein APPLE, encoded by lncRNA ASH1L-AS1, facilitates the interaction between PABPC1 (poly(A)-binding protein cytoplasmic 1) and eIF4G (eukaryotic initiation factor 4G), consequently promoting mRNA circularization and the assembly of the eIF4F initiation complex to support a specific pro-cancer translation program [108].

Regulating post-translational modification

Post-translational modifications (PTMs), including ubiquitination and phosphorylation, are critical regulatory mechanisms that govern protein abundance and function. The ubiquitination modification is a dynamic process governed by the bidirectional regulation of E3 ubiquitin ligases and deubiquitinases (e.g., USPs). sORF-encoded microproteins have been shown to interact with these enzymes, either facilitating or impeding their ability to recognize and modify specific substrates. For instance, the microprotein circZKSaa, derived from the circZKSCAN1, interacts with the E3 ubiquitin ligase FBXW7 to promote the ubiquitination and subsequent degradation of mTOR, thereby suppressing mTOR signaling [169]. In contrast, the microprotein N1DARP (Notch1 degradation-associated regulatory protein), encoded by LINC00261, disrupts the interaction between Notch1 intracellular domain (N1ICD) and the ubiquitin-specific peptidase 10 (USP10), leading to the polyubiquitination of N1ICD via K11 and K48 linkages and the inhibition of both canonical and non-canonical Notch1 signaling pathways [133]. Additionally, microproteins also regulate protein stability through phosphorylation modification. For example, the microprotein HEATR5B-881aa, encoded by circHEATR5B, directly interacts with the Jumonji C domain-containing protein (JMJD5) and reduces its stability by promoting phosphorylation at the S361 site [174]. However, the precise molecular mechanisms underlying this process remain to be fully elucidated. Given the importance of other PTMs, such as acetylation and methylation, in modulating protein stability and function, it would be interesting to explore whether and how microproteins influence protein homeostasis by regulating these modifications.

The roles of sORF-encoded microproteins in cancer

Emerging evidence has highlighted the oncogenic or tumor-suppressive roles of sORF-encoded microproteins in the onset and progression of various cancers. These microproteins contribute to cancer biology through distinct mechanisms, including: (i) the modulation of proliferative signaling pathways, (ii) the evasion of programmed cell death, (iii) the regulation of angiogenesis, (iv) the control of metastatic potential, and (v) the reprogramming of cellular metabolism.

Modulating proliferative signaling

c-Myc is a well-known oncogenic transcription factor that potently initiates and sustains tumor growth programs. Its expression is frequently upregulated in cancer. Recent studies reveal the involvement of sORF-encoded microproteins in regulating c-Myc abundance, thereby maintaining proliferative signaling for cancer growth. For instance, the tumor-associated peptide RBRP, encoded by LINC00266-1, interacts to IGFBP1 to enhance the interaction between IGFBP1 and *c-Myc* mRNA, increasing mRNA stability and driving colorectal cancer (CRC) progression [140]. Conversely, FBXW7-185aa, a microprotein encoded by the circular isoform of the E3 ligase FBXW7α transcript, shortens the half-life of c-Myc protein by accelerating FBXW7α-mediated c-Myc degradation [173]. Furthermore, a recent study identified a secretory 114-amino-acid microprotein, MPEP, encoded by an ORF within 5'-UTR of *MYC* mRNA. MPEP functions as an agonistic ligand for the TRKB receptor tyrosine kinase to promote glioblastoma stem cell growth independently of MYC protein function [189]. These findings provide novel insights into biological significance of noncanonical ORFs and their encoded microproteins in cancer progression.

The MAPK/ERK signaling pathway, a critical pro-proliferative cascade frequently dysregulated in various cancers, can be modulated by multiple microproteins. In esophageal squamous cell carcinoma (ESCC), the microprotein circUBE4B-173aa has been identified as a direct interactor of MAPK1, enhancing its phosphorylation and thereby promoting the MAPK/ERK-mediated cell proliferation [168]. In addition, the *MAPK* gene encodes a microprotein, MAPK1-109aa, which is derived from its circular transcript of *MAPK1*. In gastric cancer (GC), MAPK1-109aa exerts a tumor suppressive effect by competitively binding to MEK1, the upstream kinase of MAPK1. This interaction inhibits the activation of MAPK1 and its downstream pro-proliferative signals [175].

Resisting cell death

Programmed cell death is an essential process for organisms to maintain internal homeostasis, whereas cancer cells have evolved diverse mechanisms to evade cell death programs including apoptosis and ferroptosis. A growing body of evidence highlights the important roles of cancer-associated microproteins in cell death (Fig. 4).

Apoptosis is a highly controlled process of cell death to eliminate damaged or abnormal cells in a caspase-dependent or caspase-independent ways. The expression of caspases, key executioners of apoptosis, can be regulated by specific microproteins. For example, the microprotein YY1BM, encoded by m^6A-modified lncRNA LINC00278, has been shown to inhibit apoptosis in ESCC cells by downregulating caspase-3 expression. Mechanistically, YY1BM hinders the interaction between the transcription factor Yin Yang 1 (YY1) and the androgen receptor (AR), thereby suppressing the transcription of eukaryotic elongation factor 2 kinase (eEF2K). The downregulation of eEF2K relieves its inhibitory phosphorylation of eukaryotic elongation factor 2 (eEF2), ultimately promoting the translation and expression of caspase-3 in

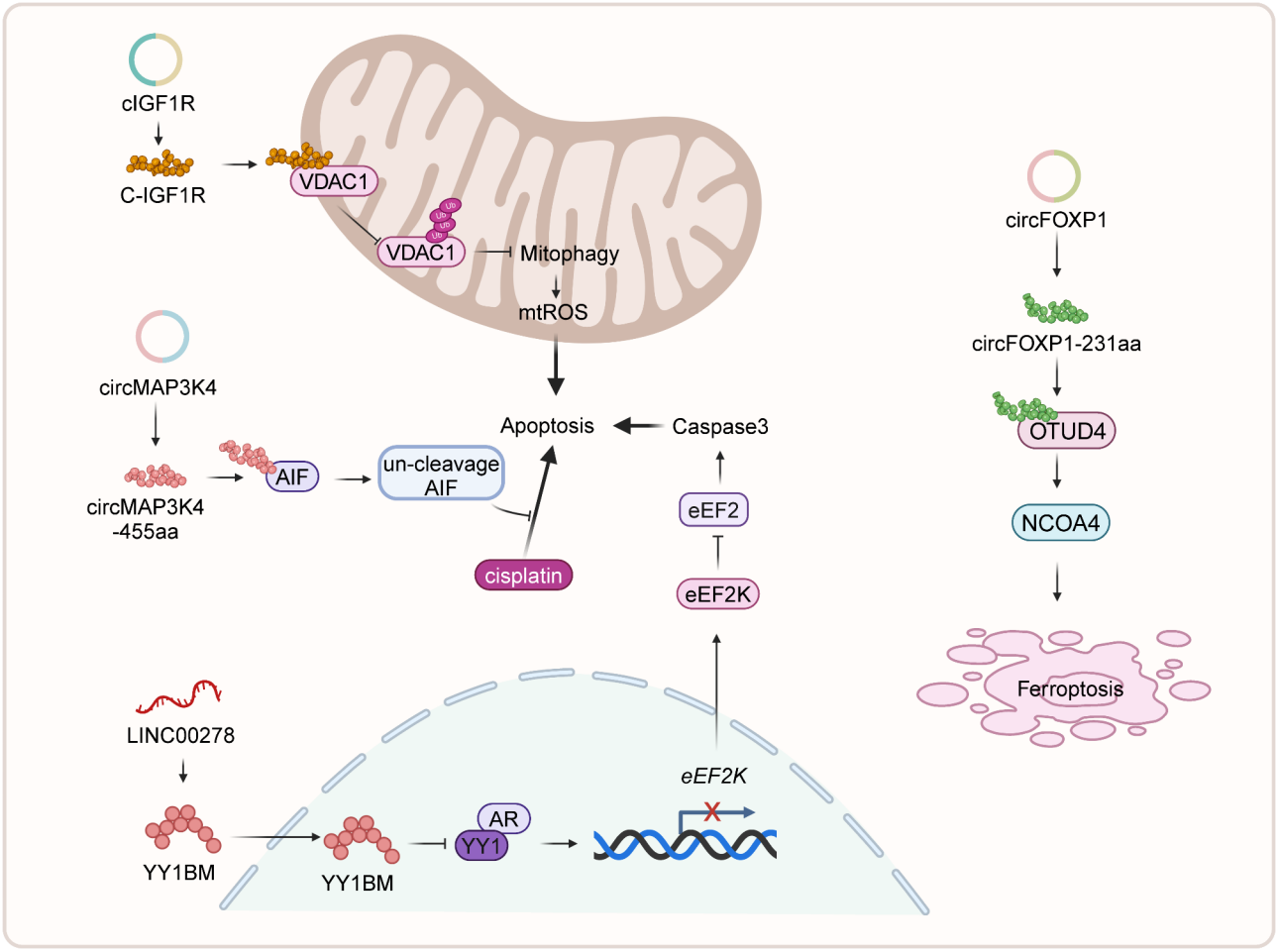

Fig. 4 Regulation of apoptosis and ferroptosis by microproteins

ESCC cells [41]. In caspase-independent apoptosis, apoptosis-inducing factor (AIF), a mitochondrial protein, undergoes N-terminal cleavage to form a soluble fragment upon apoptotic stimuli. Then the soluble AIF translocates to the nucleus to induce cell death [190]. Notably, the microprotein circMAP3K4-455aa, encoded by m^{6}A-modified circMAP3K4 and upregulated in HCC, has been demonstrated to confer resistance to cisplatin-induced apoptosis by directly interacting with AIF to prevent its N-terminal cleavage and subsequent nuclear translocation, implicating microproteins in cisplatin resistance [43].

Mitophagy is a selective form of autophagy that removes damaged mitochondria to prevent mitochondrial outer membrane permeabilization. Wang et al. revealed that C-IGF1R, a microprotein encoded by circRNA cIGF1R, suppressed mitophagy by interacting with mitochondrial membrane protein VDAC1 to inhibit its ubiquitination, thus promoting apoptosis in EGFR-TKIs-resistant non-small cell lung cancer (NSCLC) cells [155]. Hence, these findings indicate that microproteins act as a molecular switch to regulate the transition from a drug-tolerant persister state to the apoptotic process.

Distinct from apoptosis, ferroptosis is an iron-dependent cell death process characterized by uncontrolled lipid peroxidation. Nuclear receptor coactivator 4 (NCOA4) is a selective cargo receptor mediating the autophagic degradation of ferritin, a process essential for iron homeostasis. Recently, Wang et al. revealed that the microprotein circFOXP1-231aa promoted ferroptosis in ICC cells by interacting with deubiquitinating protease OTUD4 to enhance the stability and expression of NCOA4 [160]. On the contrary, the transcription factor NRF2 acts as a protective regulator against ferroptosis by targeting components of the ferroptosis cascades. Employing machine learning, Jiang et al. predicted that microprotein encoded by LINC02381 modulated ferroptosis in glioblastoma through regulating NRF2 signaling pathway [147], providing a new strategy for experimental design to validate microprotein functions. However, the broader implications of microproteins in other forms of cell death, such as necrosis and pyroptosis, remains poorly understood.

Regulating angiogenesis

The rapid growth of tumors requires a heavy supply of both oxygen and nutrients, which necessitates the formation of new blood vessels from pre-existing vasculatures via angiogenesis, a process predominately regulated by growth factors including members of VEGF family [191]. Studies have shown the regulation of VEGF expression by microproteins. For example, the microprotein XBP1SBM, encoded by the lncRNA MLLT4-AS1, promotes VEGF transcription by enhancing the nuclear localization of transcription factor XBP1s, thereby driving angiogenesis in TNBC [146]. Conversely, other microproteins, such as ASRPS peptide encoded by LINC00908 and CORO1C-47aa encoded by hsa_circ_0000437, act as negative regulators of VEGF transcription. Specifically, ASRPS directly binds to the coiled-coil domain (CCD) of STAT3 and downregulates STAT3 phosphorylation and subsequent VEGF transcription [110]. Similarly, CORO1C-47aa competitively binds to ARNT and prevents its interaction with the transcription factor TACC3, thereby inhibiting TACC3-mediated VEGF activation [171]. Collectively, these findings underscore the diverse and context-dependent roles of microproteins in the regulation of tumor angiogenesis.

Regulating metastasis

The TGF-β/Smad signaling pathway is widely characterized as a pivotal pro-metastatic signaling cascade, primarily through enhancing epithelial-mesenchymal transition (EMT), a critical early step enabling primary tumors to gain invasive and metastatic capabilities. Recent studies reveal the roles of TGF-β/Smad-regulated microproteins in EMT and tumor metastasis. In GC, activation of the TGF-β/Smad pathway upregulates the expression of the circular RNA transcript of E-cadherin (circ-E-Cad) and its encoded microprotein C-E-Cad [152]. Elevated levels of C-E-Cad subsequently enhance the expression of transcription factors Snail and Slug, leading to the downregulation of E-cadherin and the upregulation of N-cadherin and vimentin, thereby driving EMT and promoting tumor metastasis (Fig. 5, Left). Conversely, TGF-β/SMAD signaling has been reported to downregulate the expression of LINC00665-encoded microprotein CIP2A-BP in TNBC metastasis. CIP2A-BP directly binds with CIP2A to replace PP2A's B56γ subunit, thus releasing PP2A activity, which inhibits PI3K/AKT/NFκB pathway, resulting in decreased levels of MMP-2, MMP-9, and Snail and impairment of EMT process. Clinically, downregulation of CIP2A-BP in TNBC patients is associated with metastasis and poor survival in TNBC patients [115].

The Wnt/β-catenin signaling represents another critical pathway to drive tumor metastasis. The stability of cytoplasmic β-catenin is tightly controlled by the APC/AXIN/GSK-3β complex and serves as a critical regulatory switch in Wnt activation. Within this context, microproteins derived from components of this pathway have emerged as key regulators of Wnt-mediated metastasis. The microprotein AXIN1-295aa, encoded by a circular transcript of *AXIN1*, is highly expressed in GC and promotes Wnt/β-catenin signaling and metastasis [150]. Mechanistically, AXIN1-295aa competitively interacts with APC to inhibit GSK3β-mediated degradation of β-catenin, allowing cytoplasmic β-catenin to accumulate and translocate into the nucleus, where it drives the transcriptional activation of genes associated with GC metastasis (Fig. 5, Middle). Similarly, the microprotein circβ-catenin-370aa, encoded by circular β-catenin, promotes Wnt/β-catenin signaling by competitively binding to GSK3β and protected β-catenin from GSK3β-induced degradation [170]. Additionally, the abundance of GSK3β itself can be regulated by microproteins. In TNBC, the microprotein EIF6-224aa interacts with MYH9 to prevent its degradation. This stabilization enhances MYH9-mediated destruction of GSK3β via the ubiquitin-proteasome pathway, thereby amplifying β-catenin signaling and facilitating TNBC metastasis [172].

The pro-inflammatory cytokine IL-6 initiates a signaling cascade by binds to its cognate receptor (IL-6R) to activate JAK kinase, thereby leading to the phosphorylation and nuclear translocation of STAT3 [192]. In ICC, IL-6 has been shown to upregulate the circRNA GGNBP2 and its encoded microprotein, cGGNBP2-184aa [154]. This microprotein directly interacts with STAT3 to facilitate $STAT3^{Tyr705}$ phosphorylation. Thus, IL-6/cGGNBP2-184aa/STAT3 forms a positive feedback loop to sustain constitutive activation of IL-6/STAT3 signaling and promote ICC metastasis (Fig. 5, Right), underscoring the complexity of cancer metastatic networks.

Reprogramming cellular metabolisms

Metabolic reprogramming represents a hallmark of cancer, characterized by the dynamic adjustments of metabolic pathways for nutrients (e.g., glucose, fatty acid) in cancer cells to meet their demands for rapid proliferation, enhanced survival and invasive capabilities [193]. Several sORF-encoded microproteins have been identified as effective regulators of cancer metabolic reprogramming (Fig. 6).

The "Warburg effect" is a defined feature of metabolic reprogramming in cancer, characterized by the preferential reliance of cancer cells on glycolysis for energy production, even in the presence of sufficient oxygen [194]. Glycolysis, a multi-step enzymatic process involving the breakdown of glucose, is catalyzed by a series of enzymes, such as hexokinase 2 (HK2), phosphoglycerate kinase 1 (PGK1), pyruvate kinase M (PKM) and lactate dehydrogenase A (LDHA). The abundance of these

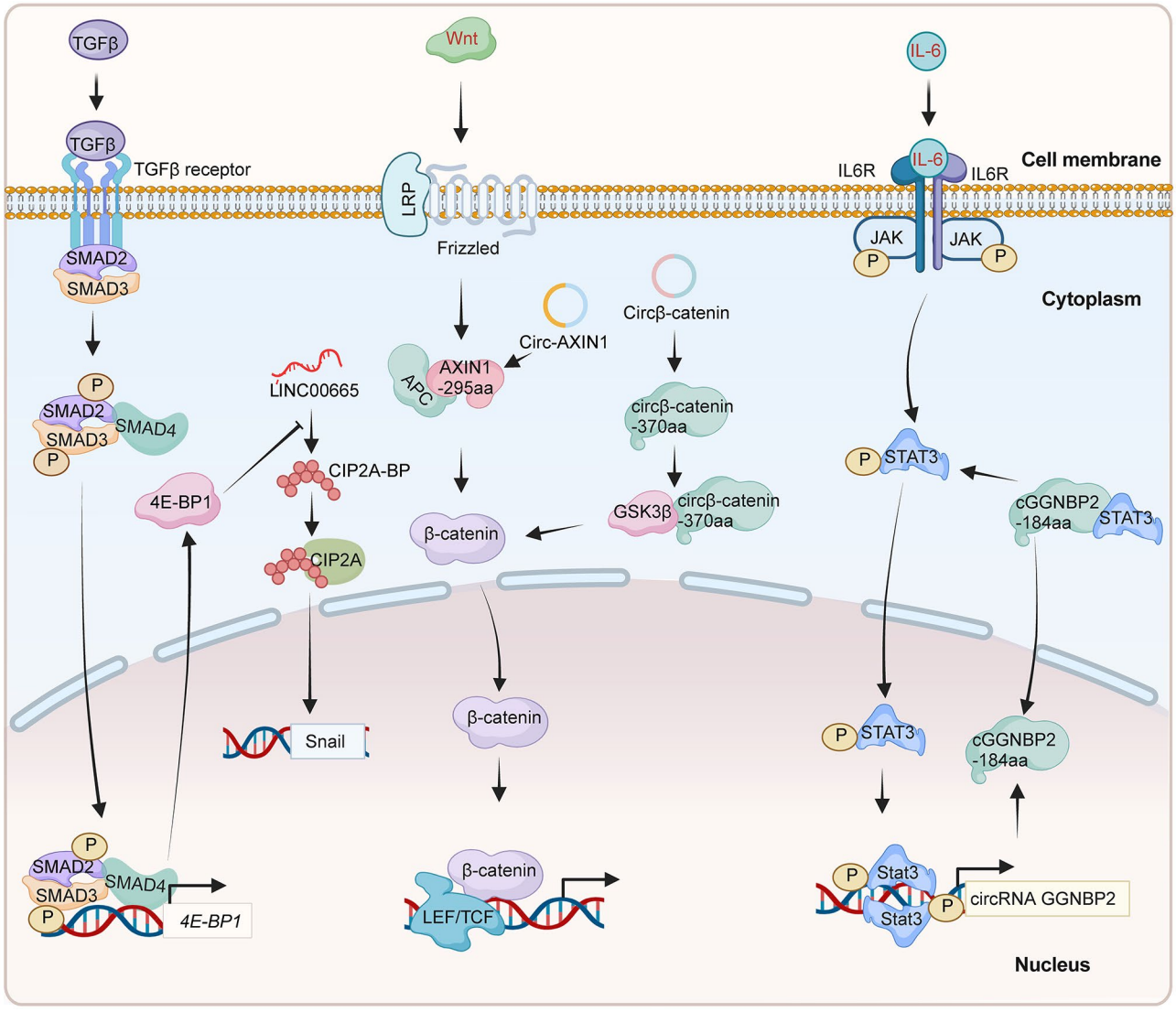

Fig. 5 Regulation of cancer metastasis by microproteins. Activation of Wnt/β-catenin and IL6/STAT3 pathways significantly promotes metastasis. TGF-β/Smad signaling pathways regulates microprotein expression, then markedly increases the expression of pro-EMT transcription factors such as Snail

enzymes can be regulated by microproteins through both transcriptional and post-transcriptional mechanisms. For example, Fang et al. revealed that the transcription factor Myeloid Zinc Finger 1 (MZF1) promoted the transcription of *HK2* and *PGK1*, thereby promoting aerobic glycolysis in neuroblastoma, whereas a 21-amino acid microprotein derived from the 5'-UTR of *MZF1* mRNA (termed MZF1-uPEP) functioned as a negative regulator of the MZF1/HK2/PGK1 signaling axis. Specifically, MZF1-uPEP interacts with YY1 to inhibit its transactivation activity, leading to the downregulation of MZF1 and its downstream glycolytic targets [195]. LDHA catalyzes the conversion of pyruvate to lactate and is frequently upregulated in various cancers. In glioblastoma, the microprotein P4-135aa, encoded by the pseudogene *MAPK6P4*, promotes the translocation of KLF15 into nucleus, where KLF15 directly binds to the promoter region of LDHA and enhanced its transcription [196]. Similarly, the microprotein circMRCKα-227aa, encoded by circMRCKα, enhances LDHA transcription and glycolysis in HCC. Mechanistically, circMRCKα-227aa enhances USP22-mediated deubiquitinating and upregulation of HIF-1α, driving HIF-1α-induced LDHA transcription [163]. Alternative splicing of *PKM* pre-mRNA determines cellular metabolic phenotypes. The inclusion of exon9 generates the PKM1 isoform, which favors oxidative phosphorylation, while the exclusion of exon 9, mediated by the splicing factor hnRNP A1, produce the PKM2 isoform, which promotes aerobic glycolysis. In CRC, the microprotein HOXB-AS3, encoded by the lncRNA HOXB-AS3, competitively binds to hnRNP A1, and this interaction antagonized the hnRNP

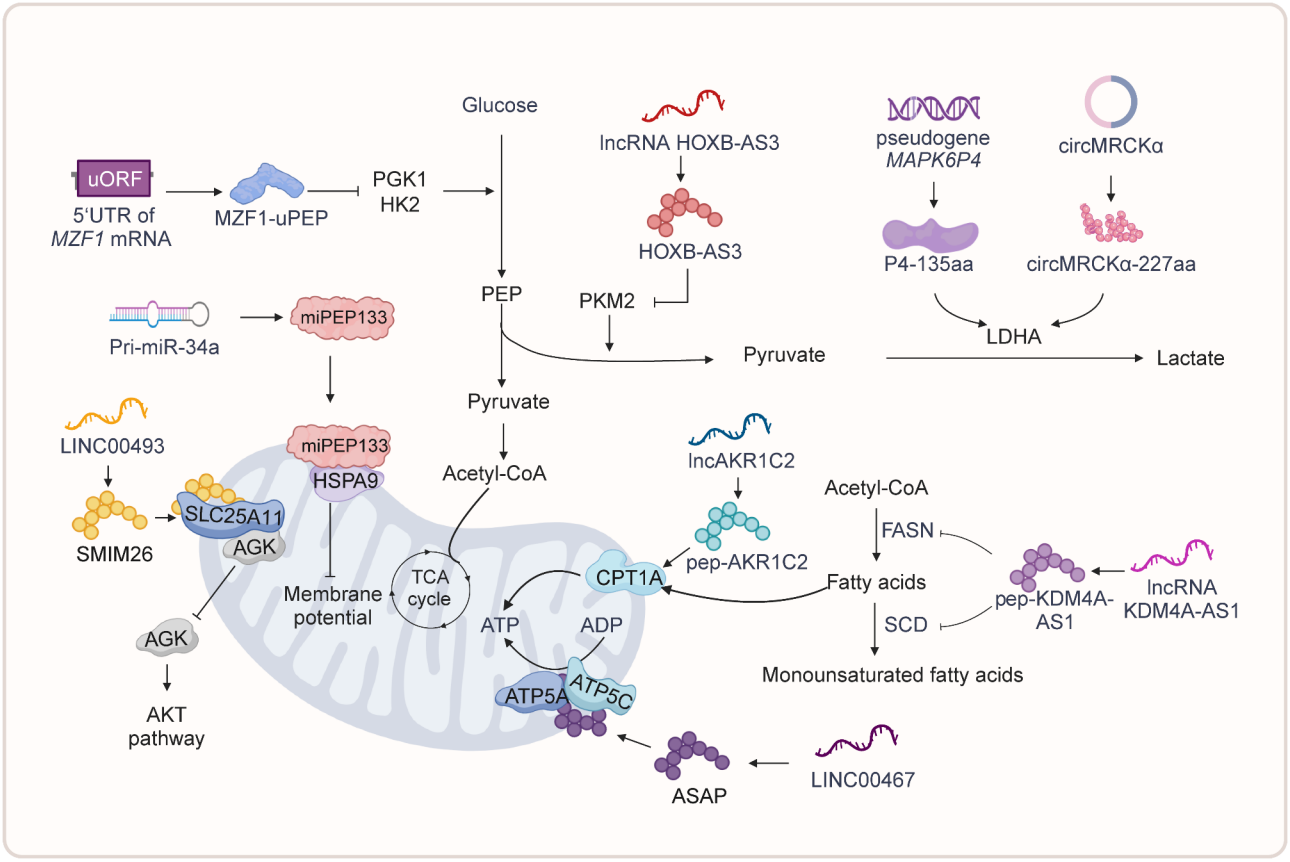

Fig. 6 Regulation of metabolic reprogramming by microproteins

A1-mediated production of PKM2 and subsequent aerobic glycolysis [121]. These findings provide novel mechanistic insights in the complex regulatory networks governing glucose metabolic reprogramming in cancer.

The abundance of fatty acids is elaborately controlled by the dynamic equilibrium between their synthesis and oxidation, with the disruptions in this metabolic balance being a recurrent feature in cancer [197]. In ESCC, the activity of key anabolic enzymes, including fatty acid synthase (FASN) and stearoyl-CoA desaturase (SCD), is significantly elevated to boost fatty acid synthesis. Conversely, the microprotein pep-KDM4A-AS1, encoded by lncRNA KDM4A-AS1, has been shown to suppress fatty acid anabolism in ESCC by inhibiting the expression of FASN and SCD [137]. During fatty acid catabolism, carnitine palmitoyl transferase I (CPT1A) facilitates the transport of long-chain fatty acids into the mitochondria for β-oxidation [198]. Zhu et al. demonstrated that the microprotein pep-AKR1C2, encoded by exosomal lncAKR1C2, promoted CPT1A expression through reducing YAP phosphorylation and subsequently enhancing YAP-induced transcriptional activation of CPT1A, leading to an accelerated fatty acid oxidation and increased ATP production in GC [135]. However, the role of sORF-encoded microproteins in regulating other enzymes involved in fatty acid metabolism remains to be further elucidated.

Mitochondria are pivotal in metabolic reprogramming, primarily through their role in generating ATP via oxidative phosphorylation (OXPHOS), a fundamental process that can be assessed by oxygen consumption rate (OCR), membrane potential, and ATP production rate [199]. Several sORF-encoded microproteins have been identified to interact with mitochondrial proteins to regulate their functions. For example, the mitochondria-located microprotein SMIM26, encoded by LINC00493, interacts with acylglycerol kinase (AGK) and the glutathione transporter regulator SLC25A11. This interaction traps AGK within mitochondria and subsequently inhibits AGK-mediated AKT phosphorylation [141]. Thus, SMIM26 exerts a tumor suppressive role in ccRCC, and its low expression correlates with poor prognosis for ccRCC patients. Similarly, the microprotein miPEP133, derived from the primary miR-34a transcript, functions as an anti-cancer agent in mitochondria. It interacts with mitochondrial heat shock protein 70 kDa (HSPA9) to prevent the binding of HSPA9 to its partner proteins, leading to a decrease in mitochondrial membrane potential across multiple cancer types [7]. In contrast, the microprotein ASAP, encoded by LINC00467, promotes

CRC progression by elevating ATP synthase activity. Specifically, ASAP interacts with ATP synthase subunits α and γ (ATP5A and ATP5C) to promote the assembly of ATP synthase, thereby increasing mitochondrial OCR and ATP production [109]. Hence, the interplay between microprotein and mitochondria highlights the complexity of energy metabolism in cancer cells.

Clinical significance of sORF-encoded microproteins in cancer

The expression profiles of sORF-encoded microproteins are closely associated with cancer progression and clinical outcomes, highlighting their potential as diagnostic and prognostic biomarkers. Furthermore, the important functions of microproteins during cancer pathogenesis underscore their therapeutic potential for cancer treatments.

sORF-encoded microproteins as potential diagnostic and prognostic biomarkers

Compared to normal tissues, sORF-encoded microproteins exhibit distinct expression patterns in cancer samples, making them ideal candidates as diagnostic biomarkers. For example, the microprotein MRP is found to be significantly upregulated in BC tissues. Its expression levels effectively distinguish patients with lymph node metastasis from those without, achieving an area under the curve (AUC) value of 0.7112, indicating its potential in predicting lymph node metastasis in BC [129]. Notably, microproteins encoded by ncRNAs are detectable not only in tumor tissues but also in various body fluids, highlighting their potential as non-invasive biomarkers. In a study by Pei et al., the levels of microproteins in the serum of NSCLC patients and healthy donors were compared, revealing that the microprotein ATMLP, encoded by lncRNA AFAP1-AS, is significantly upregulated in NSCLC patients. Importantly, ATMLP exhibits superior diagnostic efficiency (AUC = 0.852) compared to the conventional biomarker CEA (AUC = 0.746), underscoring its potential for early NSCLC detection [111]. Nevertheless, the diagnostic applicability of microproteins in other body fluids, such as saliva and urine, remains underexplored.

sORF-encoded microproteins also hold promise as prognostic markers for cancer progression and patient outcomes. In glioblastoma, high levels of microproteins such as MET404 and C-E-Cad are associated with poor overall survival (OS), suggesting their potential as prognostic indicators for glioblastoma [16, 153]. Similarly, high expression of other microproteins including ASAP [109], RBRP [140] and SRSP [143] is correlated with poor OS of CRC patients. Notably, RBRP and SRSP are further identified as independent prognostic factors associated with advanced clinical stages and higher histological grades [140, 143]. Conversely, high levels of microproteins such as ASRPS encoded by LINC00908 in TNBC and circATG4B-222aa encoded by circATG4B in CRC are correlated with improved OS [110, 156]. The prognostic significance of sORF-encoded microproteins has also been explored in other malignancies, including pancreatic ductal adenocarcinoma (PDAC) [17], HCC [167], ovarian cancer (OV) [36] and acute myeloid leukemia (AML) [108].

Although microproteins have emerged as promising biomarkers for cancer diagnosis and prognosis, their translation into clinical practice still faces several challenging issues, such as the identification of robust, reproducible and cancer type-specific biomarkers, along with the establishment of standardized detection protocols. Additionally, it is essential to develop more efficient, specific, and reliable analytical techniques for microprotein detection in clinical samples.

sORF-encoded microproteins as therapeutic targets

Given that sORF-encoded microproteins exhibit either oncogenic or tumor-suppressive functions, inhibiting oncogenic microproteins or restoring/enhancing the function of tumor-suppressive microproteins could be promising strategies for cancer therapy.

To achieve effective anti-cancer effects, many researches focus on restoring and enhancing the function of microproteins with the tumor suppressive role. For example, Zhai et al. found that LINC00261-endoced microprotein N1DARP inhibited tumor growth through regulating the USP10-Notch1 signaling axis [133]. Based on this discovery, the researchers developed SAH-mAH2-5, a cell-penetrating peptide designed to mimic the helical structure of N1DARP while exhibiting enhanced physicochemical stability. Moreover, SAH-mAH2-5 demonstrates potent anti-tumor activity against Notch1-activated PDAC cells, with minimal off-target and systemic toxicity [133]. Similarly, Dong et al. identified a conserved 9-amino acid peptide, ENSEP3, encoded by NEAT1, suppressed ESCC proliferation through inhibiting RAF expression and its downstream MAPK pathway. In patient-derived xenograft (PDX) models, synthetic ENSEP3 peptides specifically inhibits MAPK pathway activation, leading to significant suppression of ESCC tumor growth [116]. In addition, delivering plasmids encoding microproteins using nanocarriers can enhance the tumor-suppressive efficacy of microproteins. For example, Lei et al. identified that the circPDE5A-encoded microprotein PDE5A-500aa exerted the tumor-suppressive functions, and the delivery of its expression plasmid using a reduction-responsive nanoplatform (Meo-PEG-S–S-PLGA) effectively suppressed ESCC growth and metastasis both in vitro and in vivo [176].

Increasing studies have demonstrated that targeting oncogenic microproteins by shRNA or CRISPR/Cas9 system can effectively suppress tumor growth. For instance, Song et al. demonstrated that lentivirus-mediated delivery of shRNA targeting the oncogenic microprotein sPEP1, administered via intravenous injection, remarkably inhibited tumor growth [45]. Ge et al. reported that intratumoral injection of CRISPR/Cas9 vectors targeting the microprotein ASAP significantly inhibited CRC growth in PDX mouse models [109]. Similarly, an AAV-mediated Cas9/sgRNA delivery system is successfully applied for in vivo knockdown of the oncogenic HCP5-132aa microprotein, yielding substantial tumor growth inhibition in PDX models [119].

In addition to genetic modification approaches, targeting oncogenic microproteins with specific antibodies has also emerged as a promising strategy in cancer treatment. For example, Gao et al. reported that antibodies targeting the circ-E-Cad-encoded microprotein C-E-Cad effectively reduced STAT3 phosphorylation and inhibited the proliferation of glioma stem cell (GSC) [153]. Given that C-E-Cad activates EGFR signaling through its interaction with the CR2 domain of EGFR, the combination administration of C-E-Cad-targeting antibodies and EGFR -targeting antibodies dramatically suppressed tumor growth and improved survival rates in GSC-xenograft models. Similarly, the combination of MET404Ab, an antibody targeting circMET-encoded peptide MET404, with Onartuzumab, an FDA-approved MET antibody, demonstrated significant efficacy in inhibiting glioblastoma progression [16]. Owing to their small molecular weight and inherent immunogenic properties, microproteins may serve as potential antigens, enabling immune system to generate antibodies against tumors in vaccine therapy. For instance, Zeng et al. identified the RNF10 uPeptide, derived from the 5'-UTR of *RNF10* mRNA, as an immunogenic antigen in a CT26 murine tumor model, where it was specifically recognized by $CD8^+$ T cells to confer significant anti-tumor activity in mice. Notably, HLA-A2-restricted cytotoxic T lymphocytes (CTLs) isolated from pancreatic cancer patients recognizes the RNF10 uPeptide epitope (RLFGQQQRA) and then lysed HLA-$A2^+$ pancreatic cancer cells expressing the RNF10 uPeptide [200]. Similarly, Kikuchi et al. identified the PVT1 peptide, encoded by lncRNA PVT1, as a novel tumor-specific antigen in CRC. The PVT1 peptide is presented by HLA class I molecules and recognized by $CD8^+$ tumor-infiltrating lymphocytes (TILs) and peripheral blood mononuclear cells (PBMCs) from CRC patients, highlighting its potential application in CRC vaccine development [201].

Summary and perspectives

Advances in high-throughput sequencing and MS technologies have revealed a vast repertoire of previously undetected microproteins, significantly contributing to the complexity and diversity of proteomes. Many of these microproteins are encoded by sORFs located within ncRNAs or the UTRs of mRNAs. Despite substantial progress in this field, the identification of microproteins remain challenging due to their short length and low abundance. Thus, a concerted effort is required to optimize approaches such as MS and Ribo-seq to enhance detection sensitivity and specificity. Furthermore, the integration of multi-omics data holds promise for enabling a comprehensive characterization of the microprotein landscape. However, findings derived from such high-throughput screening approaches necessitate rigorous downstream validation to rule out false-positive results. Therefore, the development of highly specific and efficient antibodies against the microprotein of interest is essential to support experimental research and facilitate clinical validation.

Dysregulation of sORF-encoded microproteins has been increasingly implicated in various cancers. Although some microproteins have been demonstrated to play oncogenic or tumor-suppressive functions during tumor progression, the precise molecular mechanisms underlying their roles remain incompletely elucidated. Thus, it is essential to identify novel microproteins involved in tumorigenesis and to investigate their potential as therapeutic targets in cancer. Although multiple strategies have been explored to target microproteins for cancer therapy, proteolysis-targeting chimeras (PROTACs) has not yet been applied to oncogenic microproteins. Given the unique advantages of PROTACs, such as their ability to overcome drug resistance [202], developing highly specific and effective PROTAC degraders that selectively target cancer-associated microproteins while minimizing off-target effects represents a promising direction for future research.

Abbreviations

AML	Acute myeloid leukemia
BC	Breast cancer
ccRCC	Clear cell renal cell carcinoma
CircRNA	Circular RNA
CRC	Colorectal cancer
EC	Endometrium cancer
EGFR	Epidermal growth factor receptor
ESCC	Esophageal squamous cell carcinoma
GBM	Glioblastoma
GC	Gastric cancer
HCC	Hepatocellular carcinoma
HNSCC	Head and neck squamous-cell carcinoma
ICC	Intrahepatic cholangiocarcinoma
IRESs	Internal ribosome entry sites
LncRNA	Long non-coding RNA
m^6A	N6-methyladenosine
MS	Mass spectrometry
NB	Neuroblastoma

ncRNAs	Non-coding RNAs
NSCLC	Non-small cell lung cancer
OC	Ovarian cancer
ORF	Open reading frame
OS	Osteosarcoma
PC	Prostate cancer
PDAC	Pancreatic ductal adenocarcinoma
PDX	Patient-derived xenograft
pri-miRNA	Primary microRNA
Ribo-Seq	Ribosome profiling sequencing
sORF	Small open reading frame
TNBC	Triple-negative breast cancer
uORF	Upstream open reading frame
VEGF	Vascular endothelial growth factor

Acknowledgements
All figures in this article are created with BioRender.com.

Author contributions
Y.P. and J.L. conceived the structure of manuscript and revised the manuscript, T.Z. and Z.L. drafted initial manuscript, T.Z. generated figures.

Funding
This study is supported by the 1.3.5 Project for Disciplines of Excellence, West China Hospital of Sichuan University (ZYGD23027) and Post-Doctor Research Projects from West China Hospital and Sichuan University (2023HXBH095, 2024SCU12020).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 5 January 2025 / Accepted: 24 February 2025
Published online: 02 April 2025

References

1. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
2. Chew CL, Conos SA, Unal B, Tergaonkar V. Noncoding RNAs. Master regulators of inflammatory signaling. Trends Mol Med. 2018;24:66–84.
3. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to Forge new ones. Cell. 2014;157:77–94.
4. Wang J, Zhu S, Meng N, He Y, Lu R, Yan GR. ncRNA-Encoded peptides or proteins and Cancer. Mol Ther. 2019;27:1718–25.
5. Pang YN, Liu ZY, Han H, Wang BL, Li W, Mao CB, et al. Peptide SMIM30 promotes HCC development by inducing SRC/YES1 membrane anchoring and MAPK pathway activation. J Hepatol. 2020;73:1155–69.
6. Zhang M, Zhao K, Xu X, Yang Y, Yan S, Wei P, et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. Nat Commun. 2018;9:4475.
7. Kang M, Tang B, Li JX, Zhou ZY, Liu K, Wang RS, et al. Identification of miPEP133 as a novel tumor-suppressor microprotein encoded by miR-34a pri-miRNA. Mol Cancer. 2020;19:143.
8. Huang N, Chen Z, Yang X, Gao Y, Zhong J, Li Y, et al. Upstream open reading frame-encoded MP31 disrupts the mitochondrial quality control process and inhibits tumorigenesis in glioblastoma. Neuro Oncol. 2023;25(11):1947–62.
9. Freyer L, Hsu CW, Nowotschin S, Pauli A, Ishida J, Kuba K, et al. Loss of Apela peptide in mice causes low penetrance embryonic lethality and defects in early mesodermal derivatives. Cell Rep. 2017;20(9):2116–30.
10. Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, et al. Toddler. An embryonic signal that promotes cell movement via Apelin receptors. Science. 2014;343:1248636.
11. Niu L, Lou F, Sun Y, Sun L, Cai X, Liu Z, et al. A micropeptide encoded by LncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation. Sci Adv. 2020;6:eaaz2059.
12. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell. 2015;160:595–606.
13. Spencer HL, Sanders R, Boulberdaa M, Meloni M, Cochrane A, Spiroski AM, et al. The LINC00961 transcript and its encoded micropeptide, small regulatory polypeptide of amino acid response, regulate endothelial cell function. Cardiovasc Res. 2020;116:1981–94.
14. Deng Y, Zeng X, Lv Y, Qian Z, Guo P, Liu Y, et al. Cdyl2-60aa encoded by CircCDYL2 accelerates cardiomyocyte death by blocking APAF1 ubiquitination in rats. Exp Mol Med. 2023;55:860–9.
15. Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, et al. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. Mol Cell. 2017;66:22–e3729.
16. Zhong J, Wu X, Gao Y, Chen J, Zhang M, Zhou H, et al. Circular RNA encoded MET variant promotes glioblastoma tumorigenesis. Nat Commun. 2023;14:4467.
17. Cheng R, Li F, Zhang M, Xia X, Wu J, Gao X, et al. A novel protein RASON encoded by a LncRNA controls oncogenic RAS signaling in KRAS mutant cancers. Cell Res. 2023;33:30–45.
18. Hofman DA, Ruiz-Orera J, Yannuzzi I, Murugesan R, Brown A, Clauser KR, et al. Translation of non-canonical open reading frames as a cancer cell survival mechanism in childhood Medulloblastoma. Mol Cell. 2024;84:261–e276218.
19. Ye M, Zhang J, Wei M, Liu B, Dong K. Emerging role of long noncoding RNA-encoded micropeptides in cancer. Cancer Cell Int. 2020;20:506.
20. Liu YY, Zhang YY, Ran LY, Huang B, Ren JW, Ma Q, et al. A novel protein FNDC3B-267aa encoded by circ0003692 inhibits gastric cancer metastasis via promoting proteasomal degradation of c-Myc. J Transl Med. 2024;22:507.
21. Gross JD, Moerke NJ, von der Haar T, Lugovskoy AA, Sachs AB, McCarthy JE, et al. Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell. 2003;115:739–50.
22. Marintchev A, Edmonds KA, Marintcheva B, Hendrickson E, Oberer M, Suzuki C, et al. Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. Cell. 2009;136:447–60.
23. Pelletier J, Sonenberg N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature. 1988;334:320–5.
24. Ya-Jing HAO, Luo J-J, Zhang B, Chen R-S. The complexity of RNA translation: Non-translation, Part-translation, de Novo-translation and Over-translation. Prog Biochem Biophys. 2017;44:547–56.
25. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet. 2016;17:47–62.
26. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22:96–118.
27. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, et al. Detecting actively translated open reading frames in ribosome profiling data. Nat Methods. 2016;13:165–70.
28. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet. 2014;15:193–204.
29. Young SK, Wek RC. Upstream open reading frames differentially regulate Gene-specific translation in the integrated stress response. J Biol Chem. 2016;291:16927–35.
30. Barbosa C, Peixeiro I, Romao L. Gene expression regulation by upstream open reading frames and human disease. PLoS Genet. 2013;9:e1003529.
31. Rocha AL, Pai V, Perkins G, Chang T, Ma J, De Souza EV, et al. An inner mitochondrial membrane microprotein from the SLC35A4 upstream ORF regulates cellular metabolism. J Mol Biol. 2024;436:168559.
32. Ury B, Potelle S, Caligiore F, Whorton MR, Bommer GT. The promiscuous binding pocket of SLC35A1 ensures redundant transport of CDP-ribitol to the golgi. J Biol Chem. 2021;296:100789.
33. Na Z, Dai X, Zheng SJ, Bryant CJ, Loh KH, Su H, et al. Mapping subcellular localizations of unannotated microproteins and alternative proteins with microid. Mol Cell. 2022;82:2900–e29117.
34. Godet AC, David F, Hantelys F, Tatin F, Lacazette E, Garmy-Susini B, et al. IRES Trans-Acting factors, key actors of the stress response. Int J Mol Sci. 2019;20:924.
35. Huang B, Ren J, Ma Q, Yang F, Pan X, Zhang Y, et al. A novel peptide PDHK1-241aa encoded by circPDHK1 promotes CcRCC progression via interacting

with PPP1CA to inhibit AKT dephosphorylation and activate the AKT-mTOR signaling pathway. Mol Cancer. 2024;23:34.
36. Ren L, Qing X, Wei J, Mo H, Liu Y, Zhi Y, et al. The DDUP protein encoded by the DNA damage-induced CTBP1-DT LncRNA confers cisplatin resistance in ovarian cancer. Cell Death Dis. 2023;14:568.
37. Charpentier M, Dupre E, Fortun A, Briand F, Maillasson M, Com E, et al. hnRNP-A1 binds to the IRES of MELOE-1 antigen to promote MELOE-1 translation in stressed melanoma cells. Mol Oncol. 2022;16:594–606.
38. Yang Y, Fan XJ, Mao MW, Song XW, Wu P, Zhang Y, et al. Extensive translation of circular RNAs driven by -methyladenosine. Cell Res. 2017;27:626–41.
39. Chen Y, Lin Y, Shu Y, He J, Gao W. Interaction between N(6)-methyladenosine (m(6)A) modification and noncoding RNAs in cancer. Mol Cancer. 2020;19:94.
40. Zheng W, Wang L, Geng S, Yang L, Lv X, Xin S, et al. CircMIB2 therapy can effectively treat pathogenic infection by encoding a novel protein. Cell Death Dis. 2023;14:578.
41. Wu S, Zhang L, Deng J, Guo B, Li F, Wang Y, et al. A novel micropeptide encoded by Y-Linked LINC00278 links cigarette smoking and AR signaling in male esophageal squamous cell carcinoma. Cancer Res. 2020;80:2790–803.
42. Shen G, Li F, Wang Y, Huang Y, Aizezi G, Yuan J, et al. New insights on the interaction between m(6)A modification and non-coding RNA in cervical squamous cell carcinoma. World J Surg Oncol. 2023;21:25.
43. Duan JL, Chen W, Xie JJ, Zhang ML, Nie RC, Liang H, et al. A novel peptide encoded by N6-methyladenosine modified circMAP3K4 prevents apoptosis in hepatocellular carcinoma. Mol Cancer. 2022;21:93.
44. Zeng K, Peng J, Xing Y, Zhang L, Zeng P, Li W, et al. A positive feedback circuit driven by m(6)A-modified circular RNA facilitates colorectal cancer liver metastasis. Mol Cancer. 2023;22:202.
45. Song H, Wang J, Wang X, Yuan B, Li D, Hu A, et al. HNF4A-AS1-encoded small peptide promotes self-renewal and aggressiveness of neuroblastoma stem cells via eEF1A1-repressed SMAD4 transactivation. Oncogene. 2022;41:2505–19.
46. Rombel IT, Sykes KF, Rayner S, Johnston SA. ORF-FINDER. A vector for high-throughput gene identification. Gene. 2002;282:33–41.
47. Min XJ, Butler G, Storms R, Tsang A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 2005;33:W677–680.
48. Tang S, Lomsadze A, Borodovsky M. Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. 2015;43:e78.
49. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics. 2011;27:i275–282.
50. Pockrandt C, Steinegger M, Salzberg SL. PhyloCSF++: a fast and user-friendly implementation of phylocsf with annotation tools. Bioinformatics. 2022;38:1440–2.
51. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41:e74.
52. Hanada K, Akiyama K, Sakurai T, Toyoda T, Shinozaki K, Shiu SH. sORF finder. A program package to identify small open reading frames with high coding potential. Bioinformatics. 2010;26:399–400.
53. Zhu M, Gribskov M, MiPepid. MicroPeptide identification tool using machine learning. BMC Bioinformatics. 2019;20:559.
54. Skarshewski A, Stanton-Cook M, Huber T, Al Mansoori S, Smith R, Beatson SA, et al. uPEPperoni: an online tool for upstream open reading frame location and analysis of transcript conservation. BMC Bioinformatics. 2014;15:36.
55. Mokrejs M, Masek T, Vopalensky V, Hlubucek P, Delbos P, Pospisek M. IRESite–a tool for the examination of viral and cellular internal ribosome entry sites. Nucleic Acids Res. 2010;38:D131–136.
56. Zhao J, Wu J, Xu T, Yang Q, He J, Song X. IRESfinder. Identifying RNA internal ribosome entry site in eukaryotic cell using framed k-mer features. J Genet Genomics. 2018;45:403–6.
57. Hong JJ, Wu TY, Chang TY, Chen CY. Viral IRES prediction system - a web server for prediction of the IRES secondary structure in Silico. PLoS ONE. 2013;8:e79288.
58. Kolekar P, Pataskar A, Kulkarni-Kale U, Pal J, Kulkarni A, IRESPred. Web server for prediction of cellular and viral internal ribosome entry site (IRES). Sci Rep. 2016;6:27436.
59. Zhou Y, Zeng P, Li YH, Zhang Z, Cui Q. SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. Nucleic Acids Res. 2016;44:e91.
60. Qiang X, Chen H, Ye X, Su R, Wei L. M6AMRFS: robust prediction of N6-Methyladenosine sites with Sequence-Based features in multiple species. Front Genet. 2018;9:495.
61. Chen W, Tang H, Lin H. MethyRNA: a web server for identification of N(6)-methyladenosine sites. J Biomol Struct Dyn. 2017;35:683–7.
62. Wei L, Chen H, Su R. M6APred-EL: A Sequence-Based predictor for identifying N6-methyladenosine sites using ensemble learning. Mol Ther Nucleic Acids. 2018;12:635–44.
63. Liu M, Wang Q, Shen J, Yang BB, Ding X. Circbank: a comprehensive database for circrna with standard nomenclature. RNA Biol. 2019;16:899–905.
64. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc. 2012;7:1534–50.
65. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009;324:218–23.
66. Choudhary S, Li W. Accurate detection of short and long active ORF using Ribo-seq data. Bioinformatics. 2020;36:2053–9.
67. Xiao Z, Huang R, Xing X, Chen Y, Deng H, Yang X. De Novo annotation and characterization of the translatome with ribosome profiling data. Nucleic Acids Res. 2018;46:e61.
68. Chun SY, Rodriguez CM, Todd PK, Mills RE. SPECtre: a spectral coherence–based classifier of actively translated transcripts from ribosome profiling sequence data. BMC Bioinformatics. 2016;17:482.
69. Ji Z. RibORF: identifying Genome-Wide translated open reading frames using ribosome profiling. Curr Protoc Mol Biol. 2018;124:e67.
70. Erhard F, Halenius A, Zimmermann C, L'Hernault A, Kowalewski DJ, Weekes MP, et al. Improved Ribo-seq enables identification of cryptic translation events. Nat Methods. 2018;15:363–6.
71. Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, et al. A Regression-Based analysis of Ribosome-Profiling data reveals a conserved complexity to mammalian translation. Mol Cell. 2015;60:816–27.
72. Zhang P, He D, Xu Y, Hou J, Pan BF, Wang Y, et al. Genome-wide identification and differential analysis of translational initiation. Nat Commun. 2017;8:1749.
73. Xu Z, Hu L, Shi B, Geng S, Xu L, Wang D, et al. Ribosome elongating footprints denoised by wavelet transform comprehensively characterize dynamic cellular translation events. Nucleic Acids Res. 2018;46:e109.
74. Raj A, Wang SH, Shim H, Harpak A, Li YI, Engelmann B et al. Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. Elife. 2016; 5.
75. Calviello L, Hirsekorn A, Ohler U. Quantification of translation uncovers the functions of the alternative transcriptome. Nat Struct Mol Biol. 2020;27:717–25.
76. Malone B, Atanassov I, Aeschimann F, Li X, Grosshans H, Dieterich C. Bayesian prediction of RNA translation from ribosome profiling. Nucleic Acids Res. 2017;45:2960–72.
77. Liu Q, Shvarts T, Sliz P, Gregory RI. RiboToolkit: an integrated platform for analysis and annotation of ribosome profiling data to Decode mRNA translation at codon resolution. Nucleic Acids Res. 2020;48:W218–29.
78. Michel AM, Fox G, De Bo AMK, O'Connor C, Heaphy PB. GWIPS-viz: development of a ribo-seq genome browser. Nucleic Acids Res. 2014;42:D859–864.
79. Bartholomaus A, Kolte B, Mustafayeva A, Goebel I, Fuchs S, Benndorf D, et al. SmORFer: a modular algorithm to detect small ORF in prokaryotes. Nucleic Acids Res. 2021;49:e89.
80. van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, et al. Translational Landsc Hum Heart Cell. 2019;178:242–e26029.
81. Jackson R, Kroehling L, Khitun A, Bailis W, Jarret A, York AG, et al. The translation of non-canonical open reading frames controls mucosal immunity. Nature. 2018;564:434–8.
82. Cai T, Zhang Q, Wu B, Wang J, Li N, Zhang T, et al. LncRNA-encoded microproteins: A new form of cargo in cell culture-derived and Circulating extracellular vesicles. J Extracell Vesicles. 2021;10:e12123.
83. Fabre B, Combier JP, Plaza S. Recent advances in mass spectrometry-based peptidomics workflows to identify short-open-reading-frame-encoded peptides and explore their functions. Curr Opin Chem Biol. 2021;60:122–30.
84. Ma J, Diedrich JK, Jungreis I, Donaldson C, Vaughan J, Kellis M, et al. Improved identification and analysis of small open reading frame encoded polypeptides. Anal Chem. 2016;88:3967–75.
85. Yang Y, Wang H, Zhang Y, Chen L, Chen G, Bao Z, et al. An optimized proteomics approach reveals novel alternative proteins in mouse liver development. Mol Cell Proteom. 2023;22:100480.
86. Zhang Q, Wu E, Tang Y, Cai T, Zhang L, Wang J, et al. Deeply mining a universe of peptides encoded by long noncoding RNAs. Mol Cell Proteom. 2021;20:100109.

87. Wang B, Hao J, Pan N, Wang Z, Chen Y, Wan C. Identification and analysis of small proteins and short open reading frame encoded peptides in Hep3B cell. J Proteom. 2021;230:103965.
88. Wang B, Wang Z, Pan N, Huang J, Wan C. Improved identification of small open reading frames encoded peptides by Top-Down proteomic approaches and de Novo sequencing. Int J Mol Sci. 2021;22:5476.
89. Kaulich PT, Cassidy L, Bartel J, Schmitz RA, Tholey A. Multi-protease approach for the improved identification and molecular characterization of small proteins and short open reading Frame-Encoded peptides. J Proteome Res. 2021;20(5):2895–903.
90. D'Lima NG, Khitun A, Rosenbloom AD, Yuan P, Gassaway BM, Barber KW, et al. Comparative proteomics enables identification of nonannotated cold shock proteins in E. coli. J Proteome Res. 2017;16:3722–31.
91. Martinez TF, Lyons-Abbott S, Bookout AL, De Souza EV, Donaldson C, Vaughan JM, et al. Profiling mouse brown and white adipocytes to identify metabolically relevant small ORFs and functional microproteins. Cell Metab. 2023;35:166–e18311.
92. Bruderer R, Bernhardt OM, Gandhi T, Miladinovic SM, Cheng LY, Messner S, et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Mol Cell Proteom. 2015;14:1400–10.
93. van der Spek SJF, Gonzalez-Lozano MA, Koopmans F, Miedema SSM, Paliukhovich I, Smit AB, et al. Age-Dependent hippocampal proteomics in the APP/PS1 alzheimer mouse model: A comparative analysis with classical SWATH/DIA and directdia approaches. Cells. 2021;10:1588.
94. Li Y, Zhou H, Chen X, Zheng Y, Kang Q, Hao D, et al. SmProt: A reliable repository with comprehensive annotation of small proteins identified from ribosome profiling. Genomics Proteom Bioinf. 2021;19(4):602–10.
95. Olexiouk V, Van Criekinge W, Menschaert G. An update on sORFs.org: a repository of small ORFs identified by ribosome profiling. Nucleic Acids Res. 2018;46:D497–502.
96. Brunet MA, Lucier JF, Levesque M, Leblanc S, Jacques JF, Al-Saedi HRH, et al. OpenProt 2021: deeper functional annotation of the coding potential of eukaryotic genomes. Nucleic Acids Res. 2021;4:D380–8.
97. Sami A, Fu M, Yin H, Ali U, Tian L, Wang S, et al. NCPbook: A comprehensive database of noncanonical peptides. Plant Physiol. 2024;196:67–76.
98. Mohapatra S, Banerjee A, Rausseo P, Dragomir MP, Manyam GC, Broom BM et al. FuncPEP v2.0: An Updated Database of Functional Short Peptides Translated from Non-Coding RNAs. Noncoding RNA. 2024; 10:20.
99. Zhou X, Qin Y, Li J, Fan L, Zhang S, Zhang B, et al. LncPepAtlas: a comprehensive resource for exploring the translational landscape of long non-coding RNAs. Nucleic Acids Res. 2024;53:D468–76.
100. Huang W, Ling Y, Zhang S, Xia Q, Cao R, Fan X, et al. TransCirc: an interactive database for translatable circular RNAs based on multi-omics evidence. Nucleic Acids Res. 2021;49:D236–42.
101. Vitorino R, Guedes S, Trindade F, Correia I, Moura G, Carvalho P, et al. De Novo sequencing of proteins by mass spectrometry. Expert Rev Proteom. 2020;17:595–607.
102. Pan N, Wang Z, Wang B, Wan J, Wan C. Mapping microproteins and ncRNA-Encoded polypeptides in different mouse tissues. Front Cell Dev Biol. 2021;9:687748.
103. Yu R, Hu Y, Zhang S, Li X, Tang M, Yang M, et al. LncRNA CTBP1-DT-encoded microprotein DDUP sustains DNA damage response signalling to trigger dual DNA repair mechanisms. Nucleic Acids Res. 2022;50:8060–79.
104. Li M, Li X, Zhang Y, Wu H, Zhou H, Ding X, et al. Micropeptide MIAC inhibits HNSCC progression by interacting with Aquaporin 2. J Am Chem Soc. 2020;142:6708–16.
105. Zhang Q, Wei T, Yan L, Zhu S, Jin W, Bai Y, et al. Hypoxia-Responsive LncRNA AC115619 encodes a micropeptide that suppresses m6A modifications and hepatocellular carcinoma progression. Cancer Res. 2023;83:2496–512.
106. Zhang S, Zhang Z, Liu X, Deng Y, Zheng J, Deng J, et al. LncRNA-Encoded micropeptide ACLY-BP drives lipid deposition and cell proliferation in clear cell renal cell carcinoma via maintenance of ACLY acetylation. Mol Cancer Res. 2023;21:1064–78.
107. Du B, Zhang Z, Jia L, Zhang H, Zhang S, Wang H, et al. Micropeptide AF127577.4-ORF hidden in a LncRNA diminishes glioblastoma cell proliferation via the modulation of ERK2/METTL3 interaction. Sci Rep. 2024;14:12090.
108. Sun L, Wang W, Han C, Huang W, Sun Y, Fang K, et al. The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. Mol Cell. 2021;81:4493–e45089.
109. Ge Q, Jia D, Cen D, Qi Y, Shi C, Li J, et al. Micropeptide ASAP encoded by LINC00467 promotes colorectal cancer progression by directly modulating ATP synthase activity. J Clin Invest. 2021;131:e152911.
110. Wang Y, Wu S, Zhu X, Zhang L, Deng J, Li F, et al. LncRNA-encoded polypeptide ASRPS inhibits triple-negative breast cancer angiogenesis. J Exp Med. 2020;217:jem20190950.
111. Pei H, Dai Y, Yu Y, Tang J, Cao Z, Zhang Y, et al. The tumorigenic effect of LncRNA AFAP1-AS1 is mediated by translated peptide ATMLP under the control of m(6) A methylation. Adv Sci (Weinh). 2023;10:e2300314.
112. Zheng W, Guo Y, Zhang G, Bai J, Song Y, Song X, et al. Peptide encoded by LncRNA BVES-AS1 promotes cell viability, migration, and invasion in colorectal cancer cells via the SRC/mTOR signaling pathway. PLoS ONE. 2023;18(6):e0287133.
113. De Burbano S, Tran DDH, Allister AB, Polenkowski M, Nashan B, Koch M, et al. C20orf204, a hepatocellular carcinoma-specific protein interacts with nucleolin and promotes cell proliferation. Oncogenesis. 2021;10:31.
114. Polycarpou-Schwarz M, Gross M, Mestdagh P, Schott J, Grund SE, Hildenbrand C, et al. The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation. Oncogene. 2018;37:4750–68.
115. Guo B, Wu S, Zhu X, Zhang L, Deng J, Li F, et al. Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. EMBO J. 2020;39:e102190.
116. Dong Z, Chen X, Li J, Laster K, Zhang H, Huang Y et al. A Peptide Encoded by Long Non-coding RNA NEAT1 Suppresses Cancer Growth through Interfering RAFHSP90β Complex Stability. Preprint at https://doi.org/10.21203/rs.3.rs-3608223/v1(2023).
117. Li XL, Pongor L, Tang W, Das S, Muys BR, Jones MF, et al. A small protein encoded by a putative LncRNA regulates apoptosis and tumorigenicity in human colorectal cancer cells. Elife. 2020;9:e53734.
118. Tong X, Yu Z, Xing J, Liu H, Zhou S, Huang Y, et al. LncRNA HCP5-Encoded protein regulates ferroptosis to promote the progression of Triple-Negative breast Cancer. Cancers. 2023;15:1880.
119. Li Q, Guo G, Chen Y, Lu L, Li H, Zhou Z, et al. HCP5 derived novel microprotein triggers progression of gastric Cancer through regulating ferroptosis. Adv Sci (Weinh). 2024;11:e2407012.
120. Chen Y, Li Q, Yu X, Lu L, Zhou Z, Li M, et al. The microprotein HDSP promotes gastric cancer progression through activating the MECOM-SPINK1-EGFR signaling axis. Nat Commun. 2024;15:8381.
121. Huang JZ, Chen M, Chen D, Gao XC, Zhu S, Huang H, et al. A peptide encoded by a putative LncRNA HOXB-AS3 suppresses Colon cancer growth. Mol Cell. 2017;68:171–e184176.
122. Zhang H, Liao Z, Wang W, Liu Y, Zhu H, Liang H, et al. A micropeptide JunBP regulated by TGF-beta promotes hepatocellular carcinoma metastasis. Oncogene. 2023;42:113–23.
123. Xu W, Deng B, Lin P, Liu C, Li B, Huang Q, et al. Ribosome profiling analysis identified a KRAS-interacting microprotein that represses oncogenic signaling in hepatocellular carcinoma cells. Sci China Life Sci. 2020;63:529–42.
124. Tan Z, Zhao L, Huang S, Jiang Q, Wei Y, Wu JL, et al. Small peptide LINC00511-133aa encoded by LINC00511 regulates breast cancer cell invasion and stemness through the Wnt/β-catenin pathway. Mol Cell Probes. 2023;69:101913.
125. Pan J, Liu M, Duan X, Wang D. A short peptide LINC00665_18aa encoded by LncRNA LINC00665 suppresses the proliferation and migration of osteosarcoma cells through the regulation of the CREB1/RPS6KA3 interaction. PLoS ONE. 2023;18:e0286422.
126. Polenkowski M, Burbano de Lara S, Allister AB, Nguyen TNQ, Tamura T, Tran DDH. Identification of novel micropeptides derived from hepatocellular Carcinoma-Specific long noncoding RNA. Int J Mol Sci. 2021;23:58.
127. Tang C, Zhou Y, Sun W, Hu H, Liu Y, Chen L, et al. Oncopeptide MBOP encoded by LINC01234 promotes colorectal Cancer through MAPK signaling pathway. Cancers (Basel). 2022;14:2338.
128. Li M, Liu G, Jin X, Guo H, Setrerrahmane S, Xu X, et al. Micropeptide MIAC inhibits the tumor progression by interacting with AQP2 and inhibiting EREG/EGFR signaling in renal cell carcinoma. Mol Cancer. 2022;21:181.
129. Liu HT, Gao ZX, Li F, Guo XY, Li CL, Zhang H, et al. LncRNA LY6E-DT and its encoded metastatic-related protein play oncogenic roles via different pathways and promote breast cancer progression. Cell Death Differ. 2024;31:188–202.
130. Ye M, Gao R, Chen S, Bai J, Chen J, Lu F, et al. FAM201A encodes small protein NBASP to inhibit neuroblastoma progression via inactivating MAPK pathway mediated by FABP5. Commun Biol. 2023;6:714.

131. Yang L, Tang Y, He Y, Wang Y, Lian Y, Xiong F, et al. High expression of LINC01420 indicates an unfavorable prognosis and modulates cell migration and invasion in nasopharyngeal carcinoma. J Cancer. 2017;8:97–103.
132. D'Lima NG, Ma J, Winkler L, Chu Q, Loh KH, Corpuz EO, et al. A human microprotein that interacts with the mRNA decapping complex. Nat Chem Biol. 2017;13:174–80.
133. Zhai S, Lin J, Ji Y, Zhang R, Zhang Z, Cao Y, et al. 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. 2023;9:95.
134. Zhang C, Zhou B, Gu F, Liu H, Wu H, Yao F, et al. Micropeptide PACMP Inhibition elicits synthetic lethal effects by decreasing CtIP and poly(ADP-ribosyl) ation. Mol Cell. 2022;82:1297–e13121298.
135. Zhu KG, Yang J, Zhu Y, Zhu Q, Pan W, Deng S, et al. The microprotein encoded by Exosomal lncAKR1C2 promotes gastric cancer lymph node metastasis by regulating fatty acid metabolism. Cell Death Dis. 2023;14:708.
136. Wang X, Zhang H, Yin S, Yang Y, Yang H, Yang J, et al. lncRNA-encoded pep-AP attenuates the Pentose phosphate pathway and sensitizes colorectal cancer cells to oxaliplatin. EMBO Rep. 2022;23:e53140.
137. Zhou B, Wu Y, Cheng P, Wu C. Long noncoding RNAs with peptide-encoding potential identified in esophageal squamous cell carcinoma: KDM4A-AS1-encoded peptide weakens cancer cell viability and migratory capacity. Mol Oncol. 2023;17:1419–36.
138. Hu HF, Han L, Fu JY, He X, Tan JF, Chen QP, et al. LINC00982-encoded protein PRDM16-DT regulates CHEK2 splicing to suppress colorectal cancer metastasis and chemoresistance. Theranostics. 2024;14:3317–38.
139. Boix O, Martinez M, Vidal S, Gimenez-Alejandre M, Palenzuela L, Lorenzo-Sanz L, et al. pTINCR microprotein promotes epithelial differentiation and suppresses tumor growth through CDC42 sumoylation and activation. Nat Commun. 2022;13:6840.
140. Zhu S, Wang JZ, Chen D, He YT, Meng N, Chen M, et al. An oncopeptide regulates m(6)A recognition by the m(6)A reader IGF2BP1 and tumorigenesis. Nat Commun. 2020;11:1685.
141. Meng K, Lu S, Li YY, Hu LL, Zhang J, Cao Y, et al. LINC00493-encoded microprotein SMIM26 exerts anti-metastatic activity in renal cell carcinoma. EMBO Rep. 2023;24:e56282.
142. Li L, Shu XS, Geng H, Ying J, Guo L, Luo J, et al. A novel tumor suppressor encoded by a 1p36.3 LncRNA functions as a phosphoinositide-binding protein repressing AKT phosphorylation/activation and promoting autophagy. Cell Death Differ. 2023;30:1166–83.
143. Meng N, Chen M, Chen D, Chen XH, Wang JZ, Zhu S, et al. Small protein hidden in LncRNA LOC90024 promotes cancerous RNA splicing and tumorigenesis. Adv Sci (Weinh). 2020;7:1903233.
144. Morgado-Palacin L, Brown JA, Martinez TF, Garcia-Pedrero JM, Forouhar F, Quinn SA, et al. The TINCR ubiquitin-like microprotein is a tumor suppressor in squamous cell carcinoma. Nat Commun. 2023;14:1328.
145. Xu W, Liu C, Deng B, Lin P, Sun Z, Liu A, et al. TP53-inducible putative long noncoding RNAs encode functional polypeptides that suppress cell proliferation. Genome Res. 2022;32:1026–41.
146. Wu S, Guo B, Zhang L, Zhu X, Zhao P, Deng J, et al. A micropeptide XBP1SBM encoded by LncRNA promotes angiogenesis and metastasis of TNBC via XBP1s pathway. Oncogene. 2022;41:2163–72.
147. Jiang L, Yang J, Xu Q, Lv K, Cao Y. Machine learning for the micropeptide encoded by LINC02381 regulates ferroptosis through the glucose transporter SLC2A10 in glioblastoma. BMC Cancer. 2022;22:882.
148. Ruan X, Liu Y, Wang P, Liu L, Ma T, Xue Y, et al. RBMS3-induced circHECTD1 encoded a novel protein to suppress the vasculogenic mimicry formation in glioblastoma multiforme. Cell Death Dis. 2023;14:745.
149. Xia X, Li X, Li F, Wu X, Zhang M, Zhou H, et al. A novel tumor suppressor protein encoded by circular AKT3 RNA inhibits glioblastoma tumorigenicity by competing with active phosphoinositide-dependent Kinase-1. Mol Cancer. 2019;18:131.
150. Peng Y, Xu Y, Zhang X, Deng S, Yuan Y, Luo X, et al. A novel protein AXIN1-295aa encoded by circAXIN1 activates the Wnt/beta-catenin signaling pathway to promote gastric cancer progression. Mol Cancer. 2021;20:158.
151. Song R, Guo P, Ren X, Zhou L, Li P, Rahman NA, et al. A novel polypeptide CAPG-171aa encoded by circcapg plays a critical role in triple-negative breast cancer. Mol Cancer. 2023;22:104.
152. Li F, Tang H, Zhao S, Gao X, Yang L, Xu J. Circ-E-Cad encodes a protein that promotes the proliferation and migration of gastric cancer via the TGF-beta/Smad/C-E-Cad/PI3K/AKT pathway. Mol Carcinog. 2023;62:360–8.
153. Gao X, Xia X, Li F, Zhang M, Zhou H, Wu X, et al. Circular RNA-encoded oncogenic E-cadherin variant promotes glioblastoma tumorigenicity through activation of EGFR-STAT3 signalling. Nat Cell Biol. 2021;23:278–91.
154. Li H, Lan T, Liu H, Liu C, Dai J, Xu L, et al. IL-6-induced cGGNBP2 encodes a protein to promote cell growth and metastasis in intrahepatic cholangiocarcinoma. Hepatology. 2022;75:1402–19.
155. Wang H, Liang Y, Zhang T, Yu X, Song X, Chen Y, et al. C-IGF1R encoded by cIGF1R acts as a molecular switch to restrict mitophagy of drug-tolerant persister tumour cells in non-small cell lung cancer. Cell Death Differ. 2023;30:2365–81.
156. Pan Z, Zheng J, Zhang J, Lin J, Lai J, Lyu Z, et al. A novel protein encoded by Exosomal CircATG4B induces oxaliplatin resistance in colorectal Cancer by promoting autophagy. Adv Sci (Weinh). 2022;9:e2204513.
157. Wang Q, Cheng B, Singh S, Tao Y, Xie Z, Qin F, et al. A protein-encoding CCDC7 circular RNA inhibits the progression of prostate cancer by up-regulating FLRT3. NPJ Precis Oncol. 2024;8:11.
158. Liu T, Ma T, Xue J, Zhu L, Zhao W, Sun J, et al. Circular RNA circDDX17 suppression to gastric cancer progression via the sponging miR-1208/miR- 1279/FKBP5 axis and encodes a novel circDDX17- 63aa protein. Preprient at. 2023. https://doi.org/10.21203/rs.3.rs-3288567/v1.
159. Pan Z, Cai J, Lin J, Zhou H, Peng J, Liang J, et al. A novel protein encoded by circFNDC3B inhibits tumor progression and EMT through regulating snail in colon cancer. Mol Cancer. 2020;19:71.
160. Wang P, Hu Z, Yu S, Su S, Wu R, Chen C, et al. A novel protein encoded by circFOXP1 enhances ferroptosis and inhibits tumor recurrence in intrahepatic cholangiocarcinoma. Cancer Lett. 2024;598:217092.
161. Xiong L, Liu HS, Zhou C, Yang X, Huang L, Jie HQ, et al. A novel protein encoded by circINSIG1 reprograms cholesterol metabolism by promoting the ubiquitin-dependent degradation of INSIG1 in colorectal cancer. Mol Cancer. 2023;22:72.
162. Withdrawn. circLgr4 drives colorectal tumorigenesis and invasion through Lgr4-targeting peptide. Int J Cancer. 2022;150:E3.
163. Yu S, Su S, Wang P, Li J, Chen C, Xin H, et al. Tumor-associated macrophage-induced circmrckalpha encodes a peptide to promote Glycolysis and progression in hepatocellular carcinoma. Cancer Lett. 2024;591:216872.
164. Li P, Song R, Yin F, Liu M, Liu H, Ma S, et al. circMRPS35 promotes malignant progression and cisplatin resistance in hepatocellular carcinoma. Mol Ther. 2022;30:431–47.
165. Zheng X, Chen L, Zhou Y, Wang Q, Zheng Z, Xu B, et al. A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling. Mol Cancer. 2019;18:47.
166. Bai J, Meng X, Wu Q, Cao C, Yang W, Chu S, et al. A novel peptide encoded by circsrcap confers resistance to enzalutamide by inhibiting the Ubiquitin-Dependent degradation of AR-V7 in Castration-Resistant prostate Cancer. J Transl Med. 2025;23:108.
167. Lu J, Ru J, Chen Y, Ling Z, Liu H, Ding B, et al. N(6) -methyladenosine-modified circSTX6 promotes hepatocellular carcinoma progression by regulating the HNRNPD/ATF3 axis and encoding a 144 amino acid polypeptide. Clin Transl Med. 2023;13:e1451.
168. Lyu Y, Tan B, Li L, Liang R, Lei K, Wang K, et al. A novel protein encoded by circUBE4B promotes progression of esophageal squamous cell carcinoma by augmenting MAPK/ERK signaling. Cell Death Dis. 2023;14:346.
169. Song R, Ma S, Xu J, Ren X, Guo P, Liu H, et al. A novel polypeptide encoded by the circular RNA ZKSCAN1 suppresses HCC via degradation of mTOR. Mol Cancer. 2023;22:16.
170. Zhao W, Zhang Y, Zhu Y. Circular RNA circbeta-catenin aggravates the malignant phenotype of non-small-cell lung cancer via encoding a peptide. J Clin Lab Anal. 2021;35:e23900.
171. Li F, Cai Y, Deng S, Yang L, Liu N, Chang X, Jing L, et al. A peptide CORO1C-47aa encoded by the circular noncoding RNA circ-0000437 functions as a negative regulator in endometrium tumor angiogenesis. J Biol Chem. 2021;297:101182.
172. Li Y, Wang Z, Su P, Liang Y, Li Z, Zhang H, et al. circ-EIF6 encodes EIF6-224aa to promote TNBC progression via stabilizing MYH9 and activating the Wnt/beta-catenin pathway. Mol Ther. 2022;30:415–30.
173. Yang Y, Gao X, Zhang M, Yan S, Sun C, Xiao F, et al. Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis. J Natl Cancer Inst. 2018;110:304–15.
174. Song J, Zheng J, Liu X, Dong W, Yang C, Wang D, et al. A novel protein encoded by ZCRB1-induced circHEATR5B suppresses aerobic Glycolysis of GBM through phosphorylation of JMJD5. J Exp Clin Cancer Res. 2022;41:171.

175. Jiang T, Xia Y, Lv J, Li B, Li Y, Wang S, et al. A novel protein encoded by circMAPK1 inhibits progression of gastric cancer by suppressing activation of MAPK signaling. Mol Cancer. 2021;20:66.
176. Lei K, Liang R, Liang J, Lu N, Huang J, Xu K, et al. CircPDE5A-encoded novel regulator of the PI3K/AKT pathway inhibits esophageal squamous cell carcinoma progression by promoting USP14-mediated de-ubiquitination of PIK3IP1. J Exp Clin Cancer Res. 2024;43:124.
177. Xiang X, Fu Y, Zhao K, Miao R, Zhang X, Ma X, et al. Cellular senescence in hepatocellular carcinoma induced by a long non-coding RNA-encoded peptide PINT87aa by blocking FOXM1-mediated PHB2. Theranostics. 2021;11:4929–44.
178. Zhang M, Huang N, Yang X, Luo J, Yan S, Xiao F, et al. A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. Oncogene. 2018;37:1805–14.
179. Chang S, Ren D, Zhang L, Liu S, Yang W, Cheng H, et al. Therapeutic SHPRH-146aa encoded by circ-SHPRH dynamically upregulates P21 to inhibit CDKs in neuroblastoma. Cancer Lett. 2024;598:217120.
180. Wu X, Xiao S, Zhang M, Yang L, Zhong J, Li B, et al. A novel protein encoded by circular SMO RNA is essential for Hedgehog signaling activation and glioblastoma tumorigenicity. Genome Biol. 2021;22:33.
181. Wu X, Sun G, Fan R, Liu K, Duan C, Mao X, et al. CircSP3 encodes SP3-461aa to promote CcRCC progression via stabilizing MYH9 and activating the PI3K-Akt signaling pathway. J Cancer. 2024;15:5876–96.
182. Li Y, Wang Z, Yang J, Sun Y, He Y, Wang Y, et al. CircTRIM1 encodes TRIM1-269aa to promote chemoresistance and metastasis of TNBC via enhancing CaM-dependent MARCKS translocation and PI3K/AKT/mTOR activation. Mol Cancer. 2024;23:102.
183. Wang S, Wang Y, Li Q, Li X, Feng X, Zeng K. The novel beta-TrCP protein isoform hidden in circular RNA confers trastuzumab resistance in HER2-positive breast cancer. Redox Biol. 2023;67:102896.
184. Aktas T, Avsar Ilik I, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, et al. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. Nature. 2017;544:115–9.
185. Song Z, Lin J, Su R, Ji Y, Jia R, Li S, et al. eIF3j inhibits translation of a subset of circular RNAs in eukaryotic cells. Nucleic Acids Res. 2022;50:11529–49.
186. Huang H, Weng H, Sun W, Qin X, Shi H, Wu H, et al. Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. Nat Cell Biol. 2018;20:285–95.
187. Rajagopalan LE, Westmark CJ, Jarzembowski JA, Malter JS. HnRNP C increases amyloid precursor protein (APP) production by stabilizing APP mRNA. Nucleic Acids Res. 1998;26:3418–23.
188. Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature. 2015;518:560–4.
189. Li F, Yang K, Gao X, Zhang M, Gu D, Wu X, et al. A peptide encoded by upstream open reading frame of MYC binds to Tropomyosin receptor kinase B and promotes glioblastoma growth in mice. Sci Transl Med. 2024;16:eadk9524.
190. Susin SA, Lorenzo HK, Zamzami N, Marzo I, Snow BE, Brothers GM, et al. Molecular characterization of mitochondrial apoptosis-inducing factor. Nature. 1999;397:441–6.
191. Saharinen P, Eklund L, Pulkki K, Bono P, Alitalo K. VEGF and angiopoietin signaling in tumor angiogenesis and metastasis. Trends Mol Med. 2011;17:347–62.
192. Johnson DE, O'Keefe RA, Grandis JR. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. Nat Rev Clin Oncol. 2018;15:234–48.
193. Bergers G, Fendt SM. The metabolism of cancer cells during metastasis. Nat Rev Cancer. 2021;21:162–80.
194. Lu J. The Warburg metabolism fuels tumor metastasis. Cancer Metastasis Rev. 2019;38:157–64.
195. Fang E, Wang X, Wang J, Hu A, Song H, Yang F, et al. Therapeutic targeting of YY1/MZF1 axis by MZF1-uPEP inhibits aerobic Glycolysis and neuroblastoma progression. Theranostics. 2020;10:1555–71.
196. Zhang M, Zhao Y, Liu X, Ruan X, Wang P, Liu L, et al. Pseudogene MAPK6P4-encoded functional peptide promotes glioblastoma vasculogenic mimicry development. Commun Biol. 2023;6:1059.
197. Currie E, Schulze A, Zechner R, Walther TC, Farese RV. Jr. Cellular fatty acid metabolism and cancer. Cell Metab. 2013;18:153–61.
198. Liu Z, Liu W, Wang W, Ma Y, Wang Y, Drum DL, et al. CPT1A-mediated fatty acid oxidation confers cancer cell resistance to immune-mediated cytolytic killing. Proc Natl Acad Sci U S A. 2023;120:e2302878120.
199. Klinge CM. Estrogenic control of mitochondrial function. Redox Biol. 2020;31:101435.
200. Zeng L, Zheng W, Zhang J, Wang J, Ji Q, Wu X, et al. An epitope encoded by uORF of RNF10 elicits a therapeutic anti-tumor immune response. Mol Ther Oncolytics. 2023;31:100737.
201. Kikuchi Y, Tokita S, Hirama T, Kochin V, Nakatsugawa M, Shinkawa T, et al. CD8(+) T-cell immune surveillance against a tumor antigen encoded by the oncogenic long noncoding RNA PVT1. Cancer Immunol Res. 2021;9:1342–53.
202. Qiu X, Zheng Q, Luo D, Ming Y, Zhang T, Pu W, et al. Rational design, synthesis, and biological evaluation of novel c-Met degraders for lung cancer therapy. J Med Chem. 2025;68(3):2815–39.

Publisher's note