LncRNA-encoded peptides in cancer

Yaguang Zhang[1*]

Abstract

Long non-coding RNAs (lncRNAs), once considered transcriptional noise, have emerged as critical regulators of gene expression and key players in cancer biology. Recent breakthroughs have revealed that certain lncRNAs can encode small open reading frame (sORF)-derived peptides, which are now understood to contribute to the pathogenesis of various cancers. This review synthesizes current knowledge on the detection, functional roles, and clinical implications of lncRNA-encoded peptides in cancer. We discuss technological advancements in the detection and validation of sORFs, including ribosome profiling and mass spectrometry, which have facilitated the discovery of these peptides. The functional roles of lncRNA-encoded peptides in cancer processes such as gene transcription, translation regulation, signal transduction, and metabolic reprogramming are explored in various types of cancer. The clinical potential of these peptides is highlighted, with a focus on their utility as diagnostic biomarkers, prognostic indicators, and therapeutic targets. The challenges and future directions in translating these findings into clinical practice are also discussed, including the need for large-scale validation, development of sensitive detection methods, and optimization of peptide stability and delivery.

Keywords Long non-coding RNA, Small open reading frame, Peptide, Cancer, Application

Introduction

Long non-coding RNAs (lncRNAs) were initially defined as a class of RNAs longer than 200 nucleotides that do not encode proteins [1–3]. Initially regarded as "noise" from genome transcription, lncRNAs have increasingly been shown to play important roles in the regulation of gene expression at the epigenetic, transcriptional, and post-transcriptional levels [4–8]. They have also been found to be intimately linked with the occurrence and progression of a spectrum of human diseases, with a particularly significant association observed in the context of cancer [9–16].

With the advancement of proteomics and translation technologies, it has been discovered that some lncRNAs

have the ability to encode small peptides or micropeptides [17–21]. These peptides are encoded by small or short open reading frames (sORFs) and can range in length from tens to over a hundred amino acids (aa) [22–25]. Furthermore, there is growing evidence that peptides derived from lncRNAs have specific biological functions and can act as oncogenic drivers or tumor suppressors [26–29]. They play important roles in various cancer processes, such as transcriptional regulation, post-transcriptional regulation, translation and post-translational regulation, signal transduction, and cancer metabolism [8, 30–34].

Despite numerous reviews on lncRNA-encoded peptides published in the past, most are outdated as they were released several years ago [26, 35, 36]. LncRNA-encoded peptides have emerged as a hot topic in recent years, with many new discoveries identifying novel lncR-NAs that encode for peptides and their significant functions in cancer, along with new regulatory mechanisms. In this review, we review the methods for detecting lncRNA-encoded peptides, comparing their differences. Additionally, we systematically summarize the lncRNAs

---

*Correspondence:
Yaguang Zhang
zhangyaguang20085155@126.com
[1] Laboratory of Gastrointestinal Tumor Epigenetics and Genomics, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Chengdu 610041, People's Republic of China

known to encode peptides, their roles and mechanisms in cancer, with a particular focus on regulatory mechanisms that have not been systematically reviewed before. Finally, we explore the potential applications of these peptides. Overall, this review aims to provide a comprehensive and systematic resource for future researchers in the field of lncRNA-encoded peptides.

Biogenesis and detection of lncRNA-encoded peptides

Biogenesis of lncRNA-encoded peptides

The biogenesis of peptides encoded by lncRNAs is a multifaceted process that encompasses the transcription of lncRNAs by RNA polymerase II, followed by their maturation, which includes the addition of a 5' cap (m7G) and a polyadenylated tail. After undergoing alternative splicing, these transcripts are exported to the cytosol, where they harbor the potential to be translated into peptides [37, 38]. Notably, the work of Yu et al. has shown that DNA damage can prompt ribosomes to associate with the internal ribosome entry site (IRES) region within the lncRNA CTBP1-DT. This interaction bypasses the inhibitory effects of upstream open reading frames (uORFs) and triggers the cap-independent translation of a novel microprotein termed DNA damage-upregulated protein (DDUP) [39]. Some researchers have raised skepticism, suggesting that mere RNA structure and ribosome binding are not definitive indicators of a transcript's translatability, making the elucidation of their translational mechanisms a challenging endeavor [40]. Moreover, the translation of lncRNA-encoded peptides, despite their brevity, is also contingent upon the presence of open reading frames (ORFs). As we know, ORFs are nucleotide sequences that span from a start codon to the nearest stop codon within a nucleotide sequence. A translatable ORF is typically recognized as the coding DNA sequence (CDS) on an mRNA that gives rise to its principal protein product. In mRNA, codons-triads of nucleotides-correspond to specific amino acids, with the AUG codon typically serving as the start signal and UAA, UAG, and UGA being the traditional stop codons in eukaryotic organisms. sORFs, typically less than 100 codons in length [41], are sometimes extended to include sORFs of 200–250 codons as described in various studies [42–44]. These sORFs are distinguished by their size from all other ORFs, but not all sORFs are translated or are indeed translatable. Identifying any ORF within genomic DNA is straightforward, but differentiating between coding and non-coding sORFs is more complex. Most de novo gene prediction algorithms differentiate coding from non-coding sequences by recognizing genomic patterns indicative of features (such as start codons, stop sites, splice junctions, promoters, and polyadenylation signals) or by

analyzing intrinsic DNA sequence properties (including codon usage bias, nucleotide composition, and in-frame hexamer frequency) [45, 46]. However, these algorithms are not optimized for sORFs, as they focus on longer ORFs with a higher prevalence of these features [47, 48]. As a result, many gene annotation tools overlook ORFs shorter than 100 codons, often dismissing them as insignificant [49, 50]. However, with the advancement of technology, the challenge has begun to be addressed effectively. Several approaches have been taken to systematically predict sORFs with coding potential (Table 1). For example, Lin et al. presented PhyloCSF, a new computational method examining evolutionary conservation of a sORF across species [51]. Camargo et al. employed RNAsamba, a sophisticated bioinformatics tool that predicts the coding potential of RNA molecules from sequence information alone [52]. Utilizing a neural network-based algorithm, RNAsamba identifies patterns that distinguish coding transcripts from non-coding ones, offering a promising avenue for sORF prediction [52]. These resources have significantly expanded our understanding of the coding potential within sORFs.

While the current research has shed preliminary light on this topic, it is clear that the regulatory mechanisms by which ORFs within lncRNAs are translated into peptides are not yet fully understood. There is a pressing need for future studies to delve deeper into these mechanisms, providing a more comprehensive understanding of the translation process of lncRNA-encoded peptides. To date, a considerable number of lncRNA-encoded peptides have been identified. In response to this growing body of information, several databases have been developed and are now accessible to researchers seeking data related to lncRNA-encoded peptides (Table 2). These databases serve as valuable repositories, enabling investigators to rapidly access information on lncRNA-encoded peptides of interest and providing a rich resource for the scientific community.

Detection of lncRNA-encoded peptides

With technological progress, there are several methods available to predict and validate the coding potential of translated small open reading frames, including bioinformatics, ribosome profile sequencing (Ribo-seq), reporter tag, epitope tagging, antibody-based validation and mass spectrum (MS) (Fig. 1). These methods are often combined with protein detection procedures such as western blotting, immunocytochemistry, or immunoprecipitation steps to verify the translation of sORFs [71, 72].

Ribosome profile sequencing

Ribosomes are one of the fundamental components of the translation process in eukaryotic cells [73, 74].

Table 1 Tools for predicting sORFs

Method	Website	Year	Characteristics	References
PhyloCSF	http://compbio.mit.edu/PhyloCSF	2011	PhyloCSF is a comparative genomics method analyzing a multispecies nucleotide sequence alignment to assess the likelihood of it being a conserved region that codes for proteins, via a statistical comparison of phylogenetic codon models	[51]
CPAT	https://rna-cpat.sourceforge.net/	2013	CPAT applies a logistic regression model built with open reading frame size, open reading frame coverage, Fickett TESTCODE statistic and hexamer usage bias, to rapidly recognizes coding and noncoding transcripts from a large pool of candidates	[53]
PLEK	https://sourceforge.net/projects/plek/files/	2014	PLEK utilizes a improved k-mer framework coupled with a support vector machine technique to differentiate lncRNAs from mRNAs without the need for genomic sequences or existing annotations	[54]
riboHMM	https://github.com/rajanil/riboHMM	2016	riboHMM uses hidden Markov models to accurately resolve the precise set of RNA sequences that are being translated in a particular cell, by analyzing a ribosome profiling test, sequenced data from an RNA-seq, and the RNA sequence itself	[55]
Rp-Bp	https://github.com/dieterich-lab/rp-bp	2017	Rp-Bp is an unsupervised Bayesian approach to predict translated ORFs from ribosome profiles, characterizing with incorporating and propagates uncertainty in the prediction process, and automatic Bayesian selection of read lengths and ribosome P-site offsets	[56]
CPC2	http://cpc2.cbi.pku.edu.cn	2017	CPC2 runs fast and exhibits superior accuracy, especially for long non-coding transcripts, with species-neutral, making it feasible for ever-growing non-model organism transcriptomes	[57]
RiboCode	https://github.com/xryanglab/RiboCode	2018	RiboCode is a statistically vigorous method for the de novo annotation of the full translatome by quantitatively assessing the 3-nt periodicity, based on ribosome profiling data	[58]
ORFFinder	https://www.ncbi.nlm.nih.gov/orffinder/	2002	ORFFinder selects open reading frames from randomly fragmented genomic DNA fragments	[59]
RNAsamba	https://rnasamba.lge.ibi.unicamp.br/	2020	A tool to predict the coding potential of RNA molecules from sequence information using a neural network-based that models both the whole sequence and the ORF to identify patterns that distinguish coding from non-coding transcripts	[52]
LncCat	http://cczubio.top/lnccat	2023	LncCat combines five types of features to encode transcript sequences and employs CatBoost to build a prediction model	[60]
sORFPred	https://github.com/orangewindczw/sORFPred	2023	A method based on comprehensive features and ensemble learning to predict the sORFs in plant lncRNAs	[61]

Undoubtedly, Ribo-seq is one of the most promising scientific evidences that could point towards answering which lncRNAs are capable of encoding peptides [75–77]. It is an emerging technique that offers a glimpse into protein synthesis through the deep sequencing of RNA fragments protected by ribosomes [78, 79]. The core Ribo-seq methodology employs RNase I to digest unprotected, single-stranded RNA, leaving behind ribosome-protected fragments (RPFs). These fragments are then isolated, sequenced, and mapped to the genome, allowing for the assembly of transcripts and the discovery of novel sORFs with coding potential [80, 81]. However, this technique faces challenges, including reliance on next-generation sequencing, which can introduce false positives due to sequencing quality and depth. Additionally, Ribo-seq may not capture all sORFs due to their variable expression across conditions, stages, and tissues [82]. It also requires a significant starting material, such as 10 million cells, to meet sequencing RNA requirements [83].

Notably, Xiong et al. have recently developed an ultrasensitive Ribo-seq method, termed Ribo-lite, which can be applied to ultra-low input oocytes, even single oocytes [84]. Zou and colleagues have successfully applied this method to investigate the translational regulation during human oocyte maturation and early embryonic development [85]. However, the applicability of this method for single-cell translational regulation analysis in other cell types has yet to be reported. Furthermore, In the field of ribosome profiling, the specificity of RNase I for single-stranded RNA is a well-established fact. Notably, this specificity also introduces a potential pitfall. Given that RNase I cannot target double-stranded RNA regions, such as those found in the stem-loop structures of micro-RNA precursors [86], there exists a risk of inadvertently generating pseudo ribosomal footprints (pseudo-RPFs) from these complex structures. While it is true that double-stranded RNAs are not commonly encountered within the cellular milieu, their presence, albeit rare,

Table 2 Databases for lncRNA-encoded peptide

Database	Website	Year	Species	Characteristics	References
ncEP database	http://www.jianglab.cn/ncEP/	2020	18	ncEP contains 80 entries including 74 proteins or peptides, 22 lncRNAs, 11 circRNAs, 9 pri-miRNAs and 37 other ncRNAs across 18 species from more than 50 articles of over 2000 candidate articles	[62]
FuncPEP	https://bioinformatics.mdanderson.org/Supplements/FuncPEP/	2022	12	FuncPEP includes a comprehensive annotation of 112 functional ncPEPs from 44 studies and specific details regarding the ncRNA transcripts that encode these peptides	[63]
TransLnc	http://bio-bigdata.hrbmu.edu.cn/TransLnc/	2022	3	The current version of TransLnc contains a total of 583,840 computationally predicted peptides for 33,094 translatable lncRNAs across three different species (Human, Mouse, Rat), 381,105 experimentally verified lncRNA peptides across 34 tissues, and 391,418 candidate neoantigens	[64]
sORFs.org	http://www.sorfs.org	2016	3	A novel database for sORFs identified using ribosome profiling. At present, sORFs.org harbors 263 354 sORFs that demonstrate ribosome occupancy, originating from three different cell lines: HCT116 (human), E14_mESC (mouse) and S2 (fruit fly)	[65]
SmProt	http://bigdata.ibp.ac.cn/SmProt/	2021	8	SmProt incorporated 638,958 unique small proteins curated from 3,165,229 primary records, which were computationally predicted from 419 Ribo-seq datasets or collected from literature and other sources from 370 cell lines or tissues in 8 species	[66]
MetamORF	https://metamorf.hb.univ-amu.fr/	2021	2	MetamORF describes 664 771 and 497 904 unique ORFs in the human and mouse genomes, respectively, providing at least the information necessary to locate the ORF on the genome, its sequence and the gene it is located on (excepted for intergenic ORFs)	[67]
SPENCER	http://spencer.renlab.org	2022	1	SPENCER amassed 2,806 mass spectrometry data points from 55 studiess, comprising 1,007 cancerous and 719 healthy samples. Through its proteomics pipeline leveraging MS, it detected 29,526 ncPEPs in 15 varied cancer types, with 22,060 such ncPEPs confirmed by subsequent experimental research	[68]
cncRNAdb	http://www.rna-society.org/cncrnadb/	2021	21	This current version of cncRNAdb documents 2598 manually curated cncRNA-associated function entries with experimental evidence (including 1,936 translated ncRNA entries and 662 untranslated mRNA entries) involving 2002 coding and noncoding RNA (including 1358 translated ncRNAs and 644 untranslated mRNAs) across 21 species	[69]
LncPep	http://www.shenglilabs.com/LncPep/	2021	39	LncPep totally covers 10,580,228 peptides translated from 883,804 lncRNAs across 39 different species, and we applied 7 evidence including Mass Spectrometry, ribosome profiling, Pfam, translation initiation site, N6-methyladenosine modification of RNA sites, CPC2, and CPAT for users to explore and evaluated the predicted peptides	[70]

cannot be entirely discounted. This rarity does not eliminate the possibility that they might contribute to false-positive signals in ribosome profiling assays, thereby complicating the interpretation of the resulting data. Consequently, researchers must exercise caution when analyzing ribosome profiling data to ensure that the observed ribosomal footprints are indeed indicative of active translation events rather than artifacts stemming from the presence of double-stranded RNA structures. The short length of RPFs, approximately 30 nucleotides, complicates the differentiation of transcript isoforms resulting from alternative splicing [87]. It's important to note that ribosomal occupancy does not automatically indicate translation of the ORF [88], as it has been shown that start codons can regulate translation attenuation of a downstream ORF, mRNA availability through nonsense-mediated decay [40].
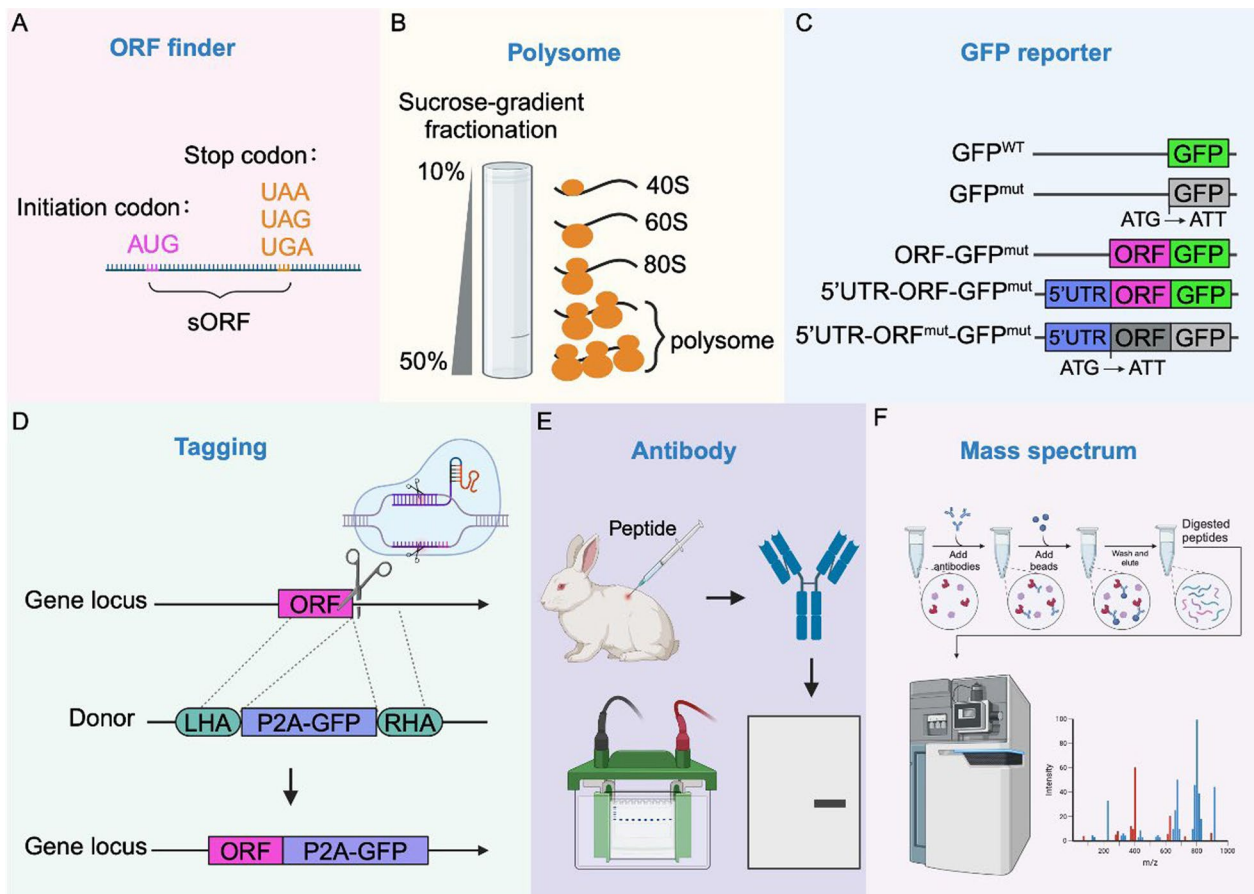
Fig. 1 Detection methods for the coding potential of lncRNAs. **A** Prediction of short open reading frames (ORFs) within lncRNAs. **B** Sucrose density gradient separation to detect ribosome enrichment on lncRNAs. **C** Detection of GFP translation using a GFP fusion with a mutated start codon within a lncRNA ORF. **D** Integration of a tag at the lncRNA ORF site using gene editing technology to assess the expression of the tagged protein. LHA, left homologous arm; RHA, right homologous arm. **E** Detection of intracellular lncRNA-encoded peptides using antibodies raised against synthetic peptides. **F** Mass spectrometry identification of peptide expression. Image created with BioRender.com

Reporter tags

The coding potential of sORFs can be evaluated by fusing them with reporter tags and then detecting the signal through immunoblotting or microscopy [34]. Specifically, a FLAG/HA-tag system is genetically engineered to be cloned immediately preceding the stop codon of the sORF under investigation. This fusion sequence, which includes the FLAG/HA-tag, is subsequently inserted into a plasmid vector, which then serves as the template for in vitro cell transfection. Upon transfection into the target cell line, the expression of the FLAG/HA-tagged micropeptide is quantified using western blotting and immunofluorescence assays with anti-FLAG/HA tag antibodies [89, 90]. As an alternative approach, sORFs derived from lncRNAs can be fused to the N-terminus of GFP vectors. The expression levels of the GFP-tagged micropeptides are then evaluated using western blotting, fluorescence microscope or immunofluorescence assays with anti-GFP antibodies, providing a visual and

quantitative assessment of the micropeptide's presence and distribution within the cells [91–93]. It should be noted that inserting a reporter tag internally or at the N-terminus of micropeptide carries the risk of disrupting the protein's function, as well as its intramolecular interactions and folding [25, 94–96]. This possibility underscores the need for careful experimental design and the interpretation of results with an awareness of potential artifacts introduced by the tagging process.

Epitope tagging

Epitope tagging is a method that incorporates a recognizable epitope tag into a protein sequence, allowing for specific and sensitive detection of sORF using available antibodies [97]. In the context of sORFs encoded within lncRNAs, the CRISPR-Cas9 system presents a powerful tool for the site-specific introduction of epitope tags. The CRISPR-Cas9 system can be programmatically designed to target the stop codon of the lncRNA locus in question

within the genome of the cells. By designing a guide RNA that directs the Cas9 nuclease to the desired location, researchers can introduce an epitope tag at the stop codon, effectively tagging the sORF for detection purposes. Once the epitope tag is integrated into the lncRNA locus, the expression of the resulting micropeptides can be assessed using Western blotting, fluorescence microscope or immunofluorescence assays with corresponding anti-tag antibodies [98, 99]. This approach effectively validates the coding potential of lncRNAs. However, several challenges must be considered when using epitope tagging. Firstly, the insertion of an epitope tag has the potential to disrupt the native structure and function of the protein, which could lead to misinterpretation of the protein's behavior in cellular assays. Secondly, the efficiency of tag integration can vary, and off-target effects may occur with the CRISPR-Cas9 system, potentially tagging unintended sites. Additionally, the detection of the tagged protein relies on the availability and specificity of antibodies, which may sometimes result in high background signals or false negatives.

Antibody-based validation

Antibody-based validation is a critical process in the identification and characterization of sORF-encoded polypeptides (SEPs). This approach involves the synthesis of antibodies that are specific to the predicted sequences of SEPs, allowing for the detection and confirmation of these peptides within complex cellular environments through western blotting [25]. For example, Faure et al. employed a monoclonal antibody directed against the Gau protein, a peptide approximately 100 amino acids long, to confirm the existence and functionality of the Gau protein [100]. Nonetheless, developing antibodies against SEPs presents significant challenges, primarily due to the small size of sORFs. Additionally, detecting SEPs can be problematic when they are expressed at low levels, as elevated antibody signals may not be easily discernible [25]. Therefore, ongoing efforts to refine antibody-based validation techniques will be essential for the discovery and characterization of new SEPs and the elucidation of their biological functions.

Mass spectrometry

Mass spectrometry is a sophisticated analytical technique that has proven indispensable in the field of proteomics, offering unparalleled capabilities for the identification and quantification of proteins and peptides, which provide direct evidence of sORFs' translation into SEPs [88, 101]. This method is often paired with the immunoprecipitation of ORF-GFP fusion peptides, leveraging anti-GFP antibodies to precipitate GFP-tagged SEPs from cell lysates. This approach not only detects unannotated proteins but also confirms the translation of sORFs into peptides, which not only detects unannotated proteins but also verifies the translation of sORFs into peptides [25, 102]. However, while mass spectrometry is adept at peptide detection, it has limitations in identifying SEPs due to their short length and a propensity for producing tryptophan-containing peptides. Furthermore, low-abundance SEPs can be overlooked during sample preparation [76, 88]. Therefore, special attention must be given to the separation and concentration steps of peptides, which are crucial for detecting small and/or low-abundance products in cell lysates [76].

The conditions, difficulty levels, and reliability of the aforementioned methods for detecting lncRNA-encoded peptides are encapsulated in Table 3. It is important to note that affirming the coding potential of an lncRNA necessitates a multifaceted approach, employing multiple methods to ascertain its function and to circumvent the possibility of false positives. This underscores the importance of a rigorous and integrated methodological strategy in validating the biological significance of lncRNA-encoded peptides.

Functions of lncRNA-encoded peptides in cancer

lncRNAs serve multifaceted roles, constructing intricate regulatory systems and engaging in a spectrum of biological activities. While numerous sORFs within lncRNAs and their corresponding short peptides have been detected using the methods previously described, the functional assignments for these peptides remain scarce. Emerging research suggests that the micropeptides derived from lncRNAs could be pivotal in tumorigenesis and tumor progression. In this section, we provide a compilation of lncRNA-encoded peptides that are associated with various cancer-related biological processes (Table 4).

Colorectal cancer

Colorectal cancer (CRC), the second most prevalent cancer in women and third in men globally, is a major contributor to cancer-related mortality, accounting for 9.2% of such deaths [133, 134]. The exploration of lncRNA-encoded peptides has unveiled their pivotal role in the molecular intricacies of CRC, influencing its development, progression, and response to treatment (Table 4 and Fig. 2). For example, the HOXB-AS3 peptide, typically down-regulated in colon cancer, can inhibit cancer growth by interfering with PKM splicing and glucose metabolism (Fig. 2A) [33], while the SRSP peptide promotes cancer cell proliferation and metastasis by affecting the splicing of transcription factor Sp4 (Fig. 2B) [105]. The RBRP peptide, upregulated in metastatic CRC, stabilizes c-Myc mRNA by binding to IGF2BP1, enhancing

Table 3 Condition, difficulty degree and reliability of methods for detecting lncrNA-encoded peptides

Method	Condition	Difficulty degree	Reliability
ORF finder	Internet	Easy	Poor
Polysome	Ultracentrifuge	Normal	Poor
	Fully automatic density gradient preparation system		
	Automatic separation system		
GFP reporter	Plasmid construction	Normal	General
	Inverted fluorescence microscope		
	Sequencing (selectable)		
Tagging	CRISPER/Cas9 system	Hard	Good
	Homologous arm construction		
	Inverted fluorescence microscopy		
	Sequencing		
	Flow cytometer (selectable)		
Antibody	Antibody preparation	Hard	Good
	Electrophoretic system		
	Chemiluminescence imager system		
Mass spectrum	Mass spectrometer	Normal	Good

tumor progression (Fig. 2C) [104]. These results underscore the post-transcriptional regulatory potential of lncRNA products in CRC. In the context of tumor metabolism, the overexpression of ASAP boosts ATP synthase activity and mitochondrial oxygen consumption, promoting CRC proliferation (Fig. 2D) [103]. Additionally, pep-AP can modulate CRC's chemotherapy sensitivity by adjusting metabolic pathways, leading to ROS accumulation and apoptosis, which may sensitize cells to treatments like Oxaliplatin (Fig. 2E) [108]. lncRNA-encoded peptides also regulate signaling pathways in CRC. BVES-AS1-201-50aa and MBOP peptides, for instance, activate the Src/mTOR and MEK1/pERK pathways, respectively, to bolster CRC cell viability, migration, and invasion (Fig. 2F–G) [106, 109]. The revelation that E3 ubiquitin ligases MAEA and RMND5A mediate MBOP degradation underscores the complex regulatory networks governing micropeptide metabolism within cells. The FORCP peptide adds another layer of complexity, inhibiting cell proliferation and inducing apoptosis in response to endoplasmic reticulum stress (Fig. 2H) [107]. These findings suggest that lncRNA-encoded peptides could serve not only as diagnostic markers but also as novel targets for therapeutic intervention in CRC, with the potential to improve treatment strategies through a deeper understanding of their mechanisms and regulatory roles.

Breast cancer
Breast cancer (BC), projected to have 310,720 new diagnoses and 42,250 deaths in the United States in 2024, is the most prevalent malignancy among women [135, 136]. Within this, triple-negative breast cancer (TNBC),

characterized by the absence of progesterone, estrogen, and human epidermal growth factor receptors, presents a particularly aggressive subtype with a lower survival rate and a complex molecular profile [137, 138]. LncRNA-encoded peptides are emerging as significant contributors to BC progression, with the peptide MRP, overexpressed in highly malignant BC cells, promoting invasion and metastasis by stabilizing EGFR mRNA and activating the PI3K pathway by binding to HNRNPC (Fig. 3A) [30]. The lncRNA product LINC00511-133aa enhances invasive properties and stem-like characteristics of BC cells by modulating the wnt/β-catenin pathway (Fig. 3B) [110], while HCP5-132aa is implicated in resistance to adriamycin and can trigger excessive autophagy through the ERK/mTOR pathway, and promote TNBC progression by regulating GPX4-induced ferroptosis (Fig. 3C) [112, 113]. Additionally, ASRPS, a peptide encoded by LINC00908, suppresses tumor angiogenesis by inhibiting the STAT3/VEGF pathway (Fig. 3D) [99], and CIP2A-BP, encoded by LINC00665, suppresses TNBC invasion and metastasis by inhibiting the PI3K/AKT/NF-κB pathway (Fig. 3E) [114]. Another peptide MAGI2-AS3-ORF5 interacts with the extracellular matrix to restrict BC cell viability and migration, though its mechanisms require further investigation (Fig. 3F) [111]. The discovery of these lncRNA-encoded peptides and their roles in BC, especially TNBC, opens new avenues for understanding disease progression and resistance to therapy. Their multifaceted influence on cellular processes suggests potential for targeted interventions. For instance, the modulation of MRP to destabilize EGFR mRNA could be a strategy to combat BC metastasis. Similarly, understanding the mechanisms

Table 4 Peptides encoded by lncRNA and their functions

Cancer	Peptide	LncRNA	Length	Role	Expression	Function	Interacting Protein	Mechanism	References
Colorectal cancer	HOXB-AS3	lncRNA HOXB-AS3	53aa	Anti-tumor	Decreased in tumor tissues	Inhibited tumor growth and metastasis in vitro and in vivo	hnRNP A1	Blocked the binding of hnRNP A1 to PKM mRNA, suppressing hnRNP A1-dependent PKM splicing and miR-18a processing	[33]
	ASAP	LINC00467	94aa	Pro-tumor	Overexpressed in tumor tissues	Promoted cell proliferation in vitro and tumor growth in vivo	ATP5A and ATP5C	Promoted interaction of ATP5A and ATP5C, increasing ATP synthase activity and mitochondrial oxygen consumption rate	[103]
	RBRP	LINC00266-1	71aa	Pro-tumor	Upregulated in highly metastatic cancer cell sublines and primary CRC tissues	Promoted tumor growth and metastasis in vitro and in vivo	IGF2BP1	Increased the binding of IGF2BP1 to the m^{6}A-modified c-Myc mRNA, increased the mRNA stability and expression of c-Myc,	[104]
	SRSP	lncRNA LOC90024	130aa	Pro-tumor	Upregulated in tumor tissues and highly metastatic cancer cell sublines	Promoted cell proliferation, migration, and invasion in vitro; stimulated CRC tumorigenesis and lung metastasis in vivo	SRSF3	Increased the binding of SRSF3 to exon 3 of SP4, resulting in the inclusion of SP4 exon 3 to induce the formation of the "cancerous" long SP4 isoform (L-SP4 protein) and inhibit the formation of the "noncancerous" short SP4 isoform (S-SP4 peptide)	[105]
	BVES-AS1-201-50aa	lncRNA BVES-AS1	50aa	Pro-tumor	–	Promoted cell viability, migration, and invasion in vitro	–	Activated the Src/mTOR signaling pathway	[106]
	FORCP	LINC00675	7aa	Anti-tumor	–	Inhibited cell proliferation and tumorigenesis, and induced apoptosis upon ER stress in vitro	–	–	[107]

Table 4 (continued)

Cancer	Peptide	LncRNA	Length	Role	Expression	Function	Interating Protein	Mechanism	References
	pep-AP	lnc-AP	37aa	Anti-tumor	–	Induced cell apoptosis, sensitized CRC cells to oxaliplatin in vitro and *in vivo*	TALDO1	Suppressed phosphorylated TALDO1 expression, attenuating the pentose phosphate pathway (PPP), and reducing NADPH/NADP+ and glutathione (GSH) levels	[108]
	MBOP	LINC01234	85aa	Pro-tumor	Upregulated in tumor tissues and cell lines	Promoted cell migration and proliferation in vitro, and tumor growth in vivo	MEK1	Promoted the expression of MEK1, and activated the MEK1/pERK/MMP2/MMP9 signaling pathway	[109]
Breast cancer	MRP	LncRNA LY6E-DT	153aa	Pro-tumor	Upregulated in highly malignant BC cells	Promoted tumor invasion and metastasis in vitro and in vivo	HNRNPC	Enhance the interaction between HNRNPC and EGFR mRNA, thus increasing EGFR mRNA stability and activating PI3K pathway	[30]
	LINC00511-133aa	LINC00511	133aa	Pro-tumor	–	Promoted the invasiveness and stemness while limiting apoptosis in vitro	–	Activated wnt/β-catenin pathway and facilitated β-catenin protein entry into the nucleus	[110]
	MAGI2-AS3-ORF5	LncRNA MAGI2-AS3	45aa	Anti-tumor	–	Restrained cell viability, proliferation, and migration in vitro	Extracellular matrix-related proteins	Might by binding to extracellular matrix-related proteins	[111]
	HCP5-132aa	lncRNA HCP5	132aa	–	–	Contributed to adriamycin resistance in vitro	–	Decreased HCP5-132aa elevated lipid ROS accumulation and Beclin 1 expression via ERK/mTOR pathway, inducing excessive autophagy	[112]

Table 4 (continued)

| Cancer	Peptide	LncRNA	Length	Role	Expression	Function	Interacting Protein	Mechanism	References
	ASRPS	LINC00908	60aa	Anti-tumor	Downregulated in tumor tissues and cell lines	Inhibited migration, invasion and angiogenesis in vitro and in vivo	STAT3	Inhibit STAT3 phosphorylation, reducing binding of STAT3 to the VEGF promoter and reduced VEGF transcription activity	[99]
	HCP5-132aa	lncRNA HCP5	132aa	Pro-tumor	Upregulated in tumor tissues	Promoted cell proliferation, migration in vitro; promoted tumor growth in vivo	—	HCP5-132aa knockdown down-regulated GPX4 to induce ferroptosis, independent of the Xc-/GSH pathway	[113]
	CIP2A-BP	LINC00665	52aa	Anti-tumor	Downregulated in tumor cell lines	Inhibited migration and invasion in vitro and in vivo	CIP2A	Replaced PP2A's B56γ subunit and released PP2A activity, inhibiting PI3K/AKT/NFκB pathway	[114]
Liver cancer	PINT87aa	LINC-PINT	87aa	Anti-tumor	Overexpressed in H2O2-induced senescent HCC cells	Induced growth inhibition, cellular senescence, and decreased mitophagy in vitro and in vivo	FOXM1	Reduced FOXM1-mediated transcription of PHB2	[115]
	SMIM30	LINC00998	59aa	Pro-tumor	Upregulation in tumor tissues	Promoted tumor growth and metastasis in vitro and in vivo	SRC and YES1	c-Myc transcribed SMIM30, which initiated the membrane anchoring of SRC/YES1 and activated the downstream MAPK signaling pathway. Promoted the G1/S transition by regulating the cyclin/CDK-Rb-E2F1 pathway, and regulating cytosolic calcium level	[89, 116]
	HBVPTPAP	lncRNA HBVPTPAP	145aa	Anti-tumor	—	Induced the apoptosis of HCC cells in vitro	PILRA	Activated JAK/STAT signaling pathway by interatccting with PILRA intracellular domain	[32]

Table 4 (continued)

Cancer	Peptide	LncRNA	Length	Role	Expression	Function	Interating Protein	Mechanism	References
	CIP2A-BP	LINC00665	52aa	Pro-tumor	–	Increased cell proliferation, invasion, and migration in vitro	–	–	[117]
	C20orf204-189AA	Linc00176	189aa	Pro-tumor	upregulated in tumor tissues	Enhanced cell proliferation in vitro	rRNA and Nucleolin	Stabilized nucleolin and promoted rRNA transcription	[31]
	Linc013026-68AA	Linc013026	68aa	Pro-tumor	–	Enhanced Cell Proliferation in vitro	–	–	[118]
	TP53LC02	TP53-regulated lncRNAs	109aa	Anti-tumor	–	Inhibited cell proliferation and controlled cell cycle in response to DNA damage in vitro	–	–	[119]
	TP53LC04	TP53-regulated lncRNAs	100aa	Anti-tumor	–	Inhibited cell proliferation in vitro	–	–	[119]
Lung cancer	ATMLP	lncRNA AFAP1-AS1	90aa	Pro-tumor	Upregulated in tumor tissues	Suppressed autolysosome formation in vitro; promoted the malignant transformation of epithelial cells in vivo	NIPSNAP1	Inhibited NIP-SNAP1 transport from the inner to the outer mitochondrial membrane	[120]
	lnc DLX6-AS1 encoded peptide	LINC DLX6-AS1	–	Pro-tumor	–	Promoted cell proliferation, migration, and invasion in vitro	–	Activated the wnt/β-catenin pathway	[121]
	UBAP1-AST6	–	–	Pro-tumor	–	Promoted cell proliferation in vitro	–	–	[122]
Esophageal squamous cell carcinoma	Pep-KDM4A-AS1	KDM4A-AS1	61aa	Anti-tumor	–	Reduced cell viability and migration in vitro	–	Might regulate the oxidation–reduction process and fatty acid metabolism	[123]
	Pep-LINC01116	LINC01116	87aa	Anti-tumor	–	Reduced cell viability and migration in vitro	–	–	[123]
	YY1BM	LINC00278	21aa	Anti-tumor	Downregulated in tumor tissues of male with smoking	Promoted apoptosis and inhibited tumor growth, adapting better to nutrient deprivation in vivo	YY1	Inhibited the interaction between YY1 and AR, which in turn decreased expression of eEF2K through the AR signaling pathway	[91]

Table 4 (continued)

Cancer	Peptide	LncRNA	Length	Role	Expression	Function	Interating Protein	Mechanism	References
Pancreatic cancer	RASON	LINC00673	78aa	Pro-tumor	Overexpressed in tumor tissues	Promoted tumor growth in vitro and in vivo	KRAS G12D/V	Suppressed both intrinsic and GTPase activating protein (GAP)-mediated GTP hydrolysis, thereby sustaining KRAS$^{G12D/V}$ in the GTP-bound hyperactive state	[124]
Renal cell carcinoma	SMIM26	LINC00493	95aa	Anti-tumor	Downregulated in tumor tissues	Inhibited tumor growth and metastasis in vitro and in vivo	AGK and SLC25A11	Increased the mitochondrial localization of AGK and inhibiting AGK-mediated AKT phosphorylation	[125]
	MIAC (micropeptide inhibiting actin cytoskeleton)	lncRNA AC025154.2	51aa	Anti-tumor	Downregulated in tumor tissues	Inhibited cell proliferation and migration ability, promoted S phase and G2 phase arrest, and cell apoptosis in vitro. MIAC overexpression inhibited the growth and metastasis of RCC in vivo	AQP2	Inhibited EREG/EGFR expression and downstream PI3K/AKT and MAPK pathways	[126]
Ovarian cancer	DDUP	lncRNA CTBP1-DT	186aa	Pro-tumor	–	Promoted DNA damage repair, and confered resistance to cisplatin in vitro	ATR, γ-H2AX, RAD18 and RAD51C	Phosphorylation of DDUP by DNA damage enhances RAD18 interaction at damage sites, facilitating DNA repair through RAD51C-dependent HRR and monoubiquitinated PCNA-driven PRR pathways	[39]
	DDUP	lncRNA CTBP1-DT	186aa	Pro-tumor	Upregulation in PDOVC with treatment with CDDP	Enhanced the capability of ovarian cancer cells to repair damaged DNA, resulting in resistance to CDDP in vitro and in vivo	–	DNA damage-induced DDUP foci sustained the RAD18/RAD51C and RAD18/PCNA complexes at the sites of DNA damage	[127]

Table 4 (continued)

Cancer	Peptide	LncRNA	Length	Role	Expression	Function	Interacting Protein	Mechanism	References
Neuroblastoma	NBASP	lncRNA FAM201A	155aa	Anti-tumor	Downregulated in tumor tissues	Inhibited cell proliferation, migration and invasion in vitro	FABP5	Reduced expression of FABP5 via the ubiquitin proteasome pathway, inactivating the MAPK pathway	[128]
	sPEP1	HNF4A-AS1	51aa	Pro-tumor	Upregulated in tumor cell lines and tissues	Promoted stemness and metastasis of NB stem cells in vitro and in vivo	eEF1A1	Facilitated eEF1A1 binding to SMAD4, resulting in repression of SMAD4 transactivation and transcriptional upregulation of stem cell genes related to tumor progression	[129]
Osteosarcoma	LINC00665_18a	LINC00665	18aa	Anti-tumor	–	Inhibited cell proliferation and migration in vitro, and diminished tumor growth in vivo	CREB1	Impaired the transcriptional activity, nuclear localization, and phosphorylation of CREB1, weakening the interaction between CREB1 and RPS6KA3	[130]
Oral squamous cell carcinoma	HOXB-AS3	LncRNAHOXB-AS3	53aa	Pro-tumor	Upregulated in tumor tissues	Promoted cell proliferation in vitro	IGF2BP2	Promoted c-Myc mRNA stability	[131]
Pan-cancer	pTINCR	TINCR lncRNA	87aa	Anti-tumor	Upregulated in epithelial-like cancer cell lines	Promoted epithelial differentiation in vitro and in vivo, reduced tumor growth in patient-derived xenografts (PDX) model	SUMO	Increased CDC42 SUMOylation and promoting its activation	[132]
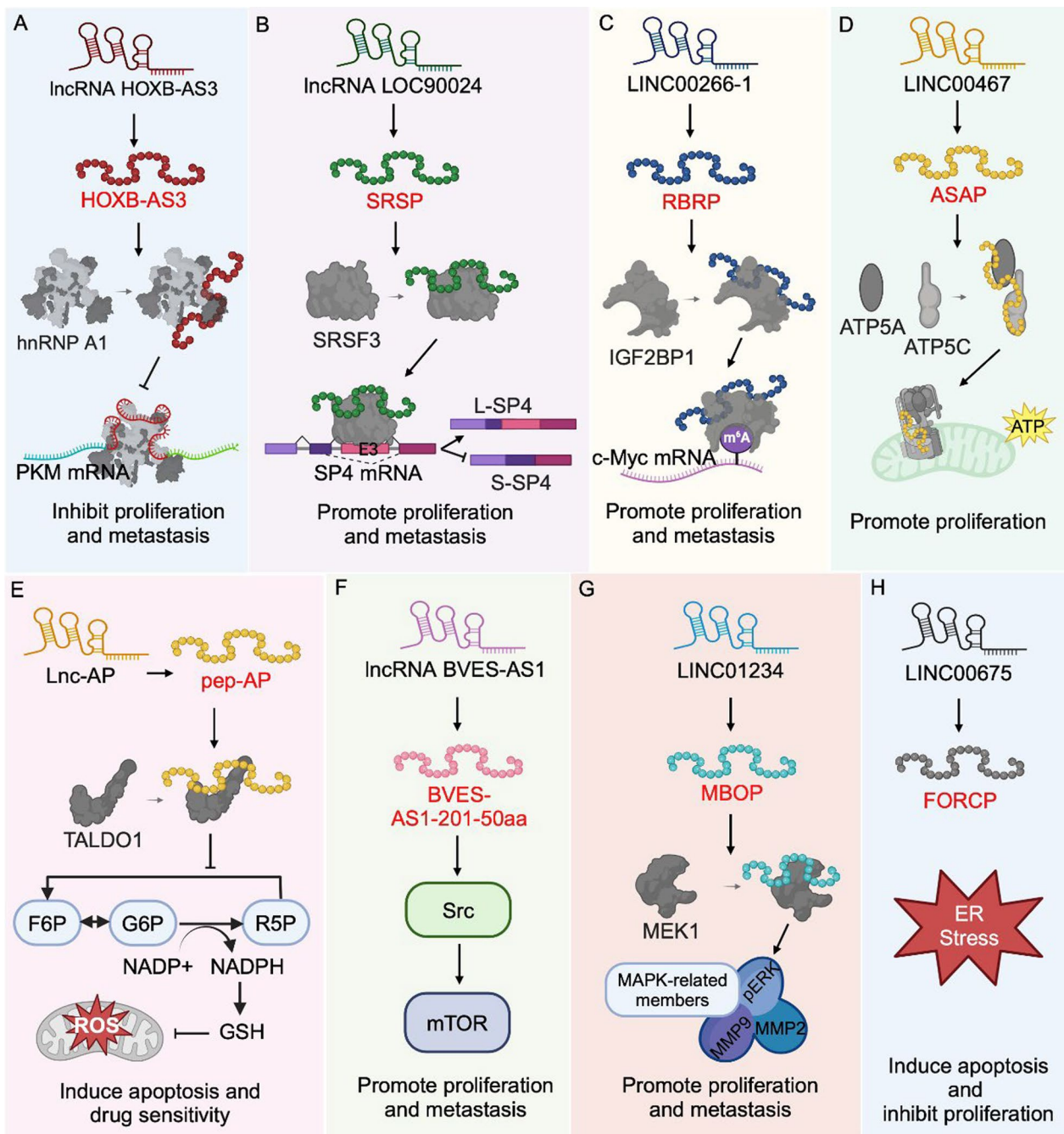
Fig. 2 The role of lncRNA-encoded peptides in colorectal cancer (CRC). **A** The peptide HOXB-AS3 encoded by LncRNA HOXB-AS3 interacts with hnRNP A1 to affect PKM mRNA splicing, inhibiting CRC growth and metastasis. **B** The peptide SRSP encoded by LncRNA LOC90024 interacts with SRSF3 to influence splicing of SP4 mRNA, promoting CRC growth and metastasis. **C** The peptide RBRP encoded by LINC00266-1 interacts with IGF2BP1 to maintain c-Myc mRNA stability, promoting CRC growth and metastasis. **D** The peptide ASAP encoded by LINC00467 enhances ATP synthase activity and mitochondrial oxygen consumption by interacting with ATP5A and ATP5C, promoting CRC growth. **E** The peptide pep-AP encoded by Lnc-AP interacts with TALDO1 to attenuate the pentose phosphate pathway (PPP), inducing apoptosis and drug sensitivity in colorectal cancer cells. **F** The peptide BVES-AS1-201-50aa encoded by LncRNA BVES-AS1 activates the Src/mTOR signaling pathway, promoting CRC proliferation, migration, and invasion. **G** The peptide MBOP encoded by LINC01234 interacts with MEK1 to regulate the MEK1/pERK/MMP2/MMP9 axis, promoting CRC proliferation and metastasis. **H** The peptide FORCP encoded by LINC00675 induces apoptosis and inhibits cell proliferation in colorectal cancer cells under endoplasmic reticulum stress. Image created with BioRender.com

by which peptides like ASRPS and CIP2A-BP inhibit key signaling pathways could lead to the development of new therapeutics that enhance the efficacy of existing treatments or overcome resistance. The interplay between lncRNA products and the extracellular matrix also presents an opportunity to explore the tumor microenvironment's role in BC progression.

Liver hepatocellular carcinoma

Liver hepatocellular carcinoma (LIHC) is the most prevalent form of primary liver cancer, constituting 90% of all hepatic cancers [139, 140]. The molecular landscape of LIHC is complex and involves lncRNAs encoded peptides, which play crucial roles in the pathogenesis and progression of the disease. For instance, HBVPTPAP induces apoptosis in LIHC cells via activation of the JAK/STAT signaling pathway, potentially through interaction with PILRA (Fig. 4A) [32]. The peptide SMIM30 is upregulated in LIHC tissues and promotes cell proliferation, migration, and invasion by interacting with SRC and YES1, activating the MAPK pathway, and being transcriptionally regulated by c-Myc (Fig. 4B) [89]. SMIM30 also enhance cell proliferation by promoting the G1/S transition via the Rb pathway and modulate the cyclin/CDK-Rb-E2F1 pathway and cytosolic calcium levels [116], which extends the impact of SMIM30 in LIHC. PINT87aa overexpressed in senescent LIHC cells, inhibits growth and induces cellular senescence by blocking FOXM1-mediated transcription of PHB2 (Fig. 4C) [115], while C20orf204-189AA enhances cell proliferation by stabilizing nucleolin and promoting ribosomal RNA transcription (Fig. 4D) [31]. The presence of additional functional lncRNA-encoded peptides such as CIP2A-BP and Linc013026-68AA in LIHC further underscores the diversity of their roles, with CIP2A-BP enhancing HCC cell proliferation and metastasis in LIHC (Fig. 4E) [117], contrasting its suppressive role in TNBC by inhibiting the PI3K/AKT/NF-κB pathway [114], as previously mentioned. The divergent roles of CIP2A-BP in LIHC and TNBC may be attributed to several factors. These include variations in the cellular microenvironment, differences in the signaling pathways active within each cancer type,

and the potential for CIP2A-BP to interact with distinct binding partners across various tissues. These considerations highlight the importance of accounting for tissue-specific and context-specific actions when assessing the contributions of lncRNA-encoded peptides to cancer pathogenesis. Additionally, Linc013026-68AA, has been shown to augment LIHC proliferation (Fig. 4F) [118], of which the precise mechanism also warrants further investigation.

Lung cancer

Lung cancer (LC) remains the primary cause of cancer-related mortality worldwide, with a grim prognosis and an estimated 234,580 new cases in the United States alone for 2024 [133, 136, 141, 142]. The disease is generally categorized into two main types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), each with distinct clinical features and treatment approaches [143–146]. Recent advances in molecular research have shed light on the role of lncRNA-encoded peptides in NSCLC, such as ATMLP, which is upregulated in NSCLC tissues and disrupts mitophagy by interacting with NIPSNAP1, thereby promoting malignant transformation and tumorigenesis (Fig. 5A) [120]. Interestingly, ATMLP's expression is regulated by N6-methyladenosine (m^{6}A) methylation of its encoding lncRNA AFAP1-AS1 [120], introducing a new perspective on the post-transcriptional regulation of lncRNA-encoded peptides, and suggesting that epigenetic modifications, such as m^{6}A methylation, may serve as key regulators in the expression and function of these peptides. Additionally, a peptide encoded by lncRNA DLX6-AS1 has been shown to activate the Wnt/β-catenin pathway, enhancing NSCLC cell proliferation and metastasis (Fig. 5B) [121]. The lncRNA product UBAP1-AST6 also enhances LC cell proliferation and clone formation, although its mechanisms of action require further investigation (Fig. 5C) [122].

Esophageal cancer

Esophageal cancer, predominantly manifesting as esophageal squamous cell carcinoma (ESCC), is the sixth most

(See figure on next page.)

Fig. 3 The role of lncRNA-encoded peptides in breast cancer. **A** The peptide MRP encoded by LncRNA LY6E-DT regulates EGFR mRNA stability and translation by interacting with HNRNPC, promoting breast cancer metastasis. **B** The peptide LINC00511-133aa encoded by LINC00511 facilitates β-catenin nuclear translocation to activate the transcription of Bax, c-Myc, and CyclinD1, promoting invasiveness and stem-like properties of breast cancer. **C** The peptide HCP5-132aa encoded by LncRNA HCP5 inhibits autophagy and ferroptosis to promote breast cancer proliferation and migration. **D** The peptide ASRPS encoded by LINC00908 inhibits STAT3 phosphorylation, leading to the suppression of VEGF transcription and thus inhibiting tumor metastasis and angiogenesis. **E** The peptide CIP2A-BP encoded by LINC00665 competes with PP2A for binding to CIP2A, reducing AKT phosphorylation to inhibit the PI3K/AKT/NFκB pathway, leading to the downregulation of MMP2, MMP9, and Snail, thus inhibiting breast cancer invasion and metastasis. **F** The peptide MAGI2-AS3-ORF5 encoded by LncRNA MAGI2-AS3 interacts with extracellular matrix proteins to inhibit breast cancer cell proliferation and migration. Image created with BioRender.com
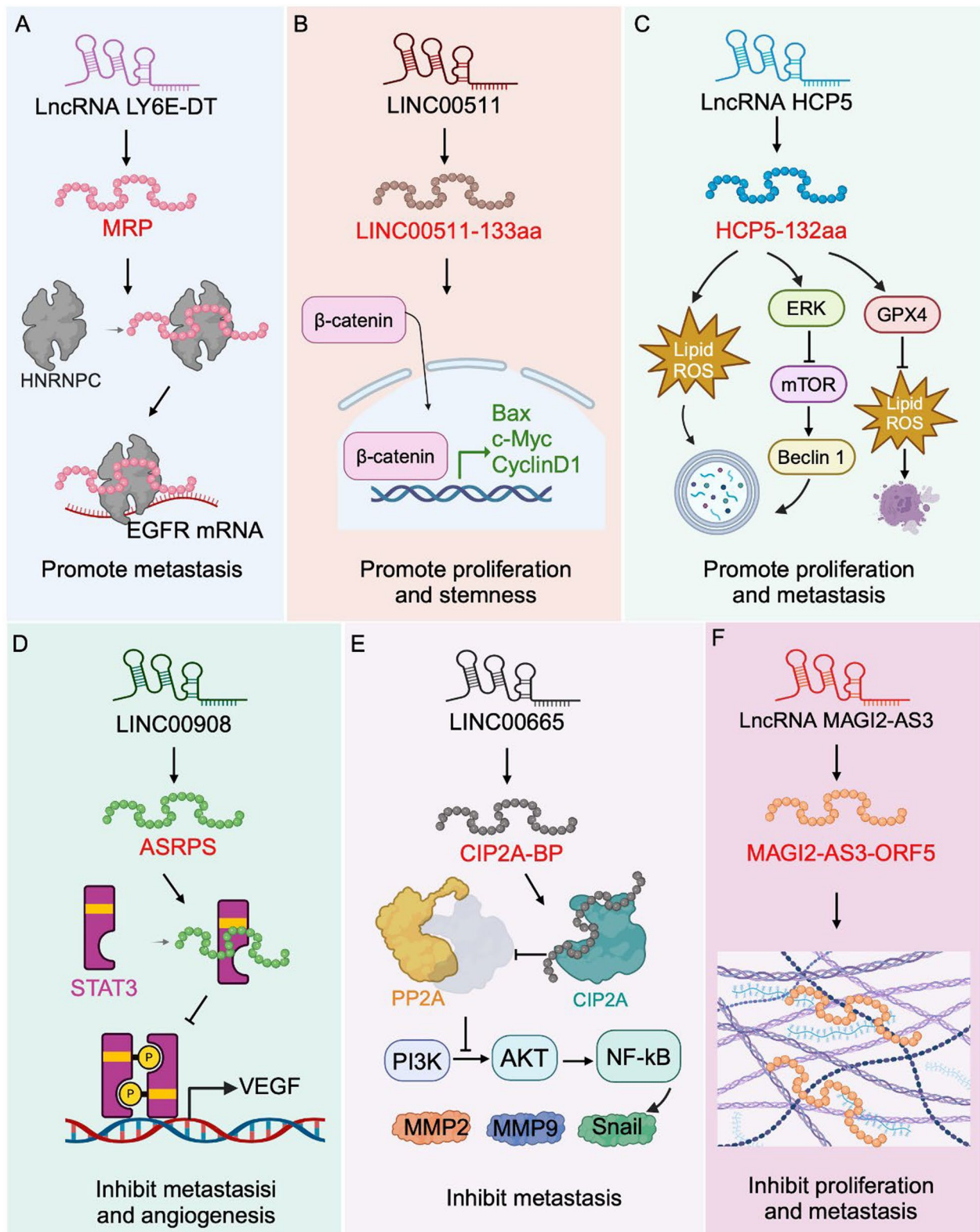
Fig. 3 (See legend on previous page.)
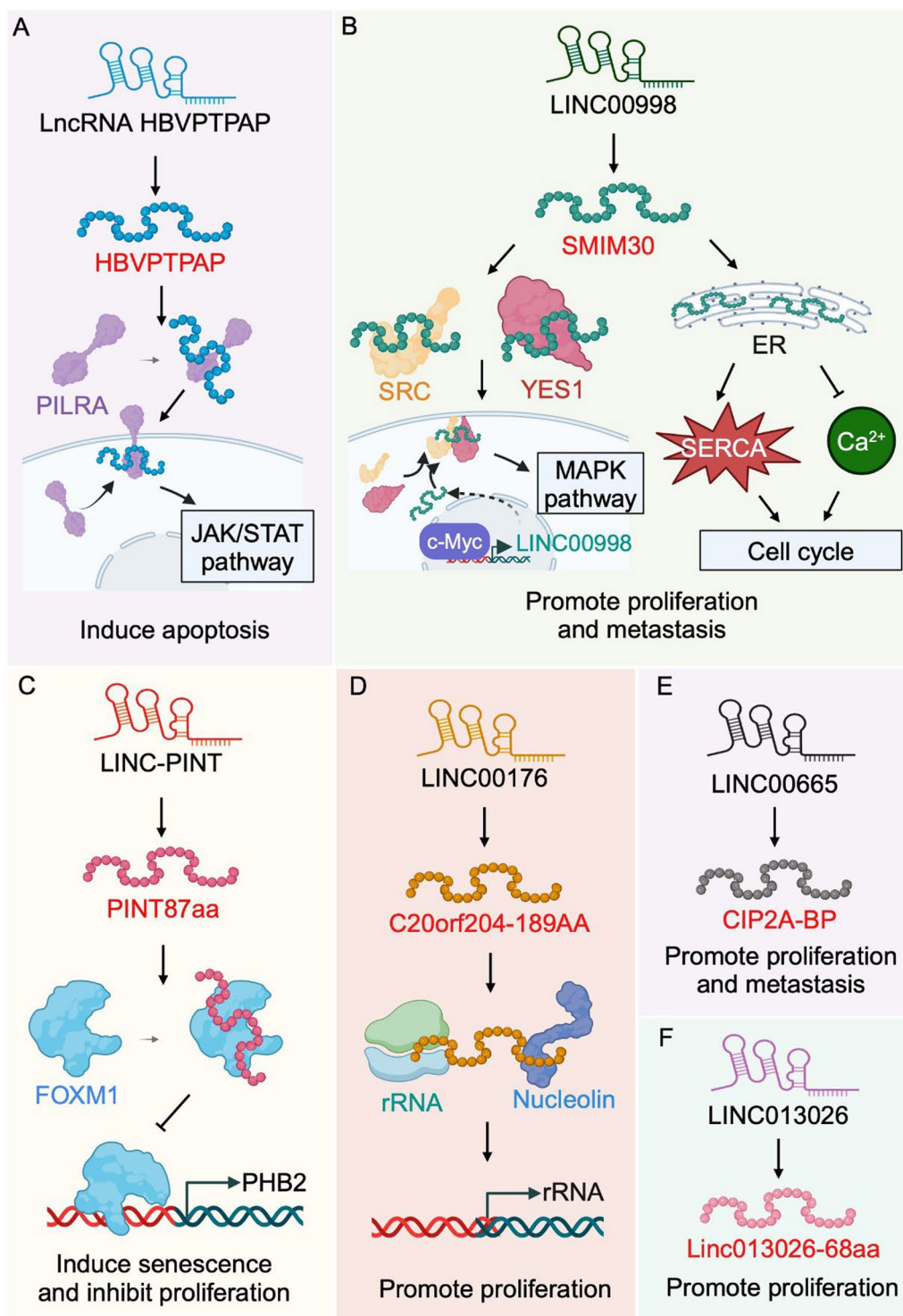
Fig. 4 The role of lncRNA-encoded peptides in liver cancer. **A** The peptide HBVPTPAP encoded by LncRNA HBVPTPAP promotes membrane localization of PILRA by interacting with it, activating the JAK/STAT signaling pathway to induce apoptosis and inhibit liver cancer development. **B** The peptide SMIM30 encoded by LINC00998 activates the MAPK signaling pathway and regulates the G1/S phase transition to promote liver cancer proliferation and metastasis. **C** The peptide PINT87aa encoded by LINC-PINT interacts with FOXM1 to inhibit PHB2 transcription, inducing cellular senescence and suppressing liver cancer growth. **D** The peptide C20orf204-189AA encoded by LINC00176 promotes liver cancer cell proliferation by stabilizing Nucleolin and enhancing rRNA transcription. **E** The peptide CIP2A-BP encoded by LINC00665 promotes liver cancer growth and metastasis. **F** The peptide Linc013026-68aa encoded by LINC013026 enhances the in vitro proliferation of HCC cells. Image created with BioRender.com
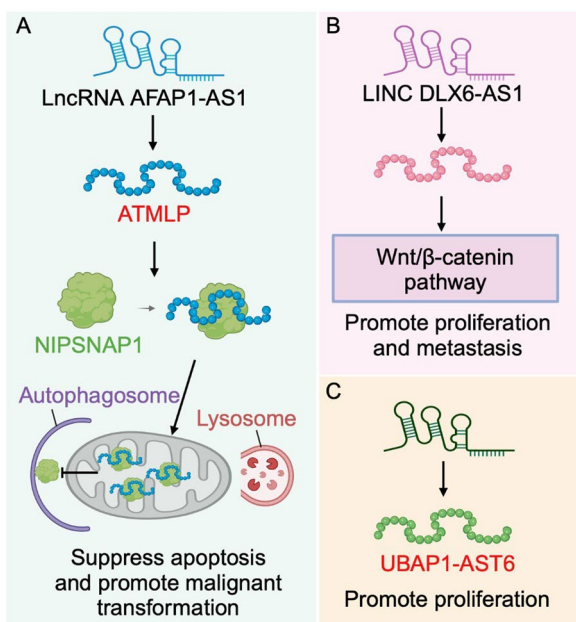
Fig. 5 The role of lncRNA-encoded peptides in lung cancer. **A** The peptide ATMLP encoded by lncRNA AFAP1-AS1 disrupts autolysosome formation by interacting with NIPSNAP1, hindering its transport, leading to lung cancer development and progression. **B** The peptide encoded by lncRNA DLX6-AS1 enhances the proliferation, migration, and invasion of NSCLC cells by activating the Wnt/β-catenin signaling pathway. **C** The peptide UBAP1-AST6 encoded by an LncRNA promotes the proliferation of lung cancer cells in vitro. Image created with BioRender.com

common cause of cancer mortality worldwide, with a significant incidence in China where it represents over 50% of global cases [147–149]. Late-stage symptoms like dysphagia and cervical lymph node enlargement contribute to a low 5-year survival rate and a poor prognosis for ESCC patients [150]. Recent studies have shed light on the role of lncRNA-encoded peptides in ESCC, offering a promising avenue in the battle against this aggressive cancer. Pep-KDM4A-AS1, a peptide encoded by the lincKDM4A-AS1, has been shown to diminish ESCC cell viability and migration by modulating the oxidation–reduction process and fatty acid metabolism. Another peptide (Fig. 6A) [123]. Pep-LINC01116, exhibits similar effects on cell viability and migration (Fig. 6B) [123]. Additionally, YY1BM, a peptide encoded by LINC00278, influences ESCC progression by disrupting the AR signaling pathway, leading to altered expression of eEF2K and impacting cell adaptability under nutrient-deprived conditions (Fig. 6C) [91].

Pancreatic cancer

The global burden of pancreatic cancer has seen a sharp escalation over recent decades, with a grim projection

that it will remain the leading cause of cancer-related mortality [151, 152]. Recent molecular research has identified the lncRNA-encoded peptide RASON, encoded by LINC00673, as a critical factor in pancreatic cancer pathology. Overexpressed in pancreatic cancer tissues, RASON promotes the proliferation of pancreatic ductal adenocarcinoma by interacting with the oncogenic KRAS$^{G12D/V}$ mutant protein (Fig. 7A) [124]. This interaction inhibits KRAS$^{G12D/V}$'S GTPase activity and GTP hydrolysis by GTPase activating protein (GAP), leading to the stabilization of KRAS$^{G12D/V}$ in a GTP-bound, hyperactive state—a key driver of pancreatic cancer [124]. The modulation of KRAS activity by RASON, considering KRAS's frequent mutation in cancer, underscores the peptide's potential as a therapeutic target.

Renal cell carcinoma

Renal cell carcinoma (RCC), with its most aggressive subtype being clear cell renal cell carcinoma (ccRCC), represents a significant health burden, contributing to an estimated 400,000 new cases and 175,000 deaths globally in 2018 [153, 154]. Recent research has shed light on the role of lncRNA-encoded peptides in the pathology of RCC. The peptide SMIM26 is downregulated in RCC tissues and has been shown to inhibit tumor proliferation and metastasis by interacting with AGK and SLC25A11, thereby affecting mitochondrial glutathione import and respiratory efficiency (Fig. 7B) [125]. Additionally, MIAC is down-expressed in ccRCC and, when overexpressed, inhibits tumor proliferation and migration while promoting apoptosis through the modulation of the PI3K/AKT and MAPK pathways by binding to the AQP2 protein and inhibiting EREG/EGFR expression (Fig. 7C) [126].

Ovarian cancer

Ovarian cancer (OV), encompassing malignancies of the ovary, fallopian tube, and peritoneum, is a significant health concern with an annual global incidence of 313,959 cases and 207,252 deaths [155]. Despite declining incidence rates and improving survival rates in regions like the United States and Europe, partly due to the use of oral contraceptives [136, 156], the prognosis for OV remains poor, with most patients diagnosed at advanced stages and lacking effective early detection strategies [157]. However, the role of the lncRNA-encoded peptide DDUP, derived from CTBP1-DT, has emerged as a key player in OV's molecular pathology, particularly in DNA damage repair. DDUP's upregulation is associated with enhanced DNA repair mechanisms and cisplatin resistance in ovarian cancer cells. The use of the ATR inhibitor Berzosertib has been shown to disrupt DDUP foci formation, thereby sensitizing these cells to DNA-damaging chemotherapeutics. The phosphorylation of DDUP
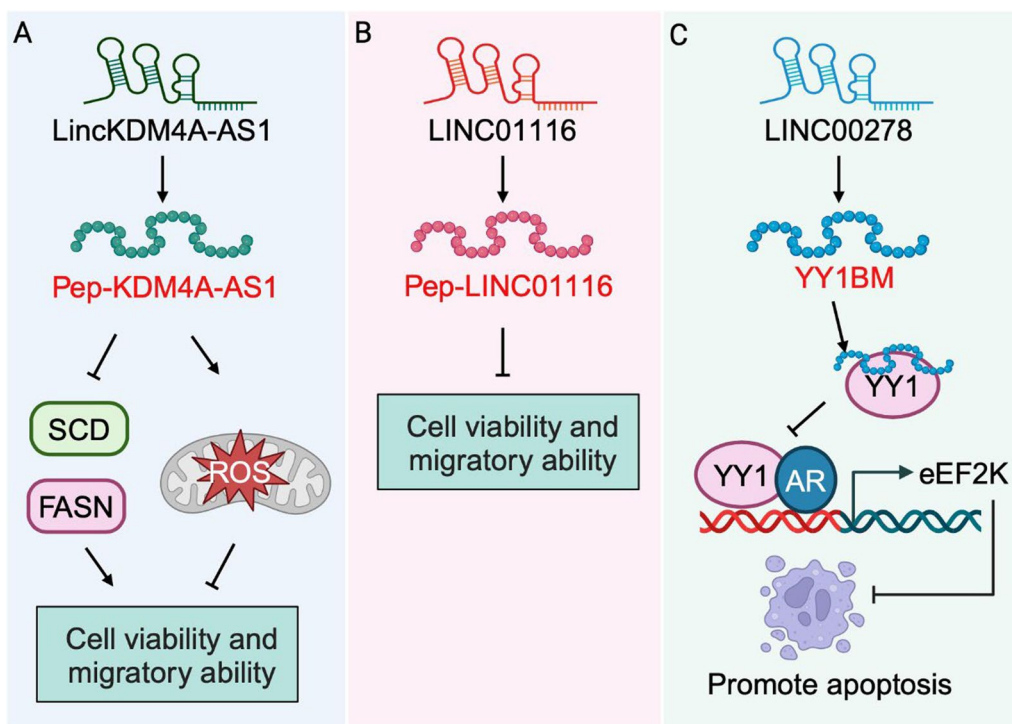
Fig. 6 The role of lncRNA-encoded peptides in esophageal cancer. **A** The peptide Pep-KDM4A-AS1 encoded by LincKDM4A-AS1 inhibits the proliferation and migration of esophageal cancer cells by regulating intracellular redox processes and fatty acid metabolism. **B** The peptide Pep-LINC01116 encoded by LINC01116 reduces the viability of ESCC cells and inhibits their migration. **C** The peptide YY1BM encoded by LINC00278 promotes apoptosis in esophageal cancer cells by disrupting the binding of YY1 and AR, leading to reduced eEF2K expression. Image created with BioRender.com

in response to DNA damage induces a conformational change that strengthens its interaction with RAD18, supporting DNA repair through homologous recombination (HR) and post-replication repair mechanisms (Fig. 8A) [39]. The upregulation of DDUP following cisplatin treatment further confirms its role in promoting cellular resistance to chemotherapy, emphasizing its significance in OV's therapeutic resistance. Another research team has revealed that DDUP is upregulated in patient-derived OV cells following cisplatin treatment, enhancing the cells' capacity for DNA repair and resulting in cisplatin resistance through RAD51C-mediated HR and PCNA-mediated post-replication repair [127], which further confirm the significant role of lncRNA-encoded peptide in DNA damage repair.

Neuroblastoma

Neuroblastoma (NB) is the most common extracranial solid tumor in children, originating from the developing peripheral sympathetic nervous system and representing approximately 8% of all childhood cancers [158–160]. Despite progress in targeted therapies, the long-term survival rate for high-risk children remains under 40%, highlighting the need for innovative treatment strategies

[160]. LncRNA-encoded peptides have emerged as potential players in NB pathology. NBASP is down-regulated in NB tissues and inhibits cell proliferation, and metastasis by interacting with FABP5 and reducing its expression through the ubiquitin proteasome pathway, resulting in the inactivation of the MAPK signaling pathway (Fig. 8B) [128]. On the other hand, sPEP1, a peptide encoded by HNF4A-AS1 and upregulated in NB stem cells, promotes tumor progression interacting with eEF1A1, enhancing its binding to SMAD4, and leading to the transcriptional upregulation of stem cell genes associated with tumor progression (Fig. 8C) [129].

Osteosarcoma

Osteosarcoma (OS), while a rare cancer, is the most common bone malignancy affecting children and adolescents [161]. It is believed to originate from osteoblastic mesenchymal cells [162]. The prognosis for patients with OS varies significantly depending on the stage of the disease; the 5 year survival rate for patients with localized OS is approximately 70%, but this figure drops to less than 30% for those with metastatic disease, indicating a poor survival outcome [163]. Recent research has highlighted the potential role of lncRNA-encoded peptides in the
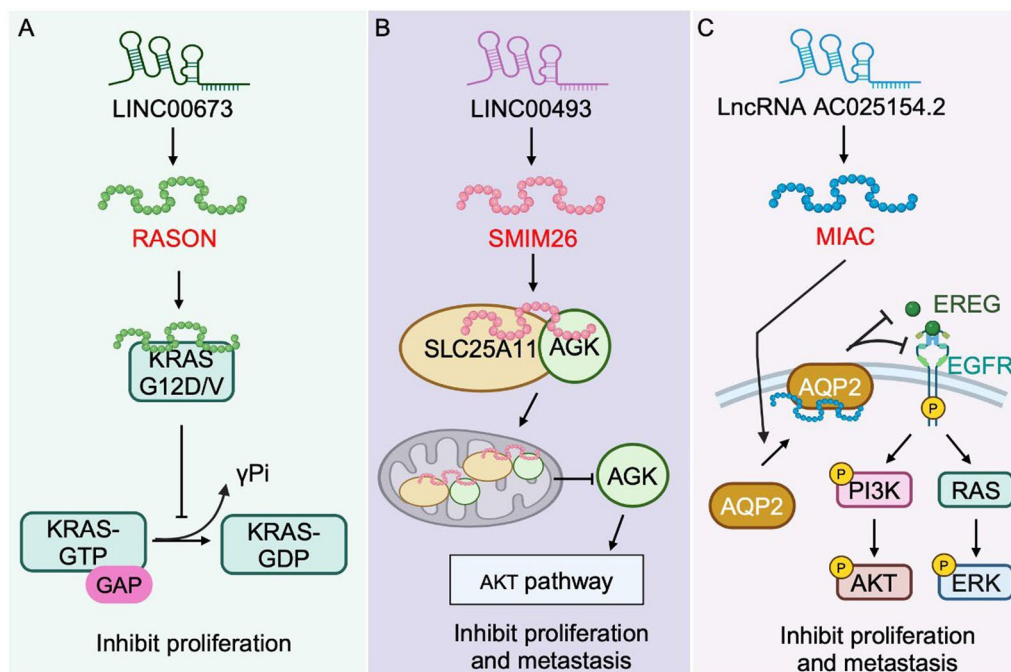
Fig. 7 The role of lncRNA-encoded peptides in pancreatic and renal cancers. **A** The peptide RASON encoded by LINC00673 promotes the growth of pancreatic cancer by stabilizing KRASG12D/V in an active GTP-bound state through interaction with KRASG12D/V. **B** The peptide SMIM26 encoded by LINC00493 inhibits the proliferation and migration of renal cell carcinoma by enhancing mitochondrial localization of AGK, thereby inhibiting AGK-mediated AKT phosphorylation. **C** The peptide MIAC encoded by LncRNA AC025154.2 inhibits the proliferation and migration of renal cell carcinoma by interacting with AQP2 to suppress the expression of EREG/EGFR. Image created with BioRender.com

pathology of OS. One such peptide, LINC00665_18aa, suppresses the viability, proliferation, and migration of human OS cells in vitro and diminishes tumor growth in vivo. The mechanistic insight behind these effects reveals that LINC00665_18aa impairs the transcriptional activity, nuclear localization, and phosphorylation of the CREB1 and disrupts the interaction between CREB1 and RPS6KA3 [130].

Oral squamous cell carcinoma
Oral cancer, predominantly oral squamous cell carcinoma (OSCC), ranks as the sixth most common malignancy globally, yet the 5-year overall survival rate remains under 50%, underscoring an urgent need for innovative therapeutic targets [133, 164]. Research into lncRNA-encoded peptides in OSCC has identified HOXB-AS3 as a significant factor; it is upregulated in OSCC tissues and facilitates cell proliferation and viability by interacting with IGF2BP2 to stabilize the mRNA of c-MYC, a key driver in cell cycle progression and cancer development [131, 165, 166]. This indicates a potential oncogenic role for HOXB-AS3 in OSCC. Interestingly, contrasting roles for HOXB-AS3 have been observed in CRC, where it is downregulated and inhibits cancer progression by interfering with PKM splicing, a key regulatory step in glucose

metabolism and the Warburg effect characteristic of cancer cells [167–169]. The dualistic behavior of HOXB-AS3 in different cancers, similar to that of CIP2A-BP in liver and breast cancers, highlights the complexity of lncRNA-encoded peptides and their tissue-specific roles in cancer.

Acute myeloid leukemia
Acute myeloid leukemia (AML) is one of the most commonly clinically fatal malignancies, characterized by differentiation block and clonal expansion of immature cells at various stages. The genetic complexity and highly heterogeneous nature of AML contribute to diverse subtypes with poor prognosis, leading to the limited effects of specific therapies [170–172]. The regulatory influence of lncRNA-encoded peptides on protein translation has been discerned in AML. The micropeptide APPLE is notably enriched in ribosomes, where it modulates the initiation phase of translation. This modulation enhances the synthesis of oncoproteins, thereby sustaining elevated rates of translation essential for the malignant phenotype. Mechanistically, APPLE fosters the interaction between PABPC1 and eIF4G, thereby facilitating mRNA circularization and the assembly of the eIF4F initiation complex. This assembly underpins a specific translational program that is conducive to cancer progression [173].While the
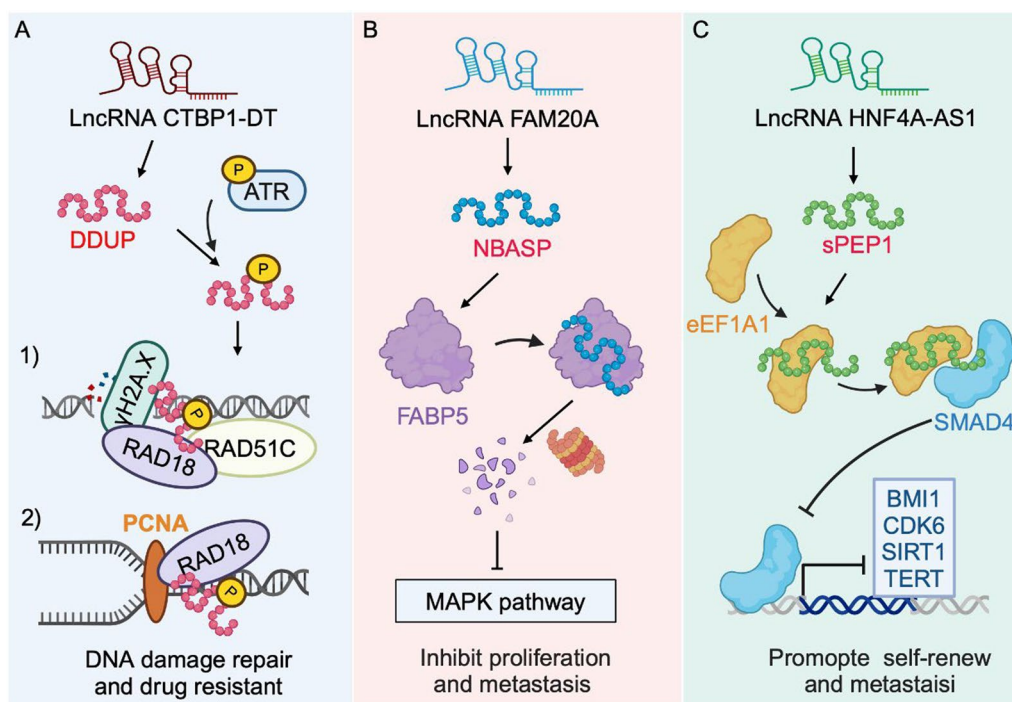
Fig. 8 The role of lncRNA-encoded peptides in ovarian and glioblastoma cancers. **A** The peptide DDUP encoded by LncRNA CTBP1-DT enhances DNA damage repair and cisplatin resistance in ovarian cancer cells by interacting with H2A.X and RAD18. **B** The peptide NBASP encoded by LncRNA increases the degradation of FABP5, leading to the inactivation of the MAPK pathway and inhibiting the proliferation and migration of glioblastoma cells. **C** The peptide sPEP1 encoded by LncRNA HNF4A-AS1 promotes the transcriptional upregulation of hepatocyte-related genes by enhancing the interaction with SMAD4, leading to the occurrence and metastasis of glioblastoma. Image created with BioRender.com

current body of research is indeed limited, the role of lncRNA-encoded peptides in other hematologic malignancies, such as chronic myeloid leukemia, remains an uncharted territory ripe for exploration.

Although research on lncRNA-encoded peptides has unveiled their potential roles in several types of cancer (Fig. 9), the precise mechanisms by which certain peptides exert their functions remain to be fully elucidated. The complexity of the role of m^{6}A modification in lncRNA-encoded peptides is also increasingly evident. For instance, as previously mentioned, m^{6}A methylation in the lncRNA AFAP1-AS1 controls the translation of the micropeptide ATMLP in lung cancer [120], while the peptides RBRP can bind to IGF2BP1 and HOXB-AS3 binds to IGF2BP2, important readers of m^{6}A modification [174, 175], to increase m^{6}A recognition in c-Myc mRNA in CRC and OSCC respectively [104, 131]. These studies suggest that lncRNA-encoded peptides can not only regulate by m^{6}A modification but also cooperate with m^{6}A modification to influence downstream molecules. However, whether other peptides are regulated by RNA modifications and the intricate interplay between them requires further investigation. Moreover, it is noteworthy that some peptides, including CIP2A-BP and

HOXB-AS3, may play opposing roles in different tumors, highlighting the importance of describing a peptide's action within the specific context of a particular cancer. The functional duality of these peptides underscores the need for a nuanced understanding of their roles in various cancerous environments. Furthermore, the exploration of these peptides in other cancer types is currently lacking, such as in the more common malignancies like gastric and prostate cancer. Expanding our research to include these prevalent cancers is crucial for gaining a comprehensive understanding of the breadth of lncRNA-encoded peptides' impact on cancer biology and their potential as therapeutic targets. The investigation into the roles of these peptides in a wider range of cancers could reveal novel insights into cancer pathogenesis and identify new opportunities for targeted cancer therapies.

Functional mechanisms of lncRNA-encoded peptides in cancer

LncRNA-encoded peptides, despite their short lengths, exert significant regulatory effects in cancer through various mechanisms.
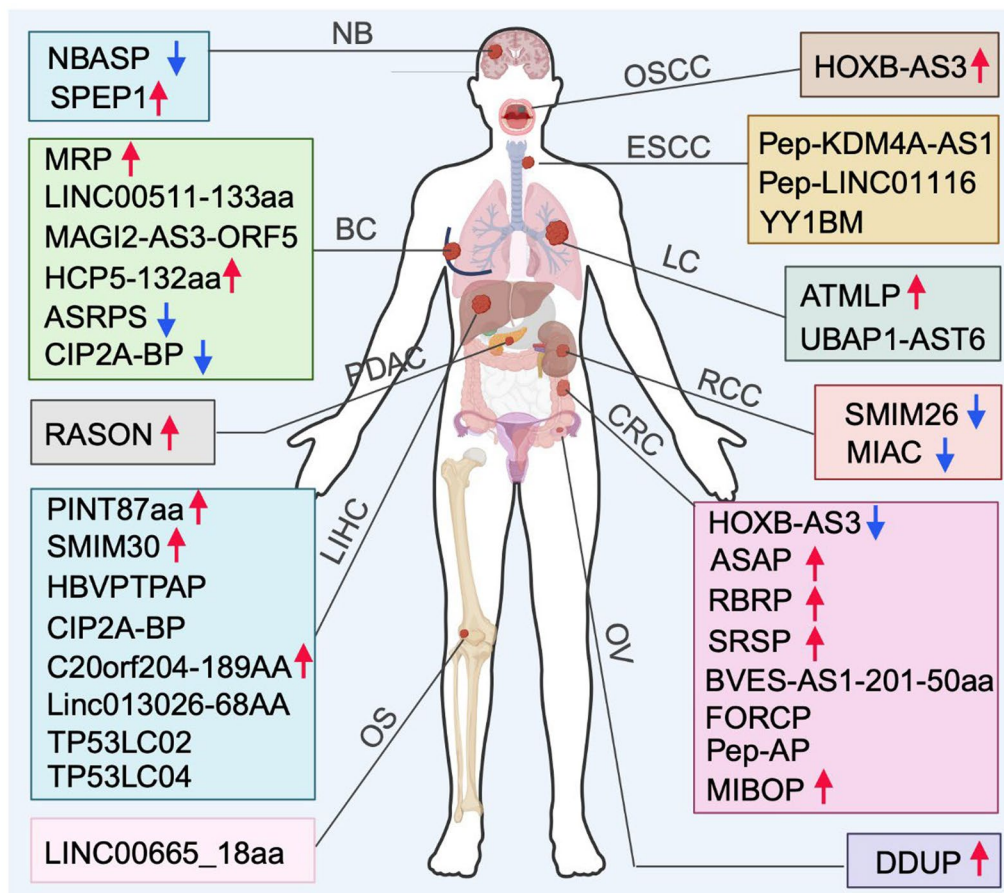
Fig. 9 lncRNA-encoded peptides identified in various human tumor types. NB, Neuroblastoma; BC, Breast cancer; PDAC, Pancreatic ductal adenocarcinoma; LIHC, Liver cancer; OS, Osteosarcoma; OSCC, Oral squamous cell carcinoma; ESCC, Esophageal squamous cell carcinoma; LC, Lung cancer; RCC, Renal cell carcinoma; CRC, Colorectal cancer; OV, Ovarian cancer. Image created with BioRender.com

Transcriptional regulation

LncRNAs engage in transcriptional regulation by interacting with transcription factors, influencing the expression of specific genes. For example, PINT87aa interacts with FOXM1 to disrupt the transcription of tumor suppressor [115], and YY1BM interacts with YY1 to affect the androgen receptor signaling pathway, influencing gene transcription [91]. Additionally, lncRNAs can indirectly participate in transcriptional regulation [129].

Post-transcriptional regulation

LncRNA-encoded peptides can directly bind to splicing factors and participate in RNA splicing. For example, SRSP interacts with SRSF3 to affect the production of different protein isoforms [105]. Some peptides can also interact with RNA-binding proteins and RNA modification enzymes, impacting RNA splicing and stability [30, 33, 104, 131]. For example, HOXB-AS3 interacts with IGF2BP2 to stabilize c-MYC mRNA stability [131].

Translation and post-translation regulation

LncRNA-encoded peptides, like APPLE in AML, are involved in the translation initiation phase, enhancing the synthesis of oncoproteins [173]. Additionally, NBASP and ATMLP illustrate how peptides can mediate protein degradation and regulate protein transport and activity, respectively [120, 128].

Bind to metabolic proteins

Moreover, lncRNA-encoded peptides regulate metabolism by binding to metabolic proteins [108, 125]. For example, ASAP promote metabolic processes by interacting with proteins like ATP synthase, affecting cellular metabolism and energy production [103].

Bind to signaling pathway-related proteins

The modulation of signaling pathways by lncRNA-encoded peptides is another critical area of influence. lncRNA-encoded peptides can both activate and inhibit signaling pathways. For instance, MBOP activates the

MEK1/pERK/MMP2/MMP9 axis [109], while CIP2A-BP inhibits the PI3K/AKT/NF-κB pathway, impacting cancer progression and metastasis [114].

Genomic stability

LncRNA-encoded peptides also involved in DNA damage repair. For example, DDUP, upon phosphorylation induced by DNA damage, interacts with RAD18 to facilitate repair mechanisms, including RAD51C-mediated homologous recombination and PCNA-mediated post-replication repair [39].

In summary, lncRNA-encoded peptides contribute to cancer development and progression through diverse regulatory roles, including transcriptional and post-transcriptional regulation, modulation of translation and protein activity, metabolic regulation, signaling pathway modulation, and maintenance of genomic stability. These functions are executed through their interactions with a range of protein partners, emphasizing their importance in cellular regulation and cancer biology.

Clinical applications of lncRNA-encoded peptides

An escalating number of studies have substantiated the pervasive involvement of lncRNA-encoded peptides in pivotal physiological processes, with an intimate connection to tumorigenesis and tumor progression. This nascent field within lncRNA research holds the key to unlocking the profound implications of these peptides in cancer biology. As such, their clinical deployment as biomarkers or targets for intervention is anticipated to shed new light on their cardinal role in oncology (Table 5, Fig. 10).

Diagnosis biomarker

The quest for novel tumor biomarkers within oncology research is driven by the need for markers that are highly sensitive, specific, reproducible, and ideally non-invasive [176–179]. In this context, circulating micropeptides encoded by lncRNAs emerge as a promising class of biomarkers with the potential to revolutionize cancer diagnostics. The discovery of functional peptides encoded by lncRNAs has opened new avenues in the search for diagnostic biomarkers. These peptides, with their differential expression patterns in malignant versus normal cells, are strong candidates for diagnostic biomarkers. ATMLP, a peptide overexpressed in tumor tissues compared to paracancerous tissues in NSCLC, exemplifies this potential. Its elevated levels in the serum of NSCLC patients, with an AUC of 0.852, suggest its effectiveness as a serum biomarker. Remarkably, ATMLP can prognosticate lung cancer development prior to PET-CT imaging, emphasizing its significant diagnostic value [120]. Similarly, MRP, which intensifies in expression in highly malignant breast cancer cells, has been shown to distinguish patients with and without lymph node metastasis, with an AUC of 0.7112 [30] indicating its potential as a diagnostic tool in breast cancer. However, the diagnostic potential of other lncRNA-encoded peptides and their utility in various biofluids, including urine, warrant further exploration. The promise of these biomarkers lies in their potential for early cancer detection, which is vital for improving patient outcomes. Future research aimed at identifying additional peptides could transform early cancer detection and provide new strategies for timely and effective intervention. As the field advances, the challenge will be to validate these biomarkers in large-scale, multicenter clinical trials to ensure their reliability and utility across diverse patient populations. The successful integration of lncRNA-encoded peptide biomarkers into routine clinical practice will require not only scientific validation but also the development of robust and accessible diagnostic platforms capable of accurately measuring these peptides in patient samples.

Prognosis biomarker

The prognostic utility of lncRNA-encoded peptides in cancer is an emerging field that offers significant promise in predicting disease progression and patient outcomes. These peptides, when identified and characterized, can serve as valuable markers that correlate with late-stage clinical pathological features and poor prognoses, thereby guiding treatment strategies and patient management. Certain peptides have been linked to tumor aggressiveness and survival rates. For instance, in TNBC, specific peptides such as ASRPS and HCP5-132aa have demonstrated a positive correlation with poor OS [99, 113], suggesting their potential as indicators of aggressive tumor behavior and treatment response [99, 113]. Conversely, the presence of the peptide CIP2A-BP has been found to inversely associate with metastasis and OS, indicating its potential as a protective factor or a marker of less aggressive cancer [114]. In CRC, peptides like ASAP, RBRP, and SRSP have been linked to poor OS, with RBRP and SRSP emerging as independent prognostic factors for survival, correlating with advanced clinical stages and higher histological grade [103–105]. The prognostic significance of lncRNA-encoded peptides extends beyond breast and colorectal cancers, with implications in pancreatic ductal adenocarcinoma [124], renal cell carcinoma [125], ovarian cancer [39], etc. These peptides enable patient stratification, leading to more personalized treatment plans and improved survival rates. As research continues to elucidate the complexities of lncRNA-encoded peptides, their role in cancer prognosis becomes increasingly clear, offering a unique perspective into tumor biology and informing clinical decision-making.
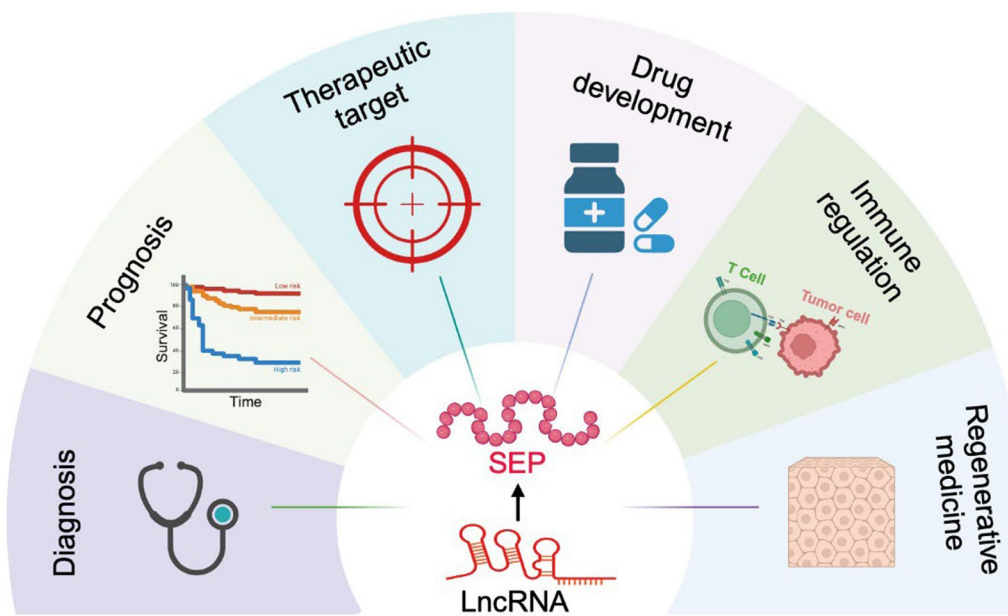
Table 5 The applications of lncRNA-encoded peptides

Cancer	Peptide	Diagnosis	Prognosis	Therapy	References
Triple-Negative Breast Cancer	ASRPS	–	Positively associated with poor OS	Intratumoral injection of ASRPS significantly improved survival in the TNBC mouse xenograft model	[99]
	HCP5-132aa	–	Positively associated with more advanced clinical stages and poor OS	–	[113]
	CIP2A-BP	–	Negatively associated with metastasis and poor OS	CIP2A-BP injection via tail vein reduced lung metastasesnand improved overall survival in lung metastasis models in the MMTV-PyMT mice	[114]
Breast cancer	MRP	May be a diagnostic biomarker for LNM, with AUC value in discriminating BC patients with or without LNM reached 0.7112	Positively correlated with LNM	–	[30]
Colorectal cancer	HOXB-AS3	–	Negatively associated with lower OS	–	[33]
	ASAP	–	Positively correlated with LNM and poor OS	Intratumoral injection of ASAP-targeted CRISPR/Cas9 vector suppressed the growth of CRC patient-derived xenografts	[103]
	RBRP	–	Positively correlated with advanced clinical stages and cancer-related death; high RBRP was an independent prognostic factor for poor survival	–	[104]
	SRSP	–	Positively associated with histological grade, pN status, clinical stage, and cancer-related death; high SRSP level was an independent prognostic factor for poor survival	–	[105]
Pancreatic ductal adenocarcinoma	RASON	–	Positively correlated with IIB-IV stage and poor OS	Peritumoral injection of sh-RASON inhibited xenograft tumor growth and sensitized KRAS mutant pancreatic cancer cells to EGFR inhibitors (cetuximab) in nude mice	[124]
Renal cell carcinoma	SMIM26	–	Negatively associated with cancer-related death	–	[125]
NSCLC	ATMLP	Could be a diagnostic serum marker with AUC of 0.852	Positively correlated with more advanced disease and poor OS	–	[120]
Pan-cancer	pTINCR	–	Posotively correlated with increased OS of patients with bladder carcinoma, PDAC, stomach adenocarcinoma, head and neck squamous cell carcinoma and lung adenocarcinoma	–	[132]

Table 5 (continued)

Cancer	Peptide	Diagnosis	Prognosis	Therapy	References
Ovarian cancer	DDUP	–	Positively correlated with CDDP resistance and relapse and inversely associated with shorter overall/relapse-free survival in patients with ovarian cancer subjected to platinum-based therapy	–	[39]
Neuroblastoma	sPEP1	–	Posotively correlated with advanced INSS stages, MYCN amplification, poor dirreren-tiation, and poor survival	–	[129]
Renal cell carcinoma	MIAC	–	Negatively correlated with advanced stage III-IV and poor OS	Intravenous injections with synthetized MIAC peptides inhibited tumor growth in subcutaneous transplanted tumor model	[126]
Esophageal squamous cell carcinoma	YY1BM	–	–	YY1BM injection intratumorally improved the survival rate of male mice, but not female mice, in ESCC tumors grafted in nude mice	[91]

Fig. 10 Potential applications of lncRNA-encoded peptides. LncRNA-encoded peptides can be utilized in various aspects of oncology, including cancer diagnosis, prognosis, therapeutic target, drug development, immune regulation, and regenerative medicine. SEP, sORF-encoded peptide. Image created with BioRender.com

Therapeutic target

Over the past few decades, cancer treatment has evolved significantly with the introduction of various therapies, including small molecule drugs that target specific signaling pathways, antiangiogenic medications, monoclonal antibodies, and gene therapy [180–183]. The specificity, efficacy, and reduced side effects associated with peptide or protein-targeted drugs make them particularly promising for clinical application. LncRNA-encoded peptides, with their diverse mechanisms of action, are attractive candidates for therapeutic intervention. The peritumoral administration of sh-RASON, which targeting the peptide RASON, exemplifies the therapeutic potential of SEPs. Studies have shown that sh-RASON can inhibit the growth of xenografted tumors and enhance the sensitivity of KRAS-mutant pancreatic cancer cells to epidermal growth factor receptor inhibitors, such as cetuximab, in a murine model [124], highlighting the potential of lncRNA-encoded peptides as therapeutic agents tailored to target specific molecular aberrations in cancer.

Other potential applications

The specificity, high activity, low cytotoxicity, and diminished immunogenicity of lncRNA-encoded peptides make them prime candidates for drug development. Intratumoral injection of ASRPS has been demonstrated to significantly improve survival in TNBC mouse xenograft models [99]. Synthetic MIAC peptides, administered intravenously, have shown promise in inhibiting

tumor growth in RCC models [126]. Additionally, cancer vaccines, which can elicit long-term immunological memory, have garnered significant attention [184–186]. Several cancer vaccines are currently utilized in clinical therapy, including Melacine for melanoma and Cima Vax EGF for lung cancer [187, 188]. Laumont et al. highlighted that tumor-specific antigens (TSA) are ideal targets for immunotherapy and found that most TSA derived from non-coding regions [189], suggesting that TSA derived from non-coding regions could be a promising avenue for cancer immunotherapy. The landscape of cancer vaccines is also being reshaped by these peptides, offering the advantage of long-term immunological memory and sustained antitumor effects. Notably, lncRNA-derived peptides have been shown to elicit a potent antigen-specific CD8+T lymphocyte response, as evidenced by Barczak et al., suggesting their utility in cancer vaccine development [190].

Other studies have explored the role of lncRNA-encoded peptides in immune modulation [191, 192]. Jackson et al. demonstrated that the translation of a novel ORF within the lncRNA Aw112010 is essential for coordinating mucosal immunity during bacterial infection and colitis [193], expanding our understanding of the protein-coding genome and the importance of proteinaceous products from lncRNA in in vivo immune responses. Kikuchi et al. identified a peptide encoded by the lncRNA PVT1 that is predominantly enriched in multiple CRC tissues. The PVT1 peptide was recognized

by patient CD8+tumor-infiltrating lymphocytes and peripheral blood mononuclear cells, indicating the presence of patient immune surveillance [194]. These findings suggest that peptides translated from lncRNAs and presented by HLA class I can be sensed by cancer patient T cells, highlighting their potential in noncoding genomic aberration detection.

As research delves deeper, the regulatory role of lncRNA-encoded peptides in tissue regeneration and stem cell differentiation is coming to light [129, 195, 196]. Matsumoto et al. found that the lncRNA encoding SPAR is downregulated in skeletal muscle upon acute injury. Using a SPAR-polypeptide-specific knockout mouse model created by CRISPR/Cas9, they established that SPAR downregulation enables efficient activation of mTORC1, promoting muscle regeneration [195, 197]. This suggests that lncRNA-encoded peptides could be applied in regenerative medicine, with significant implications for therapeutic approaches following surgical procedures such as hepatectomy.

Prospect and conclusion

The field of lncRNA-encoded peptides in cancer research is burgeoning with potential, offering new insights into the intricate mechanisms underlying tumorigenesis and progression. In the past, the misannotation of genes containing non-canonical ORFs as non-coding RNAs has obscured the significant roles these protein-coding genes play in cancer. However, recent advancements in peptide identification methods, such as Ribo-seq and mass spectrometry, have catalyzed the discovery of SEPs, shedding light on their previously underappreciated functions. To fully harness the potential of SEPs, it is imperative to experimentally validate their translation into functional proteins before delving into their functional studies. In this process, a critical consideration is point-mutation, which may lead to the creation of new ORFs [198, 199]. These peptides often interact with proteins, impacting RNA splicing, and stability, and engaging in cellular metabolism and signaling pathways, thereby participating in biological processes crucial to cancer development. Furthermore, discerning the functions of lncRNA-encoded peptides from those of their parental RNA sequences is imperative. This distinction can be achieved through the overexpression of the full-length lncRNA and its start codon mutant forms, followed by functional assays to determine whether the lncRNA itself or its encoded peptide is responsible for observed biological activities. Post-translational modifications of lncRNA-encoded peptides, analogous to those of mRNA-encoded proteins, are another area that warrants investigation [200, 201]. The interaction between these lncRNA-encoded peptides and the

tumor microenvironment, as well as their role in tumor drug resistance, is a relatively unexplored domain that could significantly enhance our understanding of cancer mechanisms and their therapeutic applications [202–205]. Despite the elucidation of the potential applications of lncRNA-encoded peptides in cancer diagnosis, prognosis, therapeutic targeting, immune modulation, drug development, and regenerative medicine, several unresolved questions and challenges must be addressed before their clinical translation. One of the primary challenges in the application of lncRNA-encoded peptides is the development of an effective delivery system. This system must effectively circumvent the possibility of provoking undesirable immune responses, which can arise from the recognition of these peptides as foreign antigens by antigen-presenting cells and T-cells via the major histocompatibility complex. Such unwanted immune reactions may undermine the therapeutic efficacy of the peptides or even result in detrimental side effects. Extracellular vesicles, with their low immunogenicity and high in vivo stability, are promising candidates for targeted drug delivery [206–208]. Furthermore, recombinant technologies and other advancements have facilitated the production of antibodies that can evade immune surveillance and response [209], which is a field that requires further exploration. The development of optimization of peptide stability and half-life to ensure sustained therapeutic effects is also important. While some lncRNA-encoded proteins have emerged as key regulatory factors in the transcriptional networks of human tumors, the functionality, regulation, and mechanisms of the majority remain elusive. Moving forward, there is a pressing need for large-scale validation to substantiate their biological relevance, development of sensitive detection methods, and optimization of peptide stability and delivery.

Effective resolution of the aforementioned issues will not only refine our understanding of the roles of lncRNA-encoded proteins but also provide a roadmap for future research methods and clues. It is undeniable that the mechanism of lncRNA-encoded micropeptides will spearhead a new wave of research enthusiasm and propel the advancement of the life sciences field. The novel perspectives offered by these findings will undoubtedly contribute to the development of future anti-cancer drugs and tumor biomarkers, offering a new frontier in the battle against cancer.

Abbreviations

lncRNA	Long non-coding RNA
sORF	Small or short open reading frame
ORF	Open reading frames
CDS	Coding DNA sequence
mRNA	Messenger RNAs
Ribo-seq	Ribosome profile sequencing
RPF	Ribosome-protected fragment

GFP	Green fluorescent protein
aa	Amino acid
SEP	SORF-encoded polypeptides
MS	Mass spectrometry
ROS	Reactive oxygen species
CRC	Colorectal cancer
hnRNP A1	Heterogeneous nuclear ribonucleoprotein A1
SRSF3	Serine and Arginine Rich Splicing Factor 3
Sp4	Sp4 Transcription Factor
IGF2BP1	Insulin Like Growth Factor 2 MRNA Binding Protein 1
TALDO1	Transaldolase 1
Src	SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase
mTOR	Mechanistic Target of Rapamycin Kinase
MEK1	Mitogen-Activated Protein Kinase Kinase 1
ERK	Mitogen-Activated Protein Kinase 1
MMP	Matrix Metallopeptidase
MAEA	Macrophage Erythroblast Attacher, E3 Ubiquitin Ligase
RMND5A	Required for Meiotic Nuclear Division 5 Homolog A
Bax	BCL2 Associated X, Apoptosis Regulator
MYC	MYC Proto-Oncogene, BHLH Transcription Factor
GPX4	Glutathione Peroxidase 4
STAT3	Signal Transducer and Activator of Transcription 3
VEGF	Vascular Endothelial Growth Factor
PP2A	Protein Phosphatase 2 Phosphatase Activator
CIP2A	Cellular Inhibitor of PP2A
AKT	AKT Serine/Threonine Kinase
LIHC	Liver hepatocellular carcinoma
JAK	Janus Kinase
PILRA	Paired Immunoglobin Like Type 2 Receptor Alpha
SRC	SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase
YES1	YES Proto-Oncogene 1, Src Family Tyrosine Kinase
Rb	RB Transcriptional Corepressor
E2F1	E2F Transcription Factor 1
FOXM1	Forkhead Box M1
PHB2	Prohibitin 2
LC	Lung cancer
NSCLC	Non-small cell lung cancer
SCLC	Small cell lung cancer
NIPSNAP1	Nipsnap Homolog 1
ESCC	Esophageal squamous cell carcinoma
AR	Androgen receptor
TNBC	Triple-negative Breast cancer
BC	Breast cancer
HNRNPC	Heterogeneous Nuclear Ribonucleoprotein C
EGFR	Epidermal Growth Factor Receptor
PI3K	Phosphatidylinositol-3-kinase
eEF2K	Eukaryotic Elongation Factor 2 Kinase
RCC	Renal cell carcinoma
AGK	Acylglycerol kinase
ccRCC	Clear cell renal cell carcinoma
SLC25A11	Solute Carrier Family 25 Member 11
EREG	Epiregulin
OV	Ovarian cancer
RAD18	RAD18 E3 Ubiquitin Protein Ligase
PCNA	Proliferating Cell Nuclear Antigen
HR	Homologous recombination
FABP5	Fatty Acid Binding Protein 5
eEF1A1	Eukaryotic translation elongation factor 1 alpha 1
SMAD4	SMAD family member 4
CREB1	CAMP response element-binding protein 1
RPS6KA3	Ribosomal protein S6 kinase A3
OSCC	Oral squamous cell carcinoma
LNM	Lymph node metastasis
OS	Overall survival
AUC	Area under the curve
TSA	Tumor-specific antigen

Availability of data and materials
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.

References

1. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. Nature. 2012;489(7414):101–8.
2. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol. 2023;24(6):430–47.
3. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011;43(6):904–14.
4. Navarro-Corcuera A, Sehrawat TS, Jalan-Sakrikar N, Gibbons HR, Pirius NE, Khanal S, et al. Long non-coding RNA ACTA2-AS1 promotes ductular reaction by interacting with the p300/ELK1 complex. J Hepatol. 2022;76(4):921–33.
5. Luo Y, Zheng S, Wu Q, Wu J, Zhou R, Wang C, et al. Long noncoding RNA (lncRNA) EIF3J-DT induces chemoresistance of gastric cancer via autophagy activation. Autophagy. 2021;17(12):4083–101.
6. Yang H, Hu Y, Weng M, Liu X, Wan P, Hu Y, et al. Hypoxia inducible lncRNA-CBSLR modulates ferroptosis through m6A-YTHDF2-dependent modulation of CBS in gastric cancer. J Adv Res. 2022;37:91–106.
7. Zhang Y, Luo M, Cui X, O'Connell D, Yang Y. Long noncoding RNA NEAT1 promotes ferroptosis by modulating the miR-362-3p/MIOX axis as a ceRNA. Cell Death Differ. 2022;29(9):1850–63.
8. Nemeth K, Bayraktar R, Ferracin M, Calin GA. Non-coding RNAs in disease: from mechanisms to therapeutics. Nat Rev Genet. 2024;25(3):211–32.
9. Li K, Wang Z. lncRNA NEAT1: key player in neurodegenerative diseases. Ageing Res Rev. 2023;86: 101878.
10. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. Cancer Cell. 2016;29(4):452–63.
11. Jusic A, Thomas PB, Wettinger SB, Dogan S, Farrugia R, Gaetano C, et al. Noncoding RNAs in age-related cardiovascular diseases. Ageing Res Rev. 2022;77: 101610.
12. DiStefano JK, Gerhard GS. Long noncoding RNAs and human liver disease. Annu Rev Pathol. 2022;17:1–21.
13. Liu SJ, Dang HX, Lim DA, Feng FY, Maher CA. Long noncoding RNAs in cancer metastasis. Nat Rev Cancer. 2021;21(7):446–60.
14. Han L, Huang D, Wu S, Liu S, Wang C, Sheng Y, et al. Lipid droplet-associated lncRNA LIPTER preserves cardiac lipid metabolism. Nat Cell Biol. 2023;25(7):1033–46.
15. Beermann J, Piccoli MT, Viereck J, Thum T. Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. Physiol Rev. 2016;96(4):1297–325.

16. Ma H, Hu T, Tao W, Tong J, Han Z, Herndler-Brandstetter D, et al. A lncRNA from an inflammatory bowel disease risk locus maintains intestinal host-commensal homeostasis. Cell Res. 2023;33(5):372–88.

17. Wang J, Wang W, Ma F, Qian H. A hidden translatome in tumors-the coding lncRNAs. Sci China Life Sci. 2023;66(12):2755–72.

18. Della Bella E, Koch J, Baerenfaller K. Translation and emerging functions of non-coding RNAs in inflammation and immunity. Allergy. 2022;77(7):2025–37.

19. Lu S, Wang T, Zhang G, He QY. Understanding the proteome encoded by "non-coding RNAs": new insights into human genome. Sci China Life Sci. 2020;63(7):986–95.

20. Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science. 2016;351(6270):271–5.

21. van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, et al. The translational landscape of the human heart. Cell. 2019;178(1):242-60 e29.

22. Rion N, Ruegg MA. LncRNA-encoded peptides: more than translational noise? Cell Res. 2017;27(5):604–5.

23. Chothani SP, Adami E, Widjaja AA, Langley SR, Viswanathan S, Pua CJ, et al. A high-resolution map of human RNA translation. Mol Cell. 2022;82(15):2885-99 e8.

24. Sandmann CL, Schulz JF, Ruiz-Orera J, Kirchner M, Ziehm M, Adami E, et al. Evolutionary origins and interactomes of human, young microproteins and small peptides translated from short open reading frames. Mol Cell. 2023;83(6):994-1011 e18.

25. Makarewich CA, Olson EN. Mining for micropeptides. Trends Cell Biol. 2017;27(9):685–96.

26. Wang J, Zhu S, Meng N, He Y, Lu R, Yan GR. ncRNA-encoded peptides or proteins and cancer. Mol Ther. 2019;27(10):1718–25.

27. Wu P, Mo Y, Peng M, Tang T, Zhong Y, Deng X, et al. Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA. Mol Cancer. 2020;19(1):22.

28. Yu J, Wang W, Yang J, Zhang Y, Gong X, Luo H, et al. LncRNA PSR regulates vascular remodeling through encoding a novel protein arteridin. Circ Res. 2022;131(9):768–87.

29. 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(5):1166–83.

30. 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(2):188–202.

31. De Burbano LS, 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(3):31.

32. Lun YZ, Pan ZP, Liu SA, Sun J, Han M, Liu B, et al. The peptide encoded by a novel putative lncRNA HBVPTPAP inducing the apoptosis of hepatocellular carcinoma cells by modulating JAK/STAT signaling pathways. Virus Res. 2020;287: 198104.

33. 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(1):171-84 e6.

34. Pan J, Wang R, Shang F, Ma R, Rong Y, Zhang Y. Functional micropeptides encoded by long non-coding RNAs: a comprehensive review. Front Mol Biosci. 2022;9: 817517.

35. Chen Q, Shen H, Nie F, Sun M. A whole new comprehension about ncRNA-encoded peptides/proteins in cancers. Cancers (Basel). 2022;14(21):5196.

36. Xing J, Liu H, Jiang W, Wang L. LncRNA-encoded peptide: functions and predicting methods. Front Oncol. 2020;10: 622294.

37. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet. 2016;17(1):47–62.

38. 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(2):96–118.

39. 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(14):8060–79.

40. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell. 2013;154(1):240–51.

41. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet. 2014;15(3):193–204.

42. Yang X, Tschaplinski TJ, Hurst GB, Jawdy S, Abraham PE, Lankford PK, et al. Discovery and annotation of small proteins using genomics, proteomics, and computational approaches. Genome Res. 2011;21(4):634–41.

43. Hayden CA, Bosco G. Comparative genomic analysis of novel conserved peptide upstream open reading frames in Drosophila melanogaster and other dipteran species. BMC Genomics. 2008;9:61.

44. Lease KA, Walker JC. The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics. Plant Physiol. 2006;142(3):831–8.

45. Brent MR, Guigo R. Recent advances in gene structure prediction. Curr Opin Struct Biol. 2004;14(3):264–72.

46. Sleator RD. An overview of the current status of eukaryote gene prediction strategies. Gene. 2010;461(1–2):1–4.

47. Hanada K, Zhang X, Borevitz JO, Li WH, Shiu SH. A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection. Genome Res. 2007;17(5):632–40.

48. Cheng H, Chan WS, Li Z, Wang D, Liu S, Zhou Y. Small open reading frames: current prediction techniques and future prospect. Curr Protein Pept Sci. 2011;12(6):503–7.

49. Basrai MA, Hieter P, Boeke JD. Small open reading frames: beautiful needles in the haystack. Genome Res. 1997;7(8):768–71.

50. Claverie JM. Computational methods for the identification of genes in vertebrate genomic sequences. Hum Mol Genet. 1997;6(10):1735–44.

51. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics. 2011;27(13):i275–82.

52. Camargo AP, Sourkov V, Pereira GAG, Carazzolle MF. RNAsamba: neural network-based assessment of the protein-coding potential of RNA sequences. NAR Genom Bioinform. 2020;2(1):lqz024.

53. 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(6): e74.

54. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinformatics. 2014;15(1):311.

55. 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:e13328.

56. 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(6):2960–72.

57. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res. 2017;45(W1):W12–6.

58. 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(10): e61.

59. Rombel IT, Sykes KF, Rayner S, Johnston SA. ORF-FINDER: a vector for high-throughput gene identification. Gene. 2002;282(1–2):33–41.

60. Feng H, Wang S, Wang Y, Ni X, Yang Z, Hu X, et al. LncCat: An ORF attention model to identify LncRNA based on ensemble learning strategy and fused sequence information. Comput Struct Biotechnol J. 2023;21:1433–47.

61. Chen Z, Meng J, Zhao S, Yin C, Luan Y. sORFPred: a method based on comprehensive features and ensemble learning to predict the sORFs in plant LncRNAs. Interdiscip Sci. 2023;15(2):189–201.

62. Liu H, Zhou X, Yuan M, Zhou S, Huang YE, Hou F, et al. ncEP: a manually curated database for experimentally validated ncRNA-encoded proteins or peptides. J Mol Biol. 2020;432(11):3364–8.

63. Dragomir MP, Manyam GC, Ott LF, Berland L, Knutsen E, Ivan C, et al. FuncPEP: a database of functional peptides encoded by non-coding RNAs. Noncoding RNA. 2020;6(4):41.

64. Lv D, Chang Z, Cai Y, Li J, Wang L, Jiang Q, et al. TransLnc: a comprehensive resource for translatable lncRNAs extends immunopeptidome. Nucleic Acids Res. 2022;50(D1):D413–20.

65. Olexiouk V, Crappe J, Verbruggen S, Verhegen K, Martens L, Menschaert G. sORFs.org: a repository of small ORFs identified by ribosome profiling. Nucleic Acids Res. 2016;44(D1):D324–9.

66. 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 Proteomics Bioinf. 2021;19(4):602–10.

67. Choteau SA, Wagner A, Pierre P, Spinelli L, Brun C. MetamORF: a repository of unique short open reading frames identified by both experimental and computational approaches for gene and metagene analyses. Database (Oxford). 2021;2021:baab032.

68. Luo X, Huang Y, Li H, Luo Y, Zuo Z, Ren J, et al. SPENCER: a comprehensive database for small peptides encoded by noncoding RNAs in cancer patients. Nucleic Acids Res. 2022;50(D1):D1373–81.

69. Huang Y, Wang J, Zhao Y, Wang H, Liu T, Li Y, et al. cncRNAdb: a manually curated resource of experimentally supported RNAs with both protein-coding and noncoding function. Nucleic Acids Res. 2021;49(D1):D65–70.

70. Liu T, Wu J, Wu Y, Hu W, Fang Z, Wang Z, et al. LncPep: a resource of translational evidences for lncRNAs. Front Cell Dev Biol. 2022;10: 795084.

71. Housman G, Ulitsky I. Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs. Biochim Biophys Acta. 2016;1859(1):31–40.

72. Yeasmin F, Yada T, Akimitsu N. Micropeptides encoded in transcripts previously identified as long noncoding RNAs: a new chapter in transcriptomics and proteomics. Front Genet. 2018;9:144.

73. Brito Querido J, Diaz-Lopez I, Ramakrishnan V. The molecular basis of translation initiation and its regulation in eukaryotes. Nat Rev Mol Cell Biol. 2024;25(3):168–86.

74. Ramakrishnan V. Ribosome structure and the mechanism of translation. Cell. 2002;108(4):557–72.

75. Sallam T, Sandhu J, Tontonoz P. Long noncoding RNA discovery in cardiovascular disease: decoding form to function. Circ Res. 2018;122(1):155–66.

76. Menschaert G, Van Criekinge W, Notelaers T, Koch A, Crappe J, Gevaert K, et al. Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. Mol Cell Proteomics. 2013;12(7):1780–90.

77. 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(5):363–6.

78. Ingolia NT. Genome-wide translational profiling by ribosome footprinting. Methods Enzymol. 2010;470:119–42.

79. Ruiz-Orera J, Alba MM. Translation of small open reading frames: roles in regulation and evolutionary innovation. Trends Genet. 2019;35(3):186–98.

80. Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MA, Brocard M, et al. Extensive translation of small open reading frames revealed by poly-ribo-seq. Elife. 2014;3: e03528.

81. Martinez TF, Chu Q, Donaldson C, Tan D, Shokhirev MN, Saghatelian A. Accurate annotation of human protein-coding small open reading frames. Nat Chem Biol. 2020;16(4):458–68.

82. Zhu M, Gribskov M. MiPepid: MICROPEPTIDE identification tool using machine learning. BMC Bioinf. 2019;20(1):559.

83. Michel AM, Baranov PV. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. Wiley Interdiscip Rev RNA. 2013;4(5):473–90.

84. Xiong Z, Xu K, Lin Z, Kong F, Wang Q, Quan Y, et al. Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. Nat Cell Biol. 2022;24(6):968–80.

85. Zou Z, Zhang C, Wang Q, Hou Z, Xiong Z, Kong F, et al. Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. Science. 2022;378(6615):abo7923.

86. Scarfiello E, Eichlinger J, Meister G. The double-stranded microRNA precursor. Postepy Biochem. 2024;70(1):57–61.

87. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009;324(5924):218–23.

88. Peeters MKR, Menschaert G. The hunt for sORFs: a multidisciplinary strategy. Exp Cell Res. 2020;391(1): 111923.

89. Pang Y, Liu Z, Han H, Wang B, Li W, Mao C, et al. Peptide SMIM30 promotes HCC development by inducing SRC/YES1 membrane anchoring and MAPK pathway activation. J Hepatol. 2020;73(5):1155–69.

90. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol. 2013;9(1):59–64.

91. 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(13):2790–803.

92. Zhang L, Tang M, Diao H, Xiong L, Yang X, Xing S. LncRNA-encoded peptides: unveiling their significance in cardiovascular physiology and pathology-current research insights. Cardiovasc Res. 2023;119(12):2165–78.

93. 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(6172):1248636.

94. Hartford CCR, Lal A. When long noncoding becomes protein coding. Mol Cell Biol. 2020;40(6):e00528.

95. 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.

96. Zordan RE, Beliveau BJ, Trow JA, Craig NL, Cormack BP. Avoiding the ends: internal epitope tagging of proteins using transposon Tn7. Genetics. 2015;200(1):47–58.

97. Lobbestael E, Reumers V, Ibrahimi A, Paesen K, Thiry I, Gijsbers R, et al. Immunohistochemical detection of transgene expression in the brain using small epitope tags. BMC Biotechnol. 2010;10:16.

98. 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(4):595–606.

99. 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(3):e20190950.

100. Faure E, Delaye L, Tribolo S, Levasseur A, Seligmann H, Barthelemy RM. Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. Biol Direct. 2011;6:56.

101. Sousa ME, Farkas MH. Micropeptide. PLoS Genet. 2018;14(12): e1007764.

102. Orr MW, Mao Y, Storz G, Qian SB. Alternative ORFs and small ORFs: shedding light on the dark proteome. Nucleic Acids Res. 2020;48(3):1029–42.

103. 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. https://doi.org/10.1172/JCI152911.

104. 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(1):1685.

105. 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(10):1903233.

106. 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.

107. 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.

108. 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(1): e53140.

109. 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(9):2338.

110. 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/beta-catenin pathway. Mol Cell Probes. 2023;69: 101913.

111. Zhang Z, Yi Y, Wang Z, Zhang H, Zhao Y, He R, et al. LncRNA MAGI2-AS3-encoded polypeptide restrains the proliferation and migration of breast cancer cells. Mol Biotechnol. 2023;66(6):1409–23.

112. Xing JN, Shang YN, Yu ZL, Zhou SH, Chen WY, Wang LH. LncRNA HCP5-encoded protein contributes to adriamycin resistance through ERK/mTOR pathway-mediated autophagy in breast cancer cells. Genes Dis. 2024;11(4): 101024.

113. 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 (Basel). 2023;15(6):1880.

114. 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(1): e102190.

115. 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(10):4929–44.

116. Yang JE, Zhong WJ, Li JF, Lin YY, Liu FT, Tian H, et al. LINC00998-encoded micropeptide SMIM30 promotes the G1/S transition of cell cycle by regulating cytosolic calcium level. Mol Oncol. 2023;17(5):901–16.

117. Li YR, Zong RQ, Zhang HY, Meng XY, Wu FX. Mechanism analysis of LINC00665 and its peptides CIP2A-BP in hepatocellular carcinoma. Front Genet. 2022;13: 861096.

118. Polenkowski M, de Burbano LS, 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(1):58.

119. 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(6):1026–41.

120. 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(13): e2300314.

121. Xu X, Zhang Y, Wang M, Zhang X, Jiang W, Wu S, et al. A peptide encoded by a long non-coding RNA DLX6-AS1 facilitates cell proliferation, migration, and invasion by activating the wnt/beta-catenin signaling pathway in non-small-cell lung cancer cell. Crit Rev Eukaryot Gene Expr. 2022;32(8):43–53.

122. Lu S, Zhang J, Lian X, Sun L, Meng K, Chen Y, et al. A hidden human proteome encoded by "non-coding" genes. Nucleic Acids Res. 2019;47(15):8111–25.

123. 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(7):1419–36.

124. 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(1):30–45.

125. 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(6): e56282.

126. 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(1):181.

127. 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(8):568.

128. 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(1):714.

129. 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(17):2505–19.

130. 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(6): e0286422.

131. Leng F, Miu YY, Zhang Y, Luo H, Lu XL, Cheng H, et al. A micro-peptide encoded by HOXB-AS3 promotes the proliferation and viability of oral squamous cell carcinoma cell lines by directly binding with IGF2BP2 to stabilize c-Myc. Oncol Lett. 2021;22(4):697.

132. 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(1):6840.

133. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.

134. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet. 2019;394(10207):1467–80.

135. Harbeck N, Gnant M. Breast cancer. Lancet. 2017;389(10074):1134–50.

136. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics. CA Cancer J Clin. 2024;74(1):12–49.

137. Derakhshan F, Reis-Filho JS. Pathogenesis of triple-negative breast cancer. Annu Rev Pathol. 2022;17:181–204.

138. Li Y, Zhang H, Merkher Y, Chen L, Liu N, Leonov S, et al. Recent advances in therapeutic strategies for triple-negative breast cancer. J Hematol Oncol. 2022;15(1):121.

139. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet. 2018;391(10127):1301–14.

140. Nagaraju GP, Dariya B, Kasa P, Peela S, El-Rayes BF. Epigenetics in hepatocellular carcinoma. Semin Cancer Biol. 2022;86(Pt 3):622–32.

141. Siegel RL, Miller KD, Jemal A. Cancer statistics. CA Cancer J Clin. 2020;70(1):7–30.

142. Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. Nat Rev Clin Oncol. 2023;20(9):624–39.

143. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. Non-small cell lung cancer, Version 3.2022, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2022;20(5):497–530.

144. Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. Nat Rev Dis Primers. 2021;7(1):3.

145. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. NCCN Guidelines(R) insights: non-small cell lung cancer, Version 2.2023. J Natl Compr Canc Netw. 2023;21(4):340–50.

146. Yin X, Li Y, Wang H, Jia T, Wang E, Luo Y, et al. Small cell lung cancer transformation: from pathogenesis to treatment. Semin Cancer Biol. 2022;86(Pt 2):595–606.

147. Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, et al. Oesophageal cancer. Nat Rev Dis Primers. 2017;3:17048.

148. Chen GZ, Zhu HC, Dai WS, Zeng XN, Luo JH, Sun XC. The mechanisms of radioresistance in esophageal squamous cell carcinoma and current strategies in radiosensitivity. J Thorac Dis. 2017;9(3):849–59.

149. Morgan E, Soerjomataram I, Rumgay H, Coleman HG, Thrift AP, Vignat J, et al. The global landscape of esophageal squamous cell carcinoma and esophageal adenocarcinoma incidence and mortality in 2020 and Projections to 2040: new estimates from GLOBOCAN 2020. Gastroenterology. 2022;163(3):649-58 e2.

150. Zhang R, Lau LHS, Wu PIC, Yip HC, Wong SH. Endoscopic diagnosis and treatment of esophageal squamous cell carcinoma. Methods Mol Biol. 2020;2129:47–62.

151. GBDPC Collaborators. The global, regional, and national burden of pancreatic cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol. 2019;4(12):934–47.

152. Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. Nat Rev Gastroenterol Hepatol. 2021;18(7):493–502.

153. Linehan WM, Ricketts CJ. The Cancer Genome Atlas of renal cell carcinoma: findings and clinical implications. Nat Rev Urol. 2019;16(9):539–52.

154. Wettersten HI, Aboud OA, Lara PN Jr, Weiss RH. Metabolic reprogramming in clear cell renal cell carcinoma. Nat Rev Nephrol. 2017;13(7):410–9.

155. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.

156. Dalmartello M, La Vecchia C, Bertuccio P, Boffetta P, Levi F, Negri E, et al. European cancer mortality predictions for the year 2022 with focus on ovarian cancer. Ann Oncol. 2022;33(3):330–9.

157. Konstantinopoulos PA, Matulonis UA. Clinical and translational advances in ovarian cancer therapy. Nat Cancer. 2023;4(9):1239–57.

158. Maris JM, Hogarty MD, Bagatell R, Cohn SL. Neuroblastoma. Lancet. 2007;369(9579):2106–20.

159. Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, et al. Neuroblastoma. Nat Rev Dis Primers. 2016;2:16078.

160. Maris JM. Recent advances in neuroblastoma. N Engl J Med. 2010;362(23):2202–11.

161. Nie Z, Peng H. Osteosarcoma in patients below 25 years of age: an observational study of incidence, metastasis, treatment and outcomes. Oncol Lett. 2018;16(5):6502–14.

162. Jafari F, Javdansirat S, Sanaie S, Naseri A, Shamekh A, Rostamzadeh D, et al. Osteosarcoma: a comprehensive review of management and treatment strategies. Ann Diagn Pathol. 2020;49: 151654.

163. Belayneh R, Fourman MS, Bhogal S, Weiss KR. Update on osteosarcoma. Curr Oncol Rep. 2021;23(6):71.

164. Li P, Fang Q, Yang Y, Chen D, Du W, Liu F, et al. Survival significance of number of positive lymph nodes in oral squamous cell carcinoma stratified by p16. Front Oncol. 2021;11: 545433.

165. Dang CV. MYC on the path to cancer. Cell. 2012;149(1):22–35.

166. Dhanasekaran R, Deutzmann A, Mahauad-Fernandez WD, Hansen AS, Gouw AM, Felsher DW. The MYC oncogene—the grand orchestrator of cancer growth and immune evasion. Nat Rev Clin Oncol. 2022;19(1):23–36.

167. Ma WK, Voss DM, Scharner J, Costa ASH, Lin KT, Jeon HY, et al. ASO-based PKM splice-switching therapy inhibits hepatocellular carcinoma growth. Cancer Res. 2022;82(5):900–15.

168. Singh S, Narayanan SP, Biswas K, Gupta A, Ahuja N, Yadav S, et al. Intragenic DNA methylation and BORIS-mediated cancer-specific splicing contribute to the Warburg effect. Proc Natl Acad Sci USA. 2017;114(43):11440–5.

169. Gao Z, Cooper TA. Reexpression of pyruvate kinase M2 in type 1 myofibers correlates with altered glucose metabolism in myotonic dystrophy. Proc Natl Acad Sci USA. 2013;110(33):13570–5.

170. Dohner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. N Engl J Med. 2015;373(12):1136–52.

171. Kelly LM, Gilliland DG. Genetics of myeloid leukemias. Annu Rev Genomics Hum Genet. 2002;3:179–98.

172. Sykes DB, Kfoury YS, Mercier FE, Wawer MJ, Law JM, Haynes MK, et al. Inhibition of dihydroorotate dehydrogenase overcomes differentiation blockade in acute myeloid leukemia. Cell. 2016;167(1):171-86 e15.

173. Sun L, Wang W, Han C, Huang W, Sun Y, Fang K, et al. The oncomicro-peptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. Mol Cell. 2021;81(21):4493-508 e9.

174. 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(3):285–95.

175. Freundlich IM, Wallace JD, Dodd GD. Thermography and the venous diameter ratio in the detection of the nonpalpable breast carcinoma. Am J Roentgenol Radium Ther Nucl Med. 1968;102(4):927–32.

176. Luo H, Wei W, Ye Z, Zheng J, Xu RH. Liquid biopsy of methylation biomarkers in cell-free DNA. Trends Mol Med. 2021;27(5):482–500.

177. Luo G, Jin K, Deng S, Cheng H, Fan Z, Gong Y, et al. Roles of CA19-9 in pancreatic cancer: biomarker, predictor and promoter. Biochim Biophys Acta Rev Cancer. 2021;1875(2): 188409.

178. Wu L, Qu X. Cancer biomarker detection: recent achievements and challenges. Chem Soc Rev. 2015;44(10):2963–97.

179. Best MG, Wesseling P, Wurdinger T. Tumor-educated platelets as a noninvasive biomarker source for cancer detection and progression monitoring. Cancer Res. 2018;78(13):3407–12.

180. Ohdo S, Koyanagi S, Matsunaga N. Implications of biological clocks in pharmacology and pharmacokinetics of antitumor drugs. J Control Release. 2023;364:490–507.

181. Wu Q, Qian W, Sun X, Jiang S. Small-molecule inhibitors, immune checkpoint inhibitors, and more: FDA-approved novel therapeutic drugs for solid tumors from 1991 to 2021. J Hematol Oncol. 2022;15(1):143.

182. Ocana A, Garcia-Alonso S, Amir E, Pandiella A. Refining early antitumoral drug development. Trends Pharmacol Sci. 2018;39(11):922–5.

183. Song X, Liu C, Wang N, Huang H, He S, Gong C, et al. Delivery of CRISPR/Cas systems for cancer gene therapy and immunotherapy. Adv Drug Deliv Rev. 2021;168:158–80.

184. Saxena M, van der Burg SH, Melief CJM, Bhardwaj N. Therapeutic cancer vaccines. Nat Rev Cancer. 2021;21(6):360–78.

185. Lorentzen CL, Haanen JB, Met O, Svane IM. Clinical advances and ongoing trials on mRNA vaccines for cancer treatment. Lancet Oncol. 2022;23(10):e450–8.

186. Sellars MC, Wu CJ, Fritsch EF. Cancer vaccines: building a bridge over troubled waters. Cell. 2022;185(15):2770–88.

187. Tagliamento M, Rijavec E, Barletta G, Biello F, Rossi G, Grossi F, et al. CIMAvax-EGF, a therapeutic non-small cell lung cancer vaccine. Expert Opin Biol Ther. 2018;18(7):829–35.

188. Sosman JA, Sondak VK. Melacine: an allogeneic melanoma tumor cell lysate vaccine. Expert Rev Vaccines. 2003;2(3):353–68.

189. Laumont CM, Vincent K, Hesnard L, Audemard E, Bonneil E, Laverdure JP, et al. Noncoding regions are the main source of targetable tumor-specific antigens. Sci Transl Med. 2018;10(470):eaau5516.

190. Barczak W, Carr SM, Liu G, Munro S, Nicastri A, Lee LN, et al. Long noncoding RNA-derived peptides are immunogenic and drive a potent anti-tumour response. Nat Commun. 2023;14(1):1078.

191. Malekos E, Carpenter S. Short open reading frame genes in innate immunity: from discovery to characterization. Trends Immunol. 2022;43(9):741–56.

192. Zhang Y, Liu Z, Zhong Z, Ji Y, Guo H, Wang W, et al. A tumor suppressor protein encoded by circKEAP1 inhibits osteosarcoma cell stemness and metastasis by promoting vimentin proteasome degradation and activating anti-tumor immunity. J Exp Clin Cancer Res. 2024;43(1):52.

193. 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(7736):434–8.

194. 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(11):1342–53.

195. Tajbakhsh S. lncRNA-encoded polypeptide SPAR(s) with mTORC1 to regulate skeletal muscle regeneration. Cell Stem Cell. 2017;20(4):428–30.

196. Jiang W, Chen Y, Sun M, Huang X, Zhang H, Fu Z, et al. LncRNA DGCR5-encoded polypeptide RIP aggravates SONFH by repressing nuclear localization of beta-catenin in BMSCs. Cell Rep. 2023;42(8): 112969.

197. Matsumoto A, Pasut A, Matsumoto M, Yamashita R, Fung J, Monteleone E, et al. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. Nature. 2017;541(7636):228–32.

198. Liu Z, Xu J, Huang S, Dai W, Zhang W, Li L, et al. Gene point mutation information translation and detection: Leveraging single base extension and CRISPR/Cas12a. Biosens Bioelectron. 2024;247: 115936.

199. Hellebrekers DM, Wolfe R, Hendrickx AT, de Coo IF, de Die CE, Geraedts JP, et al. PGD and heteroplasmic mitochondrial DNA point mutations: a systematic review estimating the chance of healthy offspring. Hum Reprod Update. 2012;18(3):341–9.

200. Lee JM, Hammaren HM, Savitski MM, Baek SH. Control of protein stability by post-translational modifications. Nat Commun. 2023;14(1):201.

201. Vu LD, Gevaert K, De Smet I. Protein language: post-translational modifications talking to each other. Trends Plant Sci. 2018;23(12):1068–80.

202. Vasan N, Baselga J, Hyman DM. A view on drug resistance in cancer. Nature. 2019;575(7782):299–309.

203. Wang X, Jiang W, Du Y, Zhu D, Zhang J, Fang C, et al. Targeting feedback activation of signaling transduction pathways to overcome drug resistance in cancer. Drug Resist Updat. 2022;65: 100884.

204. Kaymak I, Williams KS, Cantor JR, Jones RG. Immunometabolic interplay in the tumor microenvironment. Cancer Cell. 2021;39(1):28–37.

205. Wu T, Dai Y. Tumor microenvironment and therapeutic response. Cancer Lett. 2017;387:61–8.
206. Dutta D, Khan N, Wu J, Jay SM. Extracellular vesicles as an emerging frontier in spinal cord injury pathobiology and therapy. Trends Neurosci. 2021;44(6):492–506.
207. Ning J, Hou X, Hao J, Zhang W, Shi Y, Huang Y, et al. METTL3 inhibition induced by M2 macrophage-derived extracellular vesicles drives anti-PD-1 therapy resistance via M6A-CD70-mediated immune suppression in thyroid cancer. Cell Death Differ. 2023;30(10):2265–79.
208. You Q, Wang F, Du R, Pi J, Wang H, Huo Y, et al. m(6) A reader YTHDF1-targeting engineered small extracellular vesicles for gastric cancer therapy via epigenetic and immune regulation. Adv Mater. 2023;35(8): e2204910.
209. Wong TM, Ross TM. Use of computational and recombinant technologies for developing novel influenza vaccines. Expert Rev Vaccines. 2016;15(1):41–51.

Publisher's Note