



# Functional Micropeptides Encoded by Long Non-Coding RNAs: A Comprehensive Review

Jianfeng Pan<sup>1†‡</sup>, Ruijun Wang<sup>1,2,3,4‡</sup>, Fangzheng Shang<sup>1</sup>, Rong Ma<sup>1</sup>, Youjun Rong<sup>1</sup> and Yanjun Zhang<sup>1,2,3,4\*</sup>

<sup>1</sup>College of Animal Science, Inner Mongolia Agricultural University, Hohhot, China, <sup>2</sup>Key Laboratory of Mutton Sheep Genetics and Breeding, Ministry of Agriculture, Hohhot, China, <sup>3</sup>Key Laboratory of Animal Genetics, Breeding and Reproduction, Hohhot, China, <sup>4</sup>Engineering Research Center for Goat Genetics and Breeding, Hohhot, China

## OPEN ACCESS

### Edited by:

Andrea Cerase,  
Queen Mary University of London,  
United Kingdom

### Reviewed by:

Diego Cotella,  
Università degli Studi del Piemonte  
Orientale, Italy  
Bruno Dallagiovanna,  
Carlos Chagas Institute (ICC), Brazil

### \*Correspondence:

Yanjun Zhang  
imauzj@163.com

### †ORCID:

Jianfeng Pan  
orcid.org/0000-0003-0917-0949

‡These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
RNA Networks and Biology,  
a section of the journal  
Frontiers in Molecular Biosciences

Received: 18 November 2021

Accepted: 24 May 2022

Published: 13 June 2022

### Citation:

Pan J, Wang R, Shang F, Ma R, Rong Y  
and Zhang Y (2022) Functional  
Micropeptides Encoded by Long Non-  
Coding RNAs: A  
Comprehensive Review.  
Front. Mol. Biosci. 9:817517.  
doi: 10.3389/fmolb.2022.817517

Long non-coding RNAs (lncRNAs) were originally defined as non-coding RNAs (ncRNAs) which lack protein-coding ability. However, with the emergence of technologies such as ribosome profiling sequencing and ribosome-nascent chain complex sequencing, it has been demonstrated that most lncRNAs have short open reading frames hence the potential to encode functional micropeptides. Such micropeptides have been described to be widely involved in life-sustaining activities in several organisms, such as homeostasis regulation, disease, and tumor occurrence, and development, and morphological development of animals, and plants. In this review, we focus on the latest developments in the field of lncRNA-encoded micropeptides, and describe the relevant computational tools and techniques for micropeptide prediction and identification. This review aims to serve as a reference for future research studies on lncRNA-encoded micropeptides.

**Keywords:** lncRNA, micropeptide, sORF, Ribo-seq, coding potential prediction

## 1 INTRODUCTION

Non-coding RNAs (ncRNAs) are generally considered as a class of RNAs that lack protein-coding ability. Based on their regulatory functions, ncRNAs can be categorized as long non-coding RNAs (lncRNAs), primary miRNAs (pri-miRNAs), circular RNAs (circRNAs), among others (Beermann et al., 2016; Khalili-Tanha and Moghbeli, 2021). lncRNAs have transcriptional length that exceeds 200 nucleotides, being initially defined as “transcriptional noise” (Choi et al., 2019). However, with the emergence and increasing use of high-throughput technologies such as ribosome profiling sequencing (Ribo-Seq) and ribosome-nascent chain complex sequencing (RNC-Seq), it has been demonstrated that lncRNAs have short open reading frames (sORFs) encoding micropeptides (Ruiz-Orera et al., 2020). However, the function of most encoded micropeptides has been overlooked due to their low molecular weight (100 amino acid residues or fewer).

lncRNAs are mainly transcribed by RNA polymerase II (Pol II) and have a structure similar to mRNA, including a 7-methylguanosine triphosphate (m7G-cap) at the 5' end and a poly(A) tail at the 3' end (Zhang et al., 2019; Statello et al., 2021), suggesting that lncRNAs may have a translational function comparable to that of mRNAs. However, unlike mRNAs, lncRNAs have distinct transcription, processing, and modification processes (Quinn and Chang, 2016). In addition, poor conservation and spatiotemporal specificity of lncRNAs expression greatly hinder the exploration of lncRNA coding potential (Nitsche and Stadler, 2017).

Previous studies have been demonstrated that, in addition to lncRNAs with coding potential, pri-miRNAs and circRNAs also possess sORFs encoding functional micropeptides. In this

context, pri-miRNAs are a distinct type of lncRNAs of which the length is within the range of hundreds to thousands of nucleotides, being produced by Pol II. Thus, in this sense, pri-miRNAs may be similar to lncRNAs due to the micropeptide-encoding ability (Laouressergues et al., 2015; Lv et al., 2016; Wu P et al., 2020; Prasad et al., 2021). In contrast, circRNAs are transcribed by Pol II without the 5' cap and the 3' poly(A) tail, being thus resistant to digestion by RNaseR and having a ten-fold longer half-life compared to linear RNA (Lei et al., 2020). In addition, there is evidence that circRNAs possess highly conserved sORFs encoding functional micropeptides in a 5' cap-independent manner. Since circRNAs have a unique covalently closed structure sORFs therein circulate across the splicing site and even beyond their length (Shi et al., 2020; Wu S et al., 2020), and they can potentially encode micropeptides containing more than 100 amino acids in length. Collectively, these observations indicate that ncRNAs have potential applications in the field of encoding micropeptides which need to be further explored.

This review outlines the translational mechanisms of lncRNA-encoded micropeptides as well as the computational tools and techniques related to micropeptide prediction and identification. A discussion is also proposed on the latest research advancements of therapies based on lncRNA-encoded micropeptides, such as those applied to skeletal muscle, innate immunity, cancer, among others. Finally, it is summarized future outlooks on the current research landscape of lncRNA-encoded micropeptides, aiming to provide positive strategies, and novel insights for the future of micropeptide research.

## 2 TRANSLATIONAL MECHANISMS OF LNCRNA-ENCODED MICROPEPTIDES

lncRNAs with coding ability have been described as early as in 2014. Ruiz-Orera et al. (2014) found that the majority of lncRNAs expressed in cells from six different species (human, mice, fish, flies, yeast, and plant) were linked to ribosomes. In addition, the ribosomal conservation pattern was consistent with the translation of micropeptides (Ruiz-Orera et al., 2014). Moreover, lncRNAs showed coding potential and structural constraints similar to those of nascent protein-coding sequences, suggesting that lncRNAs may play an important role in the *de novo* evolution of proteins (Ruiz-Orera et al., 2014). In 2014, Pauli et al. (2014) identified a conserved peptide encoded by an lncRNA, termed Toddler, involved in zebrafish embryogenesis. It has been demonstrated that both the lack and overexpression of this peptide reduced the movement of mesodermal cells during zebrafish gastrulation (Pauli et al., 2014). In a study by Chen et al. (2020), a strategy combining ribosome profiling, mass spectrometry (MS)-based proteomics, microscopy, and CRISPR-based genetic screening was used to explore and characterize widespread translation of functional micropeptides as well as determine the protein-coding potential of complex genomes. Using this screening strategy, hundreds of non-canonical lncRNA coding DNA sequences (CDSs) encoding stable functional micropeptides were identified as essential for cell growth and whose disruption triggered specific and robust

transcriptomic and phenotypic changes in human cells (Chen et al., 2020). Thus, lncRNA-encoded micropeptides have been gaining increasing attention in research, being less considered a “translation noise” but rather functional micropeptides.

In 2015, Ji et al. (2015) identified that 40% of lncRNAs and pseudogene RNAs expressed in human cells are translated. In addition, these authors verified that approximately 35% of mRNA-encoding genes are translated upstream of primary protein-coding regions (uORFs), and 4% are translated downstream (dORFs) (Ji et al., 2015). In this same study, it has been demonstrated that translated lncRNAs are preferentially localized in the cytoplasm, while non-translated lncRNAs are preferentially found in the nucleus (Ji et al., 2015). Translation efficiency of cytoplasmic lncRNAs was shown to be comparable to that of mRNAs, indicating that sORFs of cytoplasmic lncRNAs are protected by ribosomes and involved in translation (Ji et al., 2015). Common ORFs are defined as the DNA sequence found between the start (ATG or AUG) and stop codons (TAG or TGA) (Sieber et al., 2018), whereas sORFs typically possess less than 300 nucleotides in length, and longer sORFs are more likely to be translated (Pueyo et al., 2016; Orr et al., 2020). It has also been found that regulatory elements upstream of ORFs, e.g., internal ribosome entry site (IRES), N6-methyladenosine (m6A) methylation conserved sites, can mediate micropeptide translation (Wu P et al., 2020; Charpentier et al., 2022). IRES elements are important regulatory RNA sequences that do not rely on 5' cap for translation, which mostly occur in the 5' untranslated region (5' UTR) upstream of the ORF controlled by IRES (Zhao et al., 2018). By recruiting ribosomes and then proceeding to ribosome assembly, translation of sORFs into micropeptides can occur. In addition, IRES elements may also be present between and within ORFs to mediate translation, and lncRNAs with IRES elements can be translated into micropeptides based on consecutive sORFs (Stoneley and Willis, 2004; King et al., 2010; Hanson et al., 2012; Carbonnelle et al., 2013). Furthermore, it has been demonstrated that m6A can drive endogenous ncRNA translation, in particular the translation of circRNA, and hundreds of endogenous circRNA with translation potential have been identified (Yang et al., 2017), which greatly enlarges our study. Moreover, it can be speculated that m6A could also potentially drive endogenous lncRNA translation.

The translational capacity of lncRNAs is regulated by proteins in addition to post-transcriptional regulation mechanisms (e.g., splicing, polyadenylation). The micropeptide STORM encoded by linc00689 is regulated by phosphorylation of the eukaryotic translation initiation factor 4E (eIF4E) which is mediated by TNF- $\alpha$  and mammalian Ste20-like kinase (MST1) (Min et al., 2017). eIF4E is an mRNA cap-binding protein that is a general initiation factor allowing for mRNA-ribosome interaction and cap-dependent translation in eukaryotic cells (Ross-Kaschitzka and Altmann, 2020). Phosphorylation of eIF4E was found to weaken the interaction with 5' cap while inhibiting mRNA translation, but enhanced the association of active polyribosomes with lncRNA (Min et al., 2017).

Nonsense-mediated decay (NMD) is an important mechanism for mRNA quality monitoring. NMD is triggered by long 3' UTR,

**TABLE 1** | Advantages and disadvantages of translation-omics related techniques.

Techniques	Advantages	Disadvantages	References
Polysome profiling	RNC-mRNA can be obtained; any length, sequence variation, number of ribosomes on each mRNA can be detected	It is difficult to perform in-depth analysis of all translated mRNA	Chasse et al. (2017)
RNC-Seq	It can effectively reveal the full-length information of the RNA being translated, including abundance, and type	Prone to ribosome dissociation or RNA degradation after cell lysis; low sequencing precision; no access to ribosome, ORF, uORF information	Wang L et al. (2013)
TRAP-Seq	RNC-mRNA can be obtained; avoids contamination by eliminating the need for ultracentrifugation; it has the advantage of isolating RNC-mRNA from complex tissues and specific cell types	Stably transfected cell lines need to be established to produce labeled ribosomal proteins; over-labeling of ribosomal proteins may alter the structure and properties of the ribosome	Inada et al. (2002); Heiman et al. (2014)
Ribo-Seq	Accurately locates genes under translation; accurately quantifies gene translation levels; instantaneously measures translation efficiency; obtains ribosome position, density, ORF, and uORF information	Complex experiment; expensive; can only detect ribosome-protected RNA fragments; poor reproducibility	Ingolia et al. (2009); Ingolia et al. (2019)

and intronless genes may be insensitive to NMD (Tan et al., 2021). Wery et al. (2016), using ribosomal analysis, described that actively translated lncRNA sORFs with long 3' UTR were responsive to NMD, suggesting that NMD may also be a monitoring mechanism for lncRNA translation. In addition, it has been suggested that micropeptides encoded by lncRNAs interact with the mRNA decapping protein complex which is responsible for the removal of the 5' cap from mRNA to promote 5' to 3' decay (DLima et al., 2017). Simultaneously, micropeptides encoded by lncRNAs can also be co-localized with mRNA decay-associated RNA protein granules to alter the steady-state levels of cellular NMD targets (D'Lima et al., 2017). Collectively, the above results illustrate that lncRNAs have mRNA-like translational functions of which mechanisms are regulated by a variety of regulatory proteins as well as by NMD monitoring. In addition, micropeptides encoded by lncRNAs have been shown to regulate NMD homeostasis. These findings suggest that micropeptides have a promising regulatory role, which requires further studies in order to elucidate currently unknown regulatory mechanisms.

### 3 PREDICTION AND IDENTIFICATION OF LNCRNAs CODING ABILITY

#### 3.1 Sequencing Analysis Based on “Omics” Techniques

Most of current studies on lncRNA-encoded micropeptides are based on data obtained by ribosome analysis (Ruiz-Orera and Alba, 2019). However, “omics” techniques have been considered an important tool to study the coding capacity of lncRNAs. In this context, translational omics analysis has been commonly used, and mainly relies on four techniques (Ingolia et al., 2009; Ingolia et al., 2019; Zhao et al., 2019): polysome profiling, ribosome-nascent chain complex sequencing (RNC-Seq), ribosome affinity purification (TRAP-Seq), and ribosome profiling (Ribo-Seq) (Table 1).

Ribo-seq is based on high-throughput sequencing to detect RNA translation at the whole genome level. This technique is based on the following strategies: 1) degradation of ribosome-free RNA fragments and ribosome-nascent peptide chain complexes with low concentrations of RNase; 2) removal of ribosomes; 3)

detection of small fragments (26–34 bp in length) of RNA undergoing translation whilst protected by ribosomes using second-generation sequencing technology (Ingolia et al., 2012; Ingolia et al., 2019). These ribosome-protected RNA fragments are termed ribosome footprints (RFs), which reveal the location and density of the ribosome during the translation of RNA fragments (Ingolia, 2016). Although Ribo-seq enables the detection of fragments of 26–34 bp in length undergoing translation, it usually generates 20–30 GB of data, which might represent nearly the entirety of translated sequences of an organism, thus predicting translation more accurately (Ingolia et al., 2019). Taken together, Ribo-seq has several advantages such as precise localization of genes being translated, accurate quantification of translation levels, and transient measurement of translation efficiency. In addition, compared with conventional RNC-seq, Ribo-seq enables a more accurate prediction of translated protein abundance, thus yielding more reliable results, with a lower rate of false positives.

Ribo-seq can help to unravel translational mechanisms when combined with RNA-seq, small RNA-seq, m6A-seq, single-cell RNA (scRNA)-seq, and other sequencing methods (Calviello and Ohler, 2017; La Manno, 2019; Zong et al., 2021). Thus, in the study of lncRNAs with coding ability with the aim to unravel the greatest potential for association with certain species or diseases, it is recommended to combine Ribo-seq with RNA-seq or lncRNA-seq (Yan et al., 2021). On this basis, new micropeptides encoded by lncRNAs can be further explored and validated by combined analysis with peptidomics (Zhang et al., 2014; Vitorino et al., 2021). Peptidomics comprises the study of endogenous micropeptides or small proteins in organisms and/or compartments (cells, tissues, body fluids), being generally considered proteomics of molecules of low molecular weight (Baggerman et al., 2004). Using peptidomics it is possible to effectively enrich endogenous peptides of low molecular weight and/or low abundance, thus enabling their identification by liquid chromatography-tandem mass spectrometry (LC-MS/MS), hence a more accurate micropeptide functional annotation and differential database construction (Fabre et al., 2021). Therefore, Ribo-seq can be combined with RNA-seq or lncRNA-seq and peptidomics to obtain the most comprehensive characterization of potentially

translated lncRNAs. Furthermore, considering the existence of translational regulation, correlation between transcriptome and proteome data tends to be low (Kumar et al., 2016). Thus, quantification at the translation level creates the possibility of establishing a better correlation between multi-omics data and an in-depth study of the mechanisms underlying translational regulation. Collectively, Ribo-seq can be considered an important method for the study of lncRNAs coding ability, which, when combined with multi-omics analysis, constitutes an important strategy to further validate obtained data and explore the functions of novel micropeptides encoded by lncRNAs.

### 3.2 Application of Bioinformatics to Predict the Coding Potential of lncRNAs

With the advent of high-throughput sequencing technologies, several lncRNA transcripts with coding potential have been found in different organisms. However, identification, prediction, and characterization of lncRNAs with coding ability can be challenging. Therefore, a wide variety of computational tools, software, and databases have been created for predicting and distinguishing non-coding and coding transcripts, among which can be cited sORF finder (Hanada et al., 2010), PhyloCSF (Lin et al., 2011), CNCI (Sun K et al., 2013), CPC2 (Kang et al., 2017), and CNIT (Guo et al., 2019).

Coding Potential Calculator (CPC) is a widely used method for assessing the coding potential of transcripts based on sequence features and the use of vector machines. CPC can distinguish coding and non-coding transcripts with high accuracy, but it requires sequence-to-sequence comparisons which relatively delays the analysis (Kong et al., 2007). The upgraded version CPC2 was released in 2017, which contains an accurate coding potential calculator which assesses the intrinsic features of transcript sequences, allowing for a faster and more reliable assessment of RNA coding potential (Kang et al., 2017). In addition, CPC2 is species-neutral, being thus applicable to the analysis of transcriptome data of non-model organisms (Kang et al., 2017). Furthermore, CPC2 is one of the latest lncRNA identification tools released, thus representing a considerable advancement in lncRNA coding potential identification.

In addition, predicting potential sORF in lncRNAs using bioinformatics or software is a current research trend. The ORF Finder analysis tool has been widely used and can predict all possible sORFs of lncRNAs with the corresponding amino acid sequences (Sayers et al., 2021). Subsequently, the deduced amino acid sequence can be queried against the Pfam (Mistry et al., 2021) and conserved domain database (CDD) (Lu et al., 2020) to further confirm the predicted sORFs.

In addition, conserved sequences of the coding region of lncRNAs can be determined by a variety of tools, e.g., PhyloCSF (Lin et al., 2011), RNAcode (Washietl et al., 2011), among others. A large proportion of lncRNA-encoded micropeptides are associated with intracellular membrane structures (Pang et al., 2020). The transmembrane segment of micropeptides can be predicted using the tools TMHMM or

TMpred to determine the localization of the target micropeptide in the cell (intracellular, transmembrane or extracellular) (Krogh et al., 2001; Duvaud et al., 2021). Signal peptide prediction of transmembrane micropeptides can be conducted in SignalP further helped the researchers to predict the mode of action of micropeptides (Petersen et al., 2011; Almagro Armenteros et al., 2019). Subsequently, hydropathicity or hydrophobicity mapping of micropeptides is performed using ProtScale in the ExPasy Bioinformatics Resource database (Duvaud et al., 2021), which in turn provides a reference for the identification of micropeptide transmembrane regions. In addition, the SWISS-MODEL in the ExPasy database can be applied to homology modelling of protein structures and complexes to generate reliable protein models (Waterhouse et al., 2018), which can enable an in-depth analysis of the biological functions and structural features of lncRNA-encoded micropeptides. These bioinformatics prediction tools have been widely used; however, there are several other databases and computational tools to predict protein structure and lncRNAs coding potential which have not been mentioned herein and still require further validation by the research community.

It is known that RNAs can be classified based on their protein-coding ability into ncRNA and mRNA. However, with research advancements, an increasing number of ncRNAs with coding functions and mRNAs with non-coding functions have been described, which contrasts previous knowledge of RNA classification and function. Simultaneously, the emergence of bifunctional RNAs has stretched the boundaries between coding and non-coding RNAs and prompted researchers to reconsider the specific roles and the underlying mechanisms of RNAs in function and evolution (Nam et al., 2016). This suggests that bifunctional RNAs, i.e., those with coding and non-coding functions (cncRNA), may be worth exploring further (Huang et al., 2021). In 2020, Huang et al. (2021) established a cncRNAdb database following a comprehensive characterization of cncRNA; the current version of this database contains approximately 2,600 functional entries with experimental evidence of cncRNAs, comprising over 2,000 RNAs found in more than twenty species (including over 1,300 translated ncRNAs and over 600 untranslated mRNAs). This database can be used to further elucidate the functions and mechanisms of cncRNA, thus providing a valuable resource for future studies. Other databases also allow annotation of coding-capable lncRNAs, e.g., LNCipedia (Volders et al., 2019), lncCAR (Zheng et al., 2019), among others. All relevant computational tools, software and databases cited herein are summarized in **Tables 2–4**.

### 3.3 Experimental Identification of lncRNAs Coding Potential

Through combined multi-omics analysis and bioinformatics prediction, several lncRNAs with promising application in research and coding potential have been described. After prediction, these lncRNAs require experimental identification. Firstly, RNA-fluorescence *in situ* hybridization (RNA-FISH) technology is used to determine lncRNA localization in the



**TABLE 2 |** ORF prediction and evaluation related calculation tools.

Name	Characteristics	Website	References
CPC	Use sequence features and support vector machines (SVM) to evaluate the protein coding potential of transcripts; assessing the scope, quality, integrity of ORFs	<a href="http://cpc.cbi.pku.edu.cn">http://cpc.cbi.pku.edu.cn</a>	Kong et al. (2007)
sORF finder	Package for identifying sORF with high encoding potential	<a href="http://evolver.psc.riken.jp/">http://evolver.psc.riken.jp/</a>	Hanada et al. (2010)
PhyloCSF	Based on the formal statistical comparison of phylogenetic codon models, the nucleotide sequence alignment of multiple species is analyzed to determine whether it may represent a conserved protein coding region; it can delimit likely protein-coding ORFs within transcript models that include untranslated regions	<a href="http://compbio.mit.edu/PhyloCSF">http://compbio.mit.edu/PhyloCSF</a>	Lin et al. (2011)
RNAcode	Comparison of conserved regions in coding and non-coding regions in sequence data and evaluation of coding potential; analysis of sORF or bifunctional RNAs	<a href="http://wash.github.com/rnacode">http://wash.github.com/rnacode</a>	Washietl et al. (2011)
CNCI	Classification of protein-coding and long non-coding transcripts using sequence intrinsic composition (adjacent nucleotide triplets) (SVM-based)	<a href="http://www.bioinfo.org/software/cnci">http://www.bioinfo.org/software/cnci</a>	Sun L et al. (2013)
CPAT	The coding potential assessment tool uses a permutation-free logistic regression model that can ORFs size and coverage to be assessed	<a href="http://code.google.com/p/cpat/">http://code.google.com/p/cpat/</a>	Wang T et al. (2013)
iSeeRNA	Identification of long intergenic non-coding RNA (lincRNA) transcripts in transcriptome sequencing data (SVM-based)	<a href="http://www.myogenesisdb.org/iSeeRNA">http://www.myogenesisdb.org/iSeeRNA</a>	Sun K et al. (2013)
PLEK	Efficient alignment-free computational tool for differentiating coding and non-coding transcripts in RNA-seq transcriptomes of species lacking a reference genome (SVM-based)	<a href="https://sourceforge.net/projects/plek/files/">https://sourceforge.net/projects/plek/files/</a>	Li et al. (2014)
LncRNA-ID	The tool calculates the coding potential of transcripts based on a machine learning model (random forest) and multiple features	<a href="https://github.com/zhangy72/LncRNA-ID">https://github.com/zhangy72/LncRNA-ID</a>	Achawanantakun et al. (2015)
lncRNA-MFDL	By fusing multiple features and using deep learning classification algorithms to identify human lncRNA, coding and long non-coding RNA can be quickly distinguished	<a href="http://compgenomics.utsa.edu/lncRNA_MDFL/">http://compgenomics.utsa.edu/lncRNA_MDFL/</a>	Fan and Zhang, (2015)
COME	A multi-feature-based coding potential calculation tool for lncRNA coding potential assessment	<a href="https://github.com/lulab/COME">https://github.com/lulab/COME</a>	Hu et al. (2017)
CPC2	A fast and accurate coding potential calculator based on intrinsic sequence features for ORF feature evaluation (SVM-based)	<a href="http://cpc2.cbi.pku.edu.cn">http://cpc2.cbi.pku.edu.cn</a>	Kang et al. (2017)
CNIT	A tool for identifying protein coding and long non-coding transcripts based on intrinsic sequence composition (upgraded version of CNCI)	<a href="http://cnit.noncode.org/CNIT">http://cnit.noncode.org/CNIT</a>	Guo et al. (2019)
ORF Finder	A software provided by NCBI that performs six-frame translation of a nucleotide sequence, allowing all possible ORFs to be inferred	<a href="https://www.ncbi.nlm.nih.gov/orffinder/">https://www.ncbi.nlm.nih.gov/orffinder/</a>	Sayers et al. (2021)

**TABLE 3 |** Micropeptide information and structure-related prediction tools.

Name	Characteristics	Website	References
TMHMM	Prediction software for transmembrane structural domains (using hidden Markov model to predict the topological structure of transmembrane proteins)	<a href="http://www.cbs.dtu.dk/services/TMHMM/">http://www.cbs.dtu.dk/services/TMHMM/</a>	Krogh et al. (2001)
TMpred	Predict the transmembrane regions and directions	<a href="https://embnet.vital-it.ch/software/TMPRED_form.html">https://embnet.vital-it.ch/software/TMPRED_form.html</a>	Duvaud et al. (2021)
SignalP	Signal peptide prediction tool	<a href="http://www.cbs.dtu.dk/services/SignalP/">http://www.cbs.dtu.dk/services/SignalP/</a>	Almagro Armenteros et al. (2019)
ProtScale	An online tool for mapping the hydrophilic and hydrophobic atlas of proteins	<a href="https://web.expasy.org/protscale/">https://web.expasy.org/protscale/</a>	Duvaud et al. (2021)
SWISS-MODEL	An automated protein structure homology modeling platform that uses comparative methods to generate protein 3D models	<a href="https://swissmodel.expasy.org">https://swissmodel.expasy.org</a>	Waterhouse et al. (2018)
I-TASSER	An integrated platform for automated protein structure and function prediction based on the sequence- to-structure-to-function paradigm	<a href="https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/2019-nCov/">https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/2019-nCov/</a>	Yang et al. (2015)
AlphaFold2	A tool for accurately predicting the 3D structure of a protein based on its amino acid sequence	<a href="https://github.com/deepmind/alphafold">https://github.com/deepmind/alphafold</a>	Jumper et al. (2021)
RoseTTAFold	A tool for accurate structure prediction of proteins and protein complexes using three-track neural networks	<a href="https://github.com/RosettaCommons/RoseTTAFold">https://github.com/RosettaCommons/RoseTTAFold</a>	Baek et al. (2021)

cell; since translation of micropeptides mostly occurs in the cytoplasm, determining their localization improves inferring their potential function of lncRNA-encoded micropeptides (Huang et al., 2017; Yan et al., 2021). A FLAG/HA-tag system was cloned before the stop codon of the potential sORF of this lncRNA, and the fusion sequence containing the FLAG/HA-tag was cloned into a plasmid vector for *in vitro* cell transfection (Pang et al., 2020); after transfection into target cell line or wild-

type cells, the relative expression of the micropeptide was detected by western blotting and immunofluorescence assays using anti-FLAG/HA tag antibodies (Wu S et al., 2020). Alternatively, sORFs of lncRNAs can be fused to the N-terminal end of green fluorescent protein (GFP) vectors with mutated start codons, and the relative expression of micropeptides can be detected by western blotting and immunofluorescence assays with anti-GFP antibodies (Zhu

**TABLE 4** | Commonly used databases for micropeptide research.

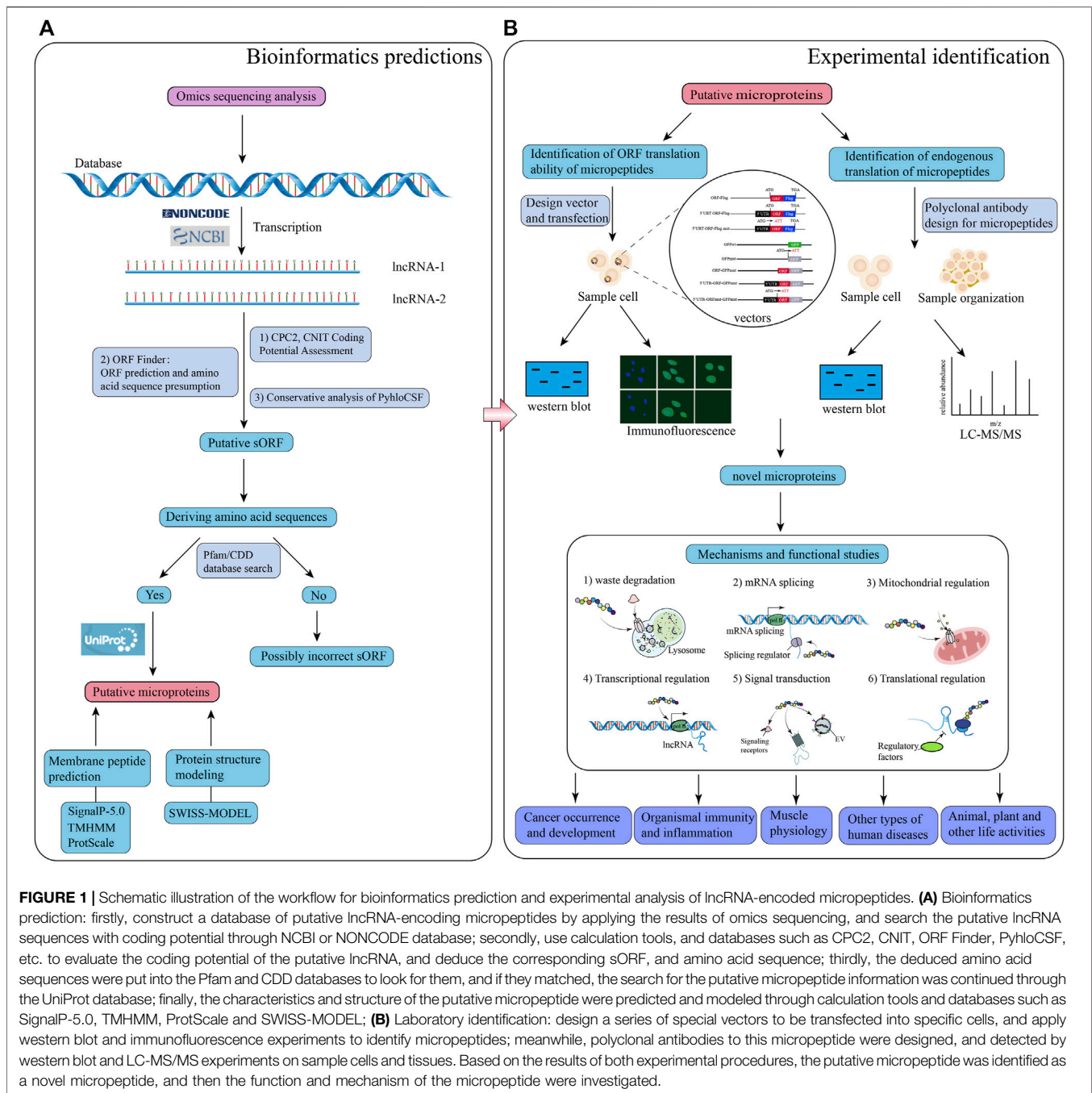
Name	Characteristics	Website	References
BLAST	A tool for similarity analysis in protein databases or gene databases to find sequences that are similar to the query sequence. This includes patterns such as blastp, blastx, etc	<a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a>	Sayers et al. (2021)
Pfam	A database that classifies protein sequences into families and domains, which can be queried for protein conserved structural domains	<a href="http://pfam.xfam.org/">http://pfam.xfam.org/</a>	Mistry et al. (2021)
CDD	NCBI conserved domain database, annotated biomolecular sequences with evolutionarily conserved protein domain footprint positions, as well as functional sites deduced from these footprints	<a href="https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml">https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml</a>	Lu et al. (2020)
cncRNADB	Manually manage a resource database of bifunctional RNA (cncRNA) with protein-coding and non-coding functions	<a href="http://www.rna-society.org/cncrnadb/">http://www.rna-society.org/cncrnadb/</a>	Huang et al. (2021)
LNCipedia	A public database for storing lncRNA sequences and annotation information	<a href="https://lncipedia.org/">https://lncipedia.org/</a>	Volders et al. (2019)
lncAR	A comprehensive resource for lncRNA from cancer arrays (including lncRNA coding information)	<a href="https://lncar.renlab.org/">https://lncar.renlab.org/</a>	Zheng et al. (2019)
NONCODE	A database annotated with a large amount of lncRNA information	<a href="http://www.noncode.org/">http://www.noncode.org/</a>	Zhao et al. (2016)
UCSC	Genome Browser database that provides high quality visualization of genomic data and genome annotation. Has tools such as BLAT, track hubs, etc. for viewing, analyzing and downloading data	<a href="https://genome.ucsc.edu">https://genome.ucsc.edu</a>	Navarro Gonzalez et al. (2021)
UniProt	The most comprehensive database of protein sequence and annotation information, consisting of UniProtKB, UniRef, and UniParc, and integrating data from three major databases, swiss-prot, TrEMBL, and PIR-PSD	<a href="https://www.uniprot.org/">https://www.uniprot.org/</a>	UniProt, (2021)
ExPasy	A database of reliable and most advanced bioinformatics service tools and resources is stored. Has tools such as protscale, TMpred, etc. for viewing, analyzing, and downloading data	<a href="https://www.expasy.org/">https://www.expasy.org/</a>	Duvaud et al. (2021)
LncPep	The lncRNA coding peptides database	<a href="http://www.shengllabs.com/LncPep/">http://www.shengllabs.com/LncPep/</a>	Liu et al. (2022)
SPENCER	A comprehensive database for small peptides encoded by noncoding RNAs in cancer patients	<a href="http://spencer.renlab.org">http://spencer.renlab.org</a>	Luo et al. (2022)

et al., 2020). Immunoprecipitation (Co-IP) in tandem with mass spectrometry (MS) analysis of ORF-GFP fusion peptides can be performed using anti-GFP antibodies to further identify lncRNA-translated micropeptides (Wang L et al., 2020). However, since most GFP-tags are larger in size than lncRNA-encoded micropeptides, and GFP-tagged micropeptides may alter the phenotype of micropeptides, FLAG-tag fused constructs are mostly used in experimental identification of lncRNAs coding potential. In addition, the CRISPR-Cas9 system can be used to knock in FLAG-tags before the stop codon of the lncRNAs locus in target cells, and the relative expression of the resulting micropeptides can be determined using by Western blotting and immunofluorescence with anti-FLAG antibodies, thus validating the coding ability of lncRNAs (Anderson et al., 2015; Wang Y et al., 2020).

Determining the endogenous expression of micropeptides is important to infer whether micropeptides play a regulatory role in the organism. The verification of micropeptide endogenous expression can be performed using the following techniques: 1) designing polyclonal antibodies based on the micropeptide, and further confirmation of micropeptide production using western blotting on target fresh tissues or cells; 2) using MS analysis to obtain the fingerprint of the target micropeptide, which can be then discovered by comparison; 3) blocking cell translation using actinomycin (CHX) or antimicropeptide antisense oligonucleotides (OMA), followed by detection of micropeptide expression over time (Walther and Mann, 2010; Li et al., 2017; Guo et al., 2020; Li et al., 2021). In addition, several micropeptides encoded by lncRNAs have been described to be associated with intracellular membrane structures (Pirkmajer et al., 2017; Pang et al., 2020). To determine whether

micropeptides are associated with cell membrane structures, in addition to the bioinformatics analysis discussed above, experimental validation is further necessary, which may include the following: 1) extraction of membrane and cytoplasmic proteins from cells followed by western blotting detection using polyclonal antibodies targeting the micropeptide; 2) imaging flow cytometry techniques (Han et al., 2016; Mikami et al., 2020; Pang et al., 2020). In addition, it has been speculated that micropeptides can act as components of structural proteins and signaling molecules, which require further demonstration.

Previous studies have revealed that lncRNAs associated with ribosomes do not necessarily encode micropeptides; furthermore, if they are coding lncRNAs, encoded micropeptides might still lack functionality. In addition, certain lncRNAs exert their regulatory effects directly rather than through their encoded micropeptides (Gaertner et al., 2020). Therefore, it is necessary to verify whether lncRNAs are inherently functional or only through their encoded micropeptides. It has also been found in earlier studies that, although most micropeptides encoded by lncRNAs may be nonfunctional and highly unstable, about 9% of lncRNA-encoded peptides are conserved in the ORFs of mice transcripts (Ji et al., 2015). Therefore, functional validation of micropeptides encoded by lncRNAs is required to confirm their functionality. Special vectors of lncRNA (knockdown or overexpression) can be designed to transfect cells to enable the impact of introduced vectors in cell fate. In addition, rescue experiments can be conducted to verify whether the lncRNA itself or the encoded micropeptide is responsible for the regulation. After demonstrating the function of the encoded micropeptide, mice models can be used to validate micropeptide activity and



**FIGURE 1** | Schematic illustration of the workflow for bioinformatics prediction and experimental analysis of lncRNA-encoded micropeptides. **(A)** Bioinformatics prediction: firstly, construct a database of putative lncRNA-encoding micropeptides by applying the results of omics sequencing, and search the putative lncRNA sequences with coding potential through NCBI or NONCODE database; secondly, use calculation tools, and databases such as CPC2, CNIT, ORF Finder, PyhloCSF, etc. to evaluate the coding potential of the putative lncRNA, and deduce the corresponding sORF, and amino acid sequence; thirdly, the deduced amino acid sequences were put into the Pfam and CDD databases to look for them, and if they matched, the search for the putative micropeptide information was continued through the UniProt database; finally, the characteristics and structure of the putative micropeptide were predicted and modeled through calculation tools and databases such as SignalP-5.0, TMHMM, ProtScale and SWISS-MODEL; **(B)** Laboratory identification: design a series of special vectors to be transfected into specific cells, and apply western blot and immunofluorescence experiments to identify micropeptides; meanwhile, polyclonal antibodies to this micropeptide were designed, and detected by western blot and LC-MS/MS experiments on sample cells and tissues. Based on the results of both experimental procedures, the putative micropeptide was identified as a novel micropeptide, and then the function and mechanism of the micropeptide were investigated.

regulatory effect *in vivo* (Zhu et al., 2020). These newly discovered functions of lncRNA-encoded micropeptides have greatly enriched the current understanding of lncRNAs. However, due to technological challenges and difficulties in synthesizing polyclonal antibodies for micropeptides, there are still relatively few studies in this field, being thus necessary to explore further. A suggested workflow for studying lncRNA-encoded micropeptides is shown in **Figure 1**.

After verifying that lncRNA-encoded micropeptides are functional micropeptides, the potential regulatory mechanisms behind these micropeptides have become a

pressing issue for subsequent research. CO-IP and MS analysis were applied to find proteins interacting with the micropeptides (Li et al., 2021); RNA-Seq of cells knocked down for micropeptides to look for differential genes and associated signalling pathways (Pang et al., 2020); the JASPAP (the open-access database of transcription factor binding profiles) was used to find the transcription factor that binds to the micropeptide, and dual-luciferase reporter gene vector and chromosomal immunoprecipitation (CHIP) assay were designed to verify the transcription factor that binds to the micropeptide (Castro-Mondragon et al., 2022).

## 4 POTENTIAL REGULATORY ROLES OF LNCRNA-ENCODED MICROPEPTIDES

With the increasing knowledge of lncRNAs encoding micropeptides, the potential regulatory mechanisms of these molecules have also been receiving increasing attention. This suggests that certain mechanisms believed to be regulated by lncRNAs might not be related to an inherent function of lncRNAs but to the micropeptides they encode. This new piece of evidence may override previous knowledge about lncRNAs, suggesting that this phenomenon should be more carefully explored to enable the discovery of appropriate regulatory factors. This will also provide more reliable information for disease and cancer treatment as well as for improving plant and animal productivity.

In 2014, Slavoff et al. (2014) identified the sORF-encoded micropeptide SEP in humans which was shown to stimulate DNA double-strand-break junctions by non-homologous end joining and be involved in DNA repair. In addition, the bifunctional gene lncRNA-Six1, located 432 bp upstream of the gene encoding the protein six homology frame 1 (Six1), was shown to cis-regulate the *Six1* gene encoding the protein; the micropeptide encoded by this lncRNA was also shown to activate the *Six1* gene, which has been shown to be associated with DNA repair (Cai et al., 2017). This indicates that lncRNA-encoded micropeptides might be involved in gene expression and DNA repair processes. Another micropeptide (namely NoBody) encoded in humans in the LINC01420/LOC550643 sORF has been shown to be involved in mRNA turnover and NMD by interacting with mRNA decapping proteins to remove the 5' cap of mRNA to promote 5' to 3' decay (D'Lima et al., 2017). Moreover, NoBody was localized in mRNA decay-associated RNA-protein granules, namely P-bodies. In addition, NoBody levels were shown to be negatively correlated with the number of cellular P-bodies and alter the steady-state levels of cellular NMD substrates (D'Lima et al., 2017), which also suggests that lncRNA-encoded micropeptides might be involved in mRNA conversion and NMD. In addition, lncRNA-encoded micropeptides were shown to interact with multiple splicing regulators to influence RNA splicing (Meng et al., 2020).

Furthermore, Pang et al. (2020) identified a conserved peptide, SMIM30, encoded by LINC00998, which activates the downstream MAPK signaling pathway by driving membrane anchoring and phosphorylation of the non-receptor tyrosine kinase SRC/YES1. This reveals a novel regulatory mechanism of lncRNA-encoded peptides related to the activation of signaling pathways. In addition, lncRNA-encoded micropeptides were shown to regulate mRNA stability and expression by interacting with m6A reader-associated proteins (Zhu et al., 2020), which may provide a guidance for future studies. However, whether these transcriptional modifications have regulatory effects on lncRNA-encoded micropeptides remains to be further explored.

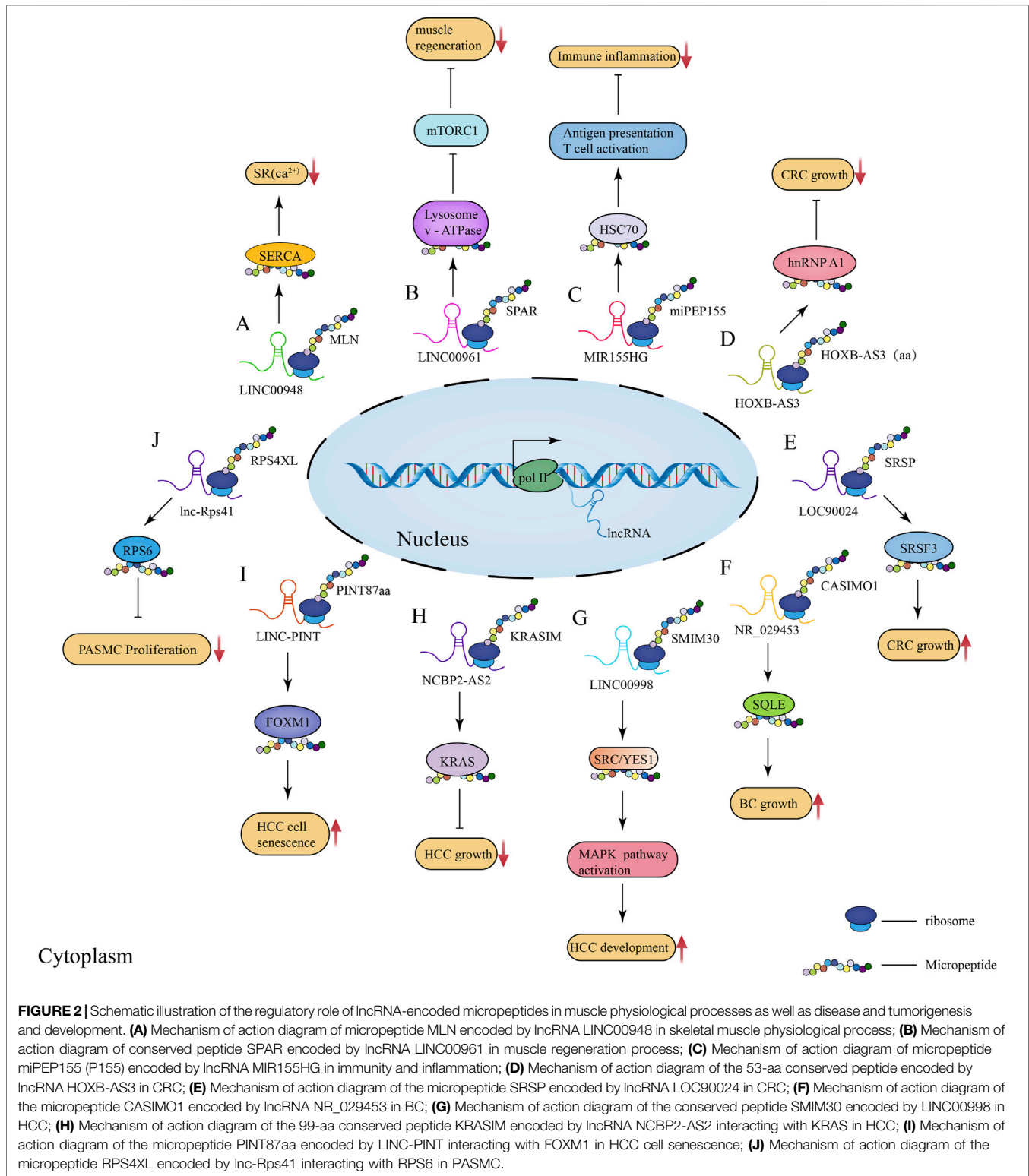
## 5 BIOLOGICAL FUNCTIONS OF LNCRNA-ENCODED MICROPEPTIDES

### 5.1 Micropeptides Associated With Skeletal Muscle Development

Skeletal muscle is the largest and most important constitutive tissue of the human locomotor system, thus playing a crucial role in locomotion and glucolipid metabolism homeostasis (Frontera and Ochala, 2015). In 2013, Magny et al. (2013) identified two peptides shorter than 30 aa in length in *Drosophila* heart tissue, and these peptides were shown to affect muscle homeostasis by regulating calcium transport. This suggests that micropeptides may be important regulators of calcium-dependent signaling in muscle tissue. In 2015, when investigating how micropeptides regulate muscle movement, Anderson et al. (2015) found that myoregulin (MLN), encoded by a skeletal muscle-specific lncRNA, could control muscle relaxation by blocking  $\text{Ca}^{2+}$  uptake into the sarcoplasmic reticulum (SR) and interaction with cardiac SR  $\text{Ca}^{2+}$ -ATPase (SERCA) (Figure 2A). Considering that SERCA plays Figure 2A an important role in the regulation of calcium homeostasis in cardiac myocytes (Anderson et al., 2015), these observations suggest that micropeptides might play an important regulatory role in skeletal muscle physiology. Subsequently, Anderson et al. (2016) further identified two additional regulatory proteins, namely endoregulin (ELN) and another-regulin (ALN) encoded by genes *1110017F19Rik/SMIM6*, and *1810037I17Rik*, which share key amino acid residues with their muscle-specific counterparts and function as direct inhibitors of SERCA pump activity. Additionally, a 34-aa-long micropeptide, DWarf Open Reading Frame (DWORF), encoded by a muscle-specific lncRNA and localized in the SR membrane, was shown to enhance SERCA activity by displacing SERCA inhibitors, phosphoproteins, myosin, and myoregulatory proteins to enhance muscle contraction (Nelson et al., 2016). These findings indicate that micropeptides act as both SERCA inhibitors and activators, thus mediating the regulation of calcium homeostasis in cardiac myocytes, and showing their importance in skeletal muscle physiology.

In addition, micropeptides can also regulate muscle regeneration by interacting with mechanistic target of rapamycin complex 1 (mTORC1). Matsumoto et al. found that SPAR, a conserved peptide encoded by LINC00961, could inhibit mTORC1 activation by interacting with lysosomal v-ATPase (Figure 2B; Matsumoto et al., 2017). Considering that activated mTORC1 promotes muscle regeneration, it can be speculated that SPAR acts as an inhibitor of muscle regeneration. Subsequently, Rion and Ruegg (2017) and Tajbakhsh (2017) further explained the mechanism underlying SPAR-mediated inhibition of mTORC1, further validating the proposed regulating mechanism of muscle regeneration. In addition, it has been proposed that lncRNA-encode micropeptides can regulate skeletal muscle movement by influencing mitochondrial metabolic processes. Makarewich et al. (2018) identified an lncRNA annotated as *1500011K16Rik* and LINC00116 in mouse and human





genomes, respectively, encoding a conserved peptide MOXI that binds to the mitochondrial trifunctional protein at the mitochondrial inner membrane, as well as affects the mitochondrial metabolism and energy homeostasis regulation.

Knockdown of MOXI reduced the ability of cardiac and skeletal muscle mitochondria to metabolize fatty acids and significantly reduced muscle motility (Makarewich et al., 2018). Another LINC00116 found enriched in skeletal muscle and heart was

shown to encode a micropeptide, Mtl<sub>n</sub>, that affects muscle motility by regulating fatty acid oxidation and mitochondrial metabolic processes (Stein et al., 2018). Chugunova et al. (2019) further investigated Mtl<sub>n</sub> and validated the important mechanism of action of this micropeptide in linking respiration and lipid metabolism, as well as its importance in the control of cell fate.

It is known that skeletal muscle development requires fusion of mononuclear progenitor cells to form multinucleated myotubes in a critical but poorly understood process (Hindi et al., 2013). In 2017, Zhang et al. (2017) discovered that the micropeptide Minion (fusion microprotein inducer) encoded by LOC10192972 controls cell fusion and muscle tissue formation by influencing myogenic progenitor cells to form syncytial myotubes. Moreover, it has been shown that Minion-deficient mice died perinatally and exhibited a significant reduction in fused muscle fibers (Zhang et al., 2017). This observation further validates the belief that skeletal muscle development requires the fusion of mononuclear progenitor cells to form multinucleated myotubes. Another micropeptide that has been shown to play a key role in muscle development is LEMP, encoded by the lncRNA MyolncR4, which is highly conserved in vertebrate species (Wang L et al., 2020). LEMP was shown to promote muscle formation and regeneration, and LEMP-deficient mutants had impaired muscle development (Wang Y et al., 2020). Collectively, these findings reveal that lncRNA-encoded micropeptides play an important regulatory role in muscle development, and that certain lncRNAs seemingly lacking coding ability may have been misannotated.

## 5.2 Micropeptides Related to Immune System Inflammatory Response

The latest research findings have revealed that lncRNA-encoded micropeptides play an important role in human innate immunity. In 2018, Jackson et al. (2018) identified a micropeptide encoded by lncRNA Aw112010, which was shown to be essential for the innate immune response *in vivo*, coordinating mucosal immunity under bacterial infections and colitis; moreover, this micropeptide is translated from a non-canonical ORF. Therefore, mis-annotation of genes containing non-canonical ORFs as non-coding RNAs may obscure the role of a large number of previously unidentified protein-coding genes in innate immunity and disease. Another study revealed that lncRNA 1810058I24Rik was downregulated in both human and murine myeloid cells exposed to lipopolysaccharides (LPS), as well as in other Toll-like receptor (TLR) ligands and inflammatory cytokines (Bhatta et al., 2020); this lncRNA encodes a 47-aa-long mitochondrial micropeptide-47 (Mm47) which might be involved in the immune response by activating the Nlrp3 inflammasome to monitor various pathogens and threatening signals (Mangan et al., 2018; Bhatta et al., 2020). Later, Niu et al. (2020) found that the lncRNA MIR155HG encodes the micropeptide miPEP155 (P155) which interacts with the heat shock cognate protein 70 (HSC70) to mediate antigen presentation and T cell initiation as well as suppress autoimmune inflammation (Figure 2C). Collectively, these

findings reveal micropeptides as modulators of antigen presentation and inhibitors of inflammatory diseases, suggesting that micropeptides play an important role in immunity and inflammation, which could offer insights for novel treatments.

## 5.3 Micropeptides Related to Cancer Development

Cancer is a major burden of human diseases. A number of functional micropeptides have been suggested to play a key regulatory role in various human diseases, including cancer, which may constitute a valuable resource for disease and cancer treatments.

Melanoma is among the most dangerous types of skin cancer. Between 2008 and 2013, multiple antigens (e.g., MELOE-1, MELOE-2, and MELOE-3) translated from multiple sORFs of lncRNAs and multiple cis-trans RNAs were found overexpressed in melanoma cells, being also involved in T cell surveillance mechanisms (Godet et al., 2008; Carbonnelle et al., 2013; Charpentier et al., 2016); these could provide optimal T cell targets and therapeutic strategies for melanoma immunotherapy. Interestingly, Huang et al. (2017) found a 53-aa-long conserved peptide encoded by lncRNA HOXB-AS3 in colorectal cancer (CRC) cells, which could inhibit the growth of CRC cells by binding to the heterogeneous nuclear ribonucleoproteins A1 (hnRNP A1) to mediate the cancer metabolic reprogramming process (Figure 2D). Meng et al. (2020) described that the micropeptide SRSP encoded by LOC90024 interacts with serine/arginine-rich splicing factor 3 (SRSF3) to promote tumorigenesis and progression in CRC (Figure 2E). Moreover, micropeptides encoded by lncRNAs have been associated with breast cancer (BC). The micropeptide CASIMO1 translated from transcripts misannotated as lncRNA was found overexpressed in hormone receptor-positive breast tumors; when it was silenced, reduced proliferation was observed in a variety of BC cell lines (Polycarpou-Schwarz et al., 2018). Moreover, CASIMO1 was found to interact with BC oncogenic gene squalene epoxidase (SQLE) in the regulation of cellular lipid homeostasis and thus cancer development (Figure 2F; Polycarpou-Schwarz et al., 2018). Other lncRNA-encoded micropeptides were also found to play a key regulatory role in BC, such as lncRNA EPR-encoded micropeptide (Rossi et al., 2019), LINC00665-encoded micropeptide CIP2A-BP (Guo et al., 2020), and LINC00908-encoded 60-aa-long micropeptide ASRPS (Wang L et al., 2020). The discovery of these key micropeptides provides valuable information on potential therapeutic targets for the treatment of BC as well as clinical research.

Recently, Pang et al. (2020) described that LINC00998 encodes the conserved peptide SMIM30 which promotes hepatocellular carcinoma (HCC) tumorigenesis by regulating cell proliferation and migration (Figure 2G). In this study, a new mechanism of HCC tumorigenesis promoted by the micropeptide has been proposed, which could potentially be used as a new target for HCC therapy as well as a biomarker for HCC diagnosis and prognosis. Xu et al. (2020) identified a 99-aa-long conserved micropeptide, KRASIM, encoded by lncRNA NCBP2-AS2, which

was shown to inhibit HCC oncogenic signals, cancer cell growth and proliferation (**Figure 2H**). These results demonstrate a novel micropeptide inhibitor and provides new insights into the regulatory mechanisms of oncogenic signaling and HCC therapy. Moreover, when exploring the mechanisms of micropeptide function in HCC cell senescence, Xiang et al. (2021) found that the micropeptide PINT87aa, encoded by LINC-PINT, could function as a biomarker and a key regulator of HCC cell senescence, being thus considered a potential therapeutic target for HCC (**Figure 2I**). In addition, it has been demonstrated that the second exon in LINC-PINT RNA can self-loop to form a circular molecule (circPINT) which encodes micropeptides and was involved in the inhibition of glioblastoma cell proliferation (Zhang et al., 2018). These interesting findings reveal that lncRNAs can self-loop and still regulate cancer progression by encoding micropeptides after self-looping, which may provide new insights for cancer and disease treatments. More recently, Cai et al. (2021) identified a micropeptide encoded by lncRNA that is abundantly present in extracellular vesicles (EVs) of glioma cancer cells, which may suggest that EVs-mediated micropeptide transfer represents a novel mechanism of intercellular communication that could potentially be applied in the diagnosis of glioma. In addition, it has been suggested that lncRNAs can encode micropeptides that form oligomers that interfere with water or ion regulation, and abnormalities in water and ion channels play an important role in cancer cell proliferation, migration, apoptosis, and differentiation (Cao et al., 2021). For instance, Cao et al. (2021) found that lncRNA DLEU1 encoding a small transmembrane peptide in glioma cells forms a pentameric channel that acts as a water channel in these cells. Furthermore, lncRNA-encoded micropeptides play an important role in other types of cancers, such as lung cancer (Lu et al., 2019) and esophageal squamous cell carcinoma (Wu P et al., 2020). Collectively, the role of lncRNA-encoded micropeptides in cancer is still poorly understood, and many regulatory mechanisms have not yet been described. Current studies have revealed that micropeptides encoded by lncRNAs, which were previously misannotated as non-coding RNAs, play an important role in cancer development and progression. However, the functions of these functional micropeptides in tumorigenesis are still poorly understood due to the limitations of current available technology for the study of lncRNAs and deserve further investigation. Moreover, the discovery of these functional micropeptides may represent a novel strategy for clinical treatment and prognosis of cancer.

## 5.4 Other Diseases

Pulmonary hypertension (PH) is a rare and fatal disease. An important pathological process in PH is related to the proliferation of pulmonary artery smooth muscle cells (PASMCs) caused by hypoxia (Hu et al., 2021). In a previous study, it was found that lnc-Rps41 with high coding capacity mediates the proliferation of PASMCs under hypoxic conditions (Liu et al., 2020); its encoded micropeptide, RPS4XL, was shown to inhibit PASMCs proliferation and reduce PH death induced by PASMCs proliferation, which

could provide a potential target for early diagnosis of PH (**Figure 2J**; Li et al., 2021).

Myocardial infarction is a severe disease in which an acute blockage of the coronary artery occurs, causing ischemic necrosis of part of the myocardium (Piamsiri et al., 2021). Spencer et al. identified that the micropeptide SPAAR encoded by LINC00961 plays an important role in angiogenesis (Spencer et al., 2020). In addition, loss of the LINC00961/SPAAR locus was found to affect development, myocardial dynamics, and myocardial infarction cardiac response in mice (Spiroski et al., 2021), which suggests that LINC00961/SPAAR contributes to growth and development as well as basal cardiovascular function in adulthood, thus mitigating the risk of myocardial infarction. Therefore, these observations may provide a novel scientific basis and strategy for clinical treatment of cardiovascular diseases.

## 6 CONCLUSION AND FUTURE PERSPECTIVES

Current research on micropeptides encoded by lncRNAs has been received increasing attention. Many computational tools, software, and databases for assessing and predicting lncRNA coding potential have been developed. Moreover, several servers for micropeptide information, and structure prediction are available, which contributes to the study of micropeptides in a more systematic and simplified manner, thus provides a solid foundation for micropeptide research. Moreover, the combined analysis of data obtained by omics techniques (transcriptomics, translomics, proteomics) constitutes a more comprehensive strategy to the analysis of processes in biological systems and to explain the complexity and the overall nature of such processes. Therefore, the progress of the field of lncRNA-encoded micropeptide research chiefly relies on establishing more systematic investigation and robust analytical tools.

Micropeptides encoded by lncRNA may be the missing part in several molecular regulatory mechanisms. Most micropeptides can regulate biological processes independently from lncRNAs and play important roles in the organism. In addition, many lncRNAs were shown to influence several disease-causing and life-sustaining processes in plants and animals; however, it remains to be elucidated whether the function of lncRNAs is related to a certain aspect of their nature or to the micropeptides they encode. Moreover, annotations in current databases of lncRNA-encoded micropeptides are available only for a few species, including human, mouse, rat, zebrafish, fly, yeast, *Caenorhabditis elegans*, *Escherichia coli*, and others. There is still a large number of species for which the lncRNA coding potential has not yet been annotated. This requires further exploration, as to enrich species database information, thus laying a solid foundation for future research.

In addition, although many lncRNAs with coding potential have been characterized, screening methods of functional micropeptides are still controversial. Considering that micropeptide screening criteria are strict and annotation is mainly based on phylogenetic conservatism analysis, a large number of non-standard translated micropeptides might have gone unnoticed, thus limiting the development of micropeptide-

based application. A more in-depth study of lncRNAs and their encoded micropeptides will significantly expand the progress of research in the life sciences and provide new insights and strategies into solving the most urgent problems of the field.

## AUTHOR CONTRIBUTIONS

Original manuscript writing and graph drawing: JP; revised edition English polishing revision, as well as manuscript content expansion and revision: RW; manuscript revision and review: FS, RM, and YR; funding, review and editing of manuscripts: YZ.

## FUNDING

The reported work was supported by the National Natural Science Foundation of China (31860627), Major science and

technology projects of Inner Mongolia Autonomous Region (2021ZD0012).

## ACKNOWLEDGMENTS

Thanks to RW for his great contribution to the revision of the revised manuscript and for his great help in expanding and enhancing the content of the manuscript. Thanks to YZ of Inner Mongolia Agricultural University for providing constructive suggestions. Thanks to FS, RM, and YR for reading the manuscript critically. Thanks to the National Natural Science Foundation of China (31860627) and Major science and technology projects of Inner Mongolia Autonomous Region (2021ZD0012) for funding. We would like to thank topedit ([www.topeditsci.com](http://www.topeditsci.com)) for its linguistic assistance during the preparation of this manuscript.

## REFERENCES

- Achawanantakun, R., Chen, J., Sun, Y., and Zhang, Y. (2015). lncrna-Id: Long Non-coding Rna Identification Using Balanced Random Forests. *Bioinformatics* 31, btv480–905. doi:10.1093/bioinformatics/btv480
- Almagro Armenteros, J. J., Tsirigos, K. D., Sonderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). Signalp 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* 37, 420–423. doi:10.1038/s41587-019-0036-z
- Anderson, D. M., Anderson, K. M., Chang, C.-L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., et al. (2015). A Micropeptide Encoded by a Putative Long Noncoding Rna Regulates Muscle Performance. *Cell* 160, 595–606. doi:10.1016/j.cell.2015.01.009
- Anderson, D. M., Makarewich, C. A., Anderson, K. M., Shelton, J. M., Bezprozvannaya, S., Bassel-Duby, R., et al. (2016). Widespread Control of Calcium Signaling by a Family of Serca-Inhibiting Micropeptides. *Sci. Signal* 9, ra119. doi:10.1126/scisignal.aaj1460
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* 373, 871–876. doi:10.1126/science.abj8754
- Baggerman, G., Verleyen, P., Clynen, E., Huybrechts, J., Deloof, A., and Schoofs, L. (2004). Peptidomics. *J. Chromatogr. B* 803, 3–16. doi:10.1016/j.jchromb.2003.07.019
- Beermann, J., Piccoli, M.-T., Viereck, J., and Thum, T. (2016). Non-Coding Rnas in Development and Disease: Background, Mechanisms, and Therapeutic Approaches. *Physiol. Rev.* 96, 1297–1325. doi:10.1152/physrev.00041.2015
- Bhatta, A., Atianand, M., Jiang, Z., Crabtree, J., Blin, J., and Fitzgerald, K. A. (2020). A Mitochondrial Micropeptide Is Required for Activation of the Nlrp3 Inflammasome. *J. I.* 204, 428–437. doi:10.4049/jimmunol.1900791
- Cai, B., Li, Z., Ma, M., Wang, Z., Han, P., Abdalla, B. A., et al. (2017). lncrna-Six1 Encodes a Micropeptide to Activate Six1 in Cis and Is Involved in Cell Proliferation and Muscle Growth. *Front. Physiol.* 8, 230. doi:10.3389/fphys.2017.00230
- Cai, T., Zhang, Q., Wu, B., Wang, J., Li, N., Zhang, T., et al. (2021). lncRNA-encoded Microproteins: A New Form of Cargo in Cell Culture-Derived and Circulating Extracellular Vesicles. *J. Extracell. Vesicles* 10, e12123. doi:10.1002/jev2.12123
- Calviello, L., and Ohler, U. (2017). Beyond Read-Counts: Ribo-Seq Data Analysis to Understand the Functions of the Transcriptome. *Trends Genet.* 33, 728–744. doi:10.1016/j.tig.2017.08.003
- Cao, Y., Yang, R., Lee, I., Zhang, W., Sun, J., Meng, X., et al. (2021). Prediction of lncrna-Encoded Small Peptides in Glioma and Oligomer Channel Functional Analysis Using In Silico Approaches. *PLoS ONE* 16, e0248634. doi:10.1371/journal.pone.0248634
- Carbannelle, D., Vignard, V., Sehedic, D., Moreau-Aubry, A., Florenceau, L., Charpentier, M., et al. (2013). The Melanoma Antigens Meloe-1 and Meloe-2 Are Translated from a Bona Fide Polycistronic Mrna Containing Functional Ires Sequences. *PLoS ONE* 8, e75233. doi:10.1371/journal.pone.0075233
- Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., Blanc-Mathieu, R., et al. (2022). Jasp2022: The 9th Release of the Open-Access Database of Transcription Factor Binding Profiles. *Nucleic Acids Res.* 50, D165–D173. doi:10.1093/nar/gkab1113
- Charpentier, M., Croyal, M., Carbannelle, D., Fortun, A., Florenceau, L., Rabu, C., et al. (2016). Ires-Dependent Translation of the Long Non Coding Rna Meloe in Melanoma Cells Produces the Most Immunogenic Meloe Antigens. *Oncotarget* 7, 59704–59713. doi:10.18632/oncotarget.10923
- Charpentier, M., Dupré, E., Fortun, A., Briand, F., Maillason, M., Com, E., et al. (2022). hnRNP-A1 Binds to the IRES of MELOE-1 Antigen to Promote MELOE-1 Translation in Stressed Melanoma Cells. *Mol. Oncol.* 16, 594–606. doi:10.1002/1878-0261.13088
- Chassé, H., Boulben, S., Costache, V., Cormier, P., and Morales, J. (2017). Analysis of Translation Using Polysome Profiling. *Nucleic Acids Res.* 45, gkw907. doi:10.1093/nar/gkw907
- Chen, J., Brunner, A.-D., Cogan, J. Z., Nuñez, J. K., Fields, A. P., Adamson, B., et al. (2020). Pervasive Functional Translation of Noncanonical Human Open Reading Frames. *Science* 367, 1140–1146. doi:10.1126/science.aay0262
- Choi, S.-W., Kim, H.-W., and Nam, J.-W. (2019). The Small Peptide World in Long Noncoding Rnas. *Brief. Bioinform* 20, 1853–1864. doi:10.1093/bib/bby055
- Chugunova, A., Loseva, E., Mazin, P., Mitina, A., Navalayev, T., Bilan, D., et al. (2019). linc00116 Codes for a Mitochondrial Peptide Linking Respiration and Lipid Metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 116, 4940–4945. doi:10.1073/pnas.1809105116
- D’Lima, N. G., Ma, J., Winkler, L., Chu, Q., Loh, K. H., Corpuz, E. O., et al. (2017). A Human Microprotein that Interacts with the Mrna Decapping Complex. *Nat. Chem. Biol.* 13, 174–180. doi:10.1038/nchembio.2249
- Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V., and Durinx, C. (2021). Expasy, the Swiss Bioinformatics Resource Portal, as Designed by its Users. *Nucleic Acids Res.* 49, W216–W227. doi:10.1093/nar/gkab225
- Fabre, B., Combié, J.-P., and Plaza, S. (2021). Recent Advances in Mass Spectrometry-Based Peptidomics Workflows to Identify Short-Open-Reading-Frame-Encoded Peptides and Explore Their Functions. *Curr. Opin. Chem. Biol.* 60, 122–130. doi:10.1016/j.cbpa.2020.12.002
- Fan, X.-N., and Zhang, S.-W. (2015). lncrna-Mfdl: Identification of Human Long Non-coding Rnas by Fusing Multiple Features and Using Deep Learning. *Mol. Biosyst.* 11, 892–897. doi:10.1039/c4mb00650j
- Frontera, W. R., and Ochala, J. (2015). Skeletal Muscle: A Brief Review of Structure and Function. *Calcif. Tissue Int.* 96, 183–195. doi:10.1007/s00223-014-9915-y



- Gaertner, B., van Heesch, S., Schneider-Lunitz, V., Schulz, J. F., Witte, F., Blachut, S., et al. (2020). A Human Esc-Based Screen Identifies a Role for the Translated lncrna Linc00261 in Pancreatic Endocrine Differentiation. *Elife* 9, e58659. doi:10.7554/eLife.58659
- Godet, Y., Moreau-Aubry, A., Guilloux, Y., Vignard, V., Khammari, A., Dreno, B., et al. (2008). Meloe-1 Is a New Antigen Overexpressed in Melanomas and Involved in Adoptive T Cell Transfer Efficiency. *J. Exp. Med.* 205, 2673–2682. doi:10.1084/jem.20081356
- Guo, B., Wu, S., Zhu, X., Zhang, L., Deng, J., Li, F., et al. (2020). Micropeptide CIP 2A-BP Encoded by LINC 00665 Inhibits Triple-Negative Breast Cancer Progression. *EMBO J.* 39, e102190. doi:10.15252/embj.2019102190
- Guo, J.-C., Fang, S.-S., Wu, Y., Zhang, J.-H., Chen, Y., Liu, J., et al. (2019). Cnit: A Fast and Accurate Web Tool for Identifying Protein-Coding and Long Non-coding Transcripts Based on Intrinsic Sequence Composition. *Nucleic Acids Res.* 47, W516–W522. doi:10.1093/nar/gkz400
- Han, Y., Gu, Y., Zhang, A. C., and Lo, Y.-H. (2016). Review: Imaging Technologies for Flow Cytometry. *Lab. Chip* 16, 4639–4647. doi:10.1039/c6lc10163f
- Hanada, K., Akiyama, K., Sakurai, T., Toyoda, T., Shinozaki, K., and Shiu, S.-H. (2010). SorfFinder: A Program Package to Identify Small Open Reading Frames with High Coding Potential. *Bioinformatics* 26, 399–400. doi:10.1093/bioinformatics/btp688
- Hanson, P. J., Zhang, H. M., Hemida, M. G., Ye, X., Qiu, Y., and Yang, D. (2012). Ires-Dependent Translational Control during Virus-Induced Endoplasmic Reticulum Stress and Apoptosis. *Front. Microbio.* 3, 92. doi:10.3389/fmicb.2012.00092
- Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P., and Heintz, N. (2014). Cell Type-specific Mrna Purification by Translating Ribosome Affinity Purification (Trap). *Nat. Protoc.* 9, 1282–1291. doi:10.1038/nprot.2014.085
- Hindi, S. M., Tajrishi, M. M., and Kumar, A. (2013). Signaling Mechanisms in Mammalian Myoblast Fusion. *Sci. Signal.* 6, re2. doi:10.1126/scisignal.2003832
- Hu, L., Wang, J., Huang, H., Yu, Y., Ding, J., Yu, Y., et al. (2021). Ythdf1 Regulates Pulmonary Hypertension through Translational Control of Maged1. *Am. J. Respir. Crit. Care Med.* 203, 1158–1172. doi:10.1164/rccm.202009-3419OC
- Hu, L., Xu, Z., Hu, B., and Lu, Z. J. (2017). Come: A Robust Coding Potential Calculation Tool for lncrna Identification and Characterization Based on Multiple Features. *Nucleic Acids Res.* 45, e2. doi:10.1093/nar/gkw798
- Huang, J.-Z., Chen, M., ChenGao, D., Gao, X.-C., Zhu, S., Huang, H., et al. (2017). A Peptide Encoded by a Putative lncrna Hoxb-As3 Suppresses Colon Cancer Growth. *Mol. Cell* 68, 171–184. doi:10.1016/j.molcel.2017.09.015
- Huang, Y., Wang, J., Zhao, Y., Wang, H., Liu, T., Li, Y., et al. (2021). Cncrnadb: A Manually Curated Resource of Experimentally Supported Rnas with Both Protein-Coding and Noncoding Function. *Nucleic Acids Res.* 49, D65–D70. doi:10.1093/nar/gkaa791
- Inada, T., Winstall, E., Tarun, S. Z., Jr., Yates, J. R., 3rd, Schieltz, D., and Sachs, A. B. (2002). One-Step Affinity Purification of the Yeast Ribosome and its Associated Proteins and Mrnas. *RNA* 8, 948–958. doi:10.1017/s1355838202026018
- Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., and Weissman, J. S. (2012). The Ribosome Profiling Strategy for Monitoring Translation *In Vivo* by Deep Sequencing of Ribosome-Protected Mrna Fragments. *Nat. Protoc.* 7, 1534–1550. doi:10.1038/nprot.2012.086
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., and Weissman, J. S. (2009). Genome-Wide Analysis *In Vivo* of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* 324, 218–223. doi:10.1126/science.1168978
- Ingolia, N. T., Hussmann, J. A., and Weissman, J. S. (2019). Ribosome Profiling: Global Views of Translation. *Cold Spring Harb. Perspect. Biol.* 11, a032698. doi:10.1101/cshperspect.a032698
- Ingolia, N. T. (2016). Ribosome Footprint Profiling of Translation throughout the Genome. *Cell* 165, 22–33. doi:10.1016/j.cell.2016.02.066
- Jackson, R., Kroehling, L., Khitun, A., Bailis, W., Jarret, A., York, A. G., et al. (2018). The Translation of Non-canonical Open Reading Frames Controls Mucosal Immunity. *Nature* 564, 434–438. doi:10.1038/s41586-018-0794-7
- Ji, Z., Song, R., Regev, A., and Struhl, K. (2015). Many lncrnas, 5'utrs, and Pseudogenes Are Translated and Some Are Likely to Express Functional Proteins. *Elife* 4, e08890. doi:10.7554/eLife.08890
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Kang, Y.-J., Yang, D.-C., Kong, L., Hou, M., Meng, Y.-Q., Wei, L., et al. (2017). Cpc2: A Fast and Accurate Coding Potential Calculator Based on Sequence Intrinsic Features. *Nucleic Acids Res.* 45, W12–W16. doi:10.1093/nar/gkx428
- Khalili-Tanha, G., and Moghbeli, M. (2021). Long Non-coding Rnas as the Critical Regulators of Doxorubicin Resistance in Tumor Cells. *Cell. Mol. Biol. Lett.* 26, 39. doi:10.1186/s11658-021-00282-9
- King, H. A., Cobbold, L. C., and Willis, A. E. (2010). The Role of Ires Trans-acting Factors in Regulating Translation Initiation. *Biochem. Soc. Trans.* 38, 1581–1586. doi:10.1042/BST0381581
- Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., et al. (2007). Cpc: Assess the Protein-Coding Potential of Transcripts Using Sequence Features and Support Vector Machine. *Nucleic Acids Res.* 35, W345–W349. doi:10.1093/nar/gkm391
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete genomes1Edited by F. Cohen. *J. Mol. Biol.* 305, 567–580. doi:10.1006/jmbi.2000.4315
- Kumar, D., Bansal, G., Narang, A., Basak, T., Abbas, T., and Dash, D. (2016). Integrating Transcriptome and Proteome Profiling: Strategies and Applications. *Proteomics* 16, 2533–2544. doi:10.1002/pmic.201600140
- La Manno, G. (2019). From Single-Cell Rna-Seq to Transcriptional Regulation. *Nat. Biotechnol.* 37, 1421–1422. doi:10.1038/s41587-019-0327-4
- Lauresergues, D., Couzigou, J.-M., Clemente, H. S., Martinez, Y., Dunand, C., Bécard, G., et al. (2015). Primary Transcripts of Micrnas Encode Regulatory Peptides. *Nature* 520, 90–93. doi:10.1038/nature14346
- Lei, M., Zheng, G., Ning, Q., Zheng, J., and Dong, D. (2020). Translation and Functional Roles of Circular Rnas in Human Cancer. *Mol. Cancer* 19, 30. doi:10.1186/s12943-020-1135-7
- Li, A., Zhang, J., and Zhou, Z. (2014). Plek: A Tool for Predicting Long Non-coding Rnas and Messenger Rnas Based on an Improved K-Mer Scheme. *BMC Bioinforma.* 15, 311. doi:10.1186/1471-2105-15-311
- Li, X., Wang, W., and Chen, J. (2017). Recent Progress in Mass Spectrometry Proteomics for Biomedical Research. *Sci. China Life Sci.* 60, 1093–1113. doi:10.1007/s11427-017-9175-2
- Li, Y., Zhang, J., Sun, H., Chen, Y., Li, W., Yu, X., et al. (2021). Lnc-Rps4l-Encoded Peptide Rps4xl Regulates Rps6 Phosphorylation and Inhibits the Proliferation of Pasmcs Caused by Hypoxia. *Mol. Ther.* 29, 1411–1424. doi:10.1016/j.jymthe.2021.01.005
- Lin, M. F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: A Comparative Genomics Method to Distinguish Protein Coding and Non-coding Regions. *Bioinformatics* 27, i275–i282. doi:10.1093/bioinformatics/btr209
- Liu, T., Wu, J., Wu, Y., Hu, W., Fang, Z., Wang, Z., et al. (2022). Lncpep: A Resource of Translational Evidences for lncrnas. *Front. Cell Dev. Biol.* 10, 795084. doi:10.3389/fcell.2022.795084
- Liu, Y., Zhang, H., Li, Y., Yan, L., Du, W., Wang, S., et al. (2020). Long Noncoding Rna Rps4l Mediates the Proliferation of Hypoxic Pulmonary Artery Smooth Muscle Cells. *Hypertension* 76, 1124–1133. doi:10.1161/HYPERTENSIONAHA.120.14644
- Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). Cdd/Sparcle: The Conserved Domain Database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi:10.1093/nar/gkz991
- Lu, S., Zhang, J., Lian, X., Sun, L., Meng, K., Chen, Y., et al. (2019). A Hidden Human Proteome Encoded by 'Non-Coding' Genes. *Nucleic Acids Res.* 47, 8111–8125. doi:10.1093/nar/gkz646
- Luo, X., Huang, Y., Li, H., Luo, Y., Zuo, Z., Ren, J., et al. (2022). Spencer: A Comprehensive Database for Small Peptides Encoded by Noncoding Rnas in Cancer Patients. *Nucleic Acids Res.* 50, D1373–D1381. doi:10.1093/nar/gkab822
- Lv, S., Pan, L., and Wang, G. (2016). Commentary: Primary Transcripts of Micrnas Encode Regulatory Peptides. *Front. Plant Sci.* 7, 1436. doi:10.3389/fpls.2016.01436
- Magny, E. G., Pueyo, J. I., Pearl, F. M. G., Cespedes, M. A., Niven, J. E., Bishop, S. A., et al. (2013). Conserved Regulation of Cardiac Calcium Uptake by Peptides Encoded in Small Open Reading Frames. *Science* 341, 1116–1120. doi:10.1126/science.1238802
- Makarewicz, C. A., Baskin, K. K., Munir, A. Z., Bezprozvannaya, S., Sharma, G., Khemtong, C., et al. (2018). MOXI Is a Mitochondrial Micropeptide that Enhances Fatty Acid  $\beta$ -Oxidation. *Cell Rep.* 23, 3701–3709. doi:10.1016/j.celrep.2018.05.058

- Mangan, M. S. J., Olhava, E. J., Roush, W. R., Seidel, H. M., Glick, G. D., and Latz, E. (2018). Targeting the Nlrp3 Inflammasome in Inflammatory Diseases. *Nat. Rev. Drug Discov.* 17, 588–606. doi:10.1038/nrd.2018.97
- Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., et al. (2017). Mtorc1 and Muscle Regeneration Are Regulated by the Linc00961-Encoded Spar Polypeptide. *Nature* 541, 228–232. doi:10.1038/nature21034
- Meng, N., Chen, M., Chen, D., Chen, X. H., Wang, J. Z., Zhu, S., et al. (2020). Small Protein Hidden in lncrna Loc90024 Promotes "Cancerous" Rna Splicing and Tumorigenesis. *Adv. Sci.* 7, 1903233. doi:10.1002/advs.201903233
- Mikami, H., Kawaguchi, M., Huang, C.-J., Matsumura, H., Sugimura, T., Huang, K., et al. (2020). Virtual-Freezing Fluorescence Imaging Flow Cytometry. *Nat. Commun.* 11, 1162. doi:10.1038/s41467-020-14929-2
- Min, K.-W., Davila, S., Zealy, R. W., Lloyd, L. T., Lee, I. Y., Lee, R., et al. (2017). Eif4e Phosphorylation by Mst1 Reduces Translation of a Subset of Mnas, but Increases Lncrna Translation. *Biochimica Biophysica Acta (BBA) - Gene Regul. Mech.* 1860, 761–772. doi:10.1016/j.bbgrm.2017.05.002
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi:10.1093/nar/gkaa913
- Nam, J. W., Choi, S. W., and You, B. H. (2016). Incredible Rna: Dual Functions of Coding and Noncoding. *Mol. Cells* 39, 367–374. doi:10.14348/molcells.2016.0039
- Navarro Gonzalez, J., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., Raney, B. J., et al. (2021). The Usc Genome Browser Database: 2021 Update. *Nucleic Acids Res.* 49, D1046–D1057. doi:10.1093/nar/gkaa1070
- Nelson, B. R., Makarewich, C. A., Anderson, D. M., Winders, B. R., Troupes, C. D., Wu, F., et al. (2016). A Peptide Encoded by a Transcript Annotated as Long Noncoding Rna Enhances Serca Activity in Muscle. *Science* 351, 271–275. doi:10.1126/science.aad4076
- Nitsche, A., and Stadler, P. F. (2017). Evolutionary Clues in lncRNAs. *WIREs RNA* 8, 1. doi:10.1002/wrna.1376
- Niu, L., Lou, F., Sun, Y., Sun, L., Cai, X., Liu, Z., et al. (2020). A Micropeptide Encoded by lncrna Mir155hg Suppresses Autoimmune Inflammation via Modulating Antigen Presentation. *Sci. Adv.* 6, eaaz2059. doi:10.1126/sciadv.aaz2059
- Orr, M. F., Mao, Y., Storz, G., and Qian, S.-B. (2020). Alternative Orfs and Small Orfs: Shedding Light on the Dark Proteome. *Nucleic Acids Res.* 48, 1029–1042. doi:10.1093/nar/gkz734
- Pang, Y., Liu, Z., Han, H., Wang, B., Li, W., Mao, C., et al. (2020). Peptide Smim30 Promotes Hcc Development by Inducing Src/Yes1 Membrane Anchoring and Mapk Pathway Activation. *J. Hepatology* 73, 1155–1169. doi:10.1016/j.jhep.2020.05.028
- Pauli, A., Norris, M. L., Valen, E., Chew, G.-L., Gagnon, J. A., Zimmerman, S., et al. (2014). Toddler: An Embryonic Signal that Promotes Cell Movement via Apelin Receptors. *Science* 343, 1248636. doi:10.1126/science.1248636
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). Signalp 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nat. Methods* 8, 785–786. doi:10.1038/nmeth.1701
- Piamsiri, C., Manechote, C., Siri-Angkul, N., Chattipakorn, S. C., and Chattipakorn, N. (2021). Targeting Necroptosis as Therapeutic Potential in Chronic Myocardial Infarction. *J. Biomed. Sci.* 28, 25. doi:10.1186/s12929-021-00722-w
- Pirkmajer, S., Kirchner, H., Lundell, L. S., Zelenin, P. V., Zierath, J. R., Makarova, K. S., et al. (2017). Early Vertebrate Origin and Diversification of Small Transmembrane Regulators of Cellular Ion Transport. *J. Physiol.* 595, 4611–4630. doi:10.1113/JP274254
- Polycarpou-Schwarz, M., Gross, M., Mestdagh, P., Schott, J., Grund, S. E., Hildenbrand, C., et al. (2018). The Cancer-Associated Microprotein Casimo1 Controls Cell Proliferation and Interacts with Squalene Epoxidase Modulating Lipid Droplet Formation. *Oncogene* 37, 4750–4768. doi:10.1038/s41388-018-0281-5
- Prasad, A., Sharma, N., and Prasad, M. (2021). Noncoding but Coding: Pri-Mirna into the Action. *Trends Plant Sci.* 26, 204–206. doi:10.1016/j.tplants.2020.12.004
- Pueyo, J. I., Magny, E. G., and Couso, J. P. (2016). New Peptides under the S(Orf) Ace of the Genome. *Trends Biochem. Sci.* 41, 665–678. doi:10.1016/j.tibs.2016.05.003
- Quinn, J. J., and Chang, H. Y. (2016). Unique Features of Long Non-coding Rna Biogenesis and Function. *Nat. Rev. Genet.* 17, 47–62. doi:10.1038/nrg.2015.10
- Rion, N., and Rüegg, M. A. (2017). lncrna-Encoded Peptides: More Than Translational Noise? *Cell Res.* 27, 604–605. doi:10.1038/cr.2017.35
- Ross-Kaschitza, D., and Altmann, M. (2020). Eif4e and Interactors from Unicellular Eukaryotes. *Ijms* 21, 2170. doi:10.3390/ijms21062170
- Rossi, M., Bucci, G., Rizzotto, D., Bordo, D., Marzi, M. J., Puppo, M., et al. (2019). lncRNA EPR Controls Epithelial Proliferation by Coordinating Cdkn1a Transcription and mRNA Decay Response to TGF- $\beta$ . *Nat. Commun.* 10, 1969. doi:10.1038/s41467-019-09754-1
- Ruiz-Orera, J., and Albà, M. M. (2019). Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. *Trends Genet.* 35, 186–198. doi:10.1016/j.tig.2018.12.003
- Ruiz-Orera, J., Messeguer, X., Subirana, J. A., and Alba, M. M. (2014). Long Non-coding Rnas as a Source of New Peptides. *Elife* 3, e03523. doi:10.7554/eLife.03523
- Ruiz-Orera, J., Villanueva-Cañas, J. L., and Albà, M. M. (2020). Evolution of New Proteins from Translated Sorfs in Long Non-coding Rnas. *Exp. Cell Res.* 391, 111940. doi:10.1016/j.yexcr.2020.111940
- Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., et al. (2021). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 49, D10–D17. doi:10.1093/nar/gkaa892
- Shi, Y., Jia, X., and Xu, J. (2020). The New Function of Circrna: Translation. *Clin. Transl. Oncol.* 22, 2162–2169. doi:10.1007/s12094-020-02371-1
- Sieber, P., Platzer, M., and Schuster, S. (2018). The Definition of Open Reading Frame Revisited. *Trends Genet.* 34, 167–170. doi:10.1016/j.tig.2017.12.009
- Slavoff, S. A., Heo, J., Budnik, B. A., Hanakahi, L. A., and Saghatelian, A. (2014). A Human Short Open Reading Frame (Sorf)-Encoded Polypeptide that Stimulates DNA End Joining. *J. Biol. Chem.* 289, 10950–10957. doi:10.1074/jbc.C113.533968
- Spencer, H. L., Sanders, R., Boulberdaa, M., Meloni, M., Cochrane, A., Spiroski, A.-M., et al. (2020). The Linc00961 Transcript and its Encoded Micropeptide, Small Regulatory Polypeptide of Amino Acid Response, Regulate Endothelial Cell Function. *Cardiovasc. Res.* 116, 1981–1994. doi:10.1093/cvr/cvaa008
- Spiroski, A.-M., Sanders, R., Meloni, M., McCracken, I. R., Thomson, A., Brittan, M., et al. (2021). The Influence of the Linc00961/Spar Locus Loss on Murine Development, Myocardial Dynamics, and Cardiac Response to Myocardial Infarction. *Ijms* 22, 969. doi:10.3390/ijms22020969
- Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene Regulation by Long Non-coding Rnas and its Biological Functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi:10.1038/s41580-020-00315-9
- Stein, C. S., Jadiya, P., Zhang, X., McLendon, J. M., Abouassaly, G. M., Witmer, N. H., et al. (2018). Mitoregulin: A lncrna-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. *Cell Rep.* 23, 3710–3720. e8. doi:10.1016/j.celrep.2018.06.002
- Stoneley, M., and Willis, A. E. (2004). Cellular Internal Ribosome Entry Segments: Structures, Trans-acting Factors and Regulation of Gene Expression. *Oncogene* 23, 3200–3207. doi:10.1038/sj.onc.1207551
- Sun, K., Chen, X., Jiang, P., Song, X., Wang, H., and Sun, H. (2013). Iseerna: Identification of Long Intergenic Non-coding Rna Transcripts from Transcriptome Sequencing Data. *BMC Genomics* 14 (Suppl. 2), S7. doi:10.1186/1471-2164-14-S2-S7
- Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing Sequence Intrinsic Composition to Classify Protein-Coding and Long Non-coding Transcripts. *Nucleic Acids Res.* 41, e166. doi:10.1093/nar/gkt646
- Tajbakhsh, S. (2017). lncrna-Encoded Polypeptide Spar(S) with Mtorc1 to Regulate Skeletal Muscle Regeneration. *Cell Stem Cell* 20, 428–430. doi:10.1016/j.stem.2017.03.016
- Tan, L., Cheng, W., Liu, F., Wang, D. O., Wu, L., Cao, N., et al. (2021). Positive Natural Selection of N6-Methyladenosine on the Rnas of Processed Pseudogenes. *Genome Biol.* 22, 180. doi:10.1186/s13059-021-02402-2
- UniProt, C. (2021). Uniprot: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D49. doi:10.1093/nar/gkaa1100
- Vitorino, R., Guedes, S., Amado, F., Santos, M., and Akimitsu, N. (2021). The Role of Micropeptides in Biology. *Cell. Mol. Life Sci.* 78, 3285–3298. doi:10.1007/s00018-020-03740-3

- Volders, P.-J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., et al. (2019). Lncpedia 5: Towards a Reference Set of Human Long Non-coding RNAs. *Nucleic Acids Res.* 47, D135–D139. doi:10.1093/nar/gky1031
- Walther, T. C., and Mann, M. (2010). Mass Spectrometry-Based Proteomics in Cell Biology. *J. Cell Biol.* 190, 491–500. doi:10.1083/jcb.201004052
- Wang, L., Fan, J., Han, L., Qi, H., Wang, Y., Wang, H., et al. (2020). The Micropeptide Lemp Plays an Evolutionarily Conserved Role in Myogenesis. *Cell Death Dis.* 11, 357. doi:10.1038/s41419-020-2570-5
- Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013). Cpat: Coding-Potential Assessment Tool Using an Alignment-free Logistic Regression Model. *Nucleic Acids Res.* 41, e74. doi:10.1093/nar/gkt006
- Wang, T., Cui, Y., Jin, J., Guo, J., Wang, G., Yin, X., et al. (2013). Translating Mrnas Strongly Correlate to Proteins in a Multivariate Manner and Their Translation Ratios Are Phenotype Specific. *Nucleic Acids Res.* 41, 4743–4754. doi:10.1093/nar/gkt178
- Wang, Y., Wu, S., Zhu, X., Zhang, L., Deng, J., Li, F., et al. (2020). Lncrna-Encoded Polypeptide Asrps Inhibits Triple-Negative Breast Cancer Angiogenesis. *J. Exp. Med.* 217, 1. doi:10.1084/jem.20190950
- Washietl, S., Findeiss, S., Müller, S. A., Kalkhof, S., von Bergen, M., Hofacker, I. L., et al. (2011). Rnacode: Robust Discrimination of Coding and Noncoding Regions in Comparative Sequence Data. *RNA* 17, 578–594. doi:10.1261/rna.2536111
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). Swiss-model: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427
- Wery, M., Descrimes, M., Vogt, N., Dallongeville, A.-S., Gautheret, D., and Morillon, A. (2016). Nonsense-Mediated Decay Restricts Lncrna Levels in Yeast unless Blocked by Double-Stranded Rna Structure. *Mol. Cell* 61, 379–392. doi:10.1016/j.molcel.2015.12.020
- Wu, P., Mo, Y., Peng, M., Tang, T., Zhong, Y., Deng, X., et al. (2020). Emerging Role of Tumor-Related Functional Peptides Encoded by Lncrna and Circrna. *Mol. Cancer* 19, 22. doi:10.1186/s12943-020-1147-3
- Wu, S., Zhang, L., Deng, J., Guo, B., Li, F., Wang, Y., et al. (2020). A Novel Micropeptide Encoded by Y-Linked Linc00278 Links Cigarette Smoking and Ar Signaling in Male Esophageal Squamous Cell Carcinoma. *Cancer Res.* 80, 2790–2803. doi:10.1158/0008-5472.CAN-19-3440
- Xiang, X., Fu, Y., Zhao, K., Miao, R., Zhang, X., Ma, X., et al. (2021). Cellular Senescence in Hepatocellular Carcinoma Induced by a Long Non-coding Rna-Encoded Peptide Pint87aa by Blocking Foxm1-Mediated Phb2. *Theranostics* 11, 4929–4944. doi:10.7150/tno.55672
- Xu, W., Deng, B., Lin, P., Liu, C., Li, B., Huang, Q., et al. (2020). Ribosome Profiling Analysis Identified a Kras-Interacting Microprotein that Represses Oncogenic Signaling in Hepatocellular Carcinoma Cells. *Sci. China Life Sci.* 63, 529–542. doi:10.1007/s11427-019-9580-5
- Yan, Y., Tang, R., Li, B., Cheng, L., Ye, S., Yang, T., et al. (2021). The Cardiac Translational Landscape Reveals that Micropeptides Are New Players Involved in Cardiomyocyte Hypertrophy. *Mol. Ther.* 29, 2253–2267. doi:10.1016/j.ymthe.2021.03.004
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-Tasser Suite: Protein Structure and Function Prediction. *Nat. Methods* 12, 7–8. doi:10.1038/nmeth.3213
- Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., et al. (2017). Extensive Translation of Circular RNAs Driven by N6-Methyladenosine. *Cell Res.* 27, 626–641. doi:10.1038/cr.2017.31
- Zhang, M., Zhao, K., Xu, X., Yang, Y., Yan, S., Wei, P., et al. (2018). A Peptide Encoded by Circular Form of Linc-Pint Suppresses Oncogenic Transcriptional Elongation in Glioblastoma. *Nat. Commun.* 9, 4475. doi:10.1038/s41467-018-06862-2
- Zhang, Q., Vashisht, A. A., O'Rourke, J., Corbel, S. Y., Moran, R., Romero, A., et al. (2017). The Microprotein Minion Controls Cell Fusion and Muscle Formation. *Nat. Commun.* 8, 15664. doi:10.1038/ncomms15664
- Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., et al. (2019). Mechanisms and Functions of Long Non-coding RNAs at Multiple Regulatory Levels. *Ijms* 20, 5573. doi:10.3390/ijms20225573
- Zhang, Z., Wu, S., Stenoien, D. L., and Paša-Tolić, L. (2014). High-Throughput Proteomics. *Annu. Rev. Anal. Chem.* 7, 427–454. doi:10.1146/annurev-anchem-071213-020216
- Zhao, J., Qin, B., Nikolay, R., Spahn, C. M. T., and Zhang, G. (2019). Translatomics: The Global View of Translation. *Ijms* 20, 212. doi:10.3390/ijms20010212
- Zhao, J., Wu, J., Xu, T., Yang, Q., He, J., and Song, X. (2018). Iresfinder: Identifying Rna Internal Ribosome Entry Site in Eukaryotic Cell Using Framed K-Mer Features. *J. Genet. Genomics* 45, 403–406. doi:10.1016/j.jgg.2018.07.006
- Zhao, Y., Li, H., Fang, S., Kang, Y., Wu, W., Hao, Y., et al. (2016). Noncode 2016: An Informative and Valuable Data Source of Long Non-coding RNAs. *Nucleic Acids Res.* 44, D203–D208. doi:10.1093/nar/gkv1252
- Zheng, Y., Xu, Q., Liu, M., Hu, H., Xie, Y., Zuo, Z., et al. (2019). Lncar: A Comprehensive Resource for Lncrnas from Cancer Arrays. *Cancer Res.* 79, 2076–2083. doi:10.1158/0008-5472.CAN-18-2169
- Zhu, S., Wang, J.-Z., Chen, D., He, Y.-T., Meng, N., Chen, M., et al. (2020). An Oncopeptide Regulates m6A Recognition by the m6A Reader IGF2BP1 and Tumorigenesis. *Nat. Commun.* 11, 1685. doi:10.1038/s41467-020-15403-9
- Zong, X., Xiao, X., Shen, B., Jiang, Q., Wang, H., Lu, Z., et al. (2021). The N6-methyladenosine RNA-Binding Protein YTHDF1 Modulates the Translation of TRAF6 to Mediate the Intestinal Immune Response. *Nucleic Acids Res.* 49, 5537–5552. doi:10.1093/nar/gkab343

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Pan, Wang, Shang, Ma, Rong and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.