


REVIEW

Open Access



Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA

Pan Wu^{1,2,3}, Yongzhen Mo², Miao Peng², Ting Tang², Yu Zhong², Xiangying Deng², Fang Xiong², Can Guo², Xu Wu^{1,2}, Yong Li⁴, Xiaoling Li², Guiyuan Li^{1,2,3}, Zhaoyang Zeng^{1,2,3} and Wei Xiong^{1,2,3*} 

Abstract

Non-coding RNAs do not encode proteins and regulate various oncological processes. They are also important potential cancer diagnostic and prognostic biomarkers. Bioinformatics and translation omics have begun to elucidate the roles and modes of action of the functional peptides encoded by ncRNA. Here, recent advances in long non-coding RNA (lncRNA) and circular RNA (circRNA)-encoded small peptides are compiled and synthesized. We introduce both the computational and analytical methods used to forecast prospective ncRNAs encoding oncologically functional oligopeptides. We also present numerous specific lncRNA and circRNA-encoded proteins and their cancer-promoting or cancer-inhibiting molecular mechanisms. This information may expedite the discovery, development, and optimization of novel and efficacious cancer diagnostic, therapeutic, and prognostic protein-based tools derived from non-coding RNAs. The role of ncRNA-encoding functional peptides has promising application perspectives and potential challenges in cancer research. The aim of this review is to provide a theoretical basis and relevant references, which may promote the discovery of more functional peptides encoded by ncRNAs, and further develop novel anticancer therapeutic targets, as well as diagnostic and prognostic cancer markers.

Keywords: cancer, circRNA, lncRNA, peptide

Background

The mammalian genome produces tens of thousands of non-coding transcripts during the transcription process. About 98% of the RNA in the human transcriptome is non-coding [1–4]. Non-coding RNA (ncRNA) is transcribed from the genome but not translated into protein. It controls various levels of gene expression during physiological and developmental processes, including epigenetic modification [5], transcription [6], RNA splicing [7], scaffold assembly [8], and others. ncRNAs have tissue-specific expression patterns and are potential biomarkers. Thus, they could serve as clinical diagnostic and

prognostic indicators [9, 10]. Additionally, ncRNAs that were heretofore considered non-coding may, in fact, be able to encode small biologically active peptides [11–13]. Functional peptides are usually encoded by short open reading frames (sORFs) in ncRNAs [14–23]. ncRNAs can have one or more sORFs that can be translated into small peptides < 100 amino acids long. Previously, the traditional gene annotation process filtered out proteins < 100 amino acids by default and treated them as noise or false positives. Thus, they were always ignored [24]. However, as proteomics and translation technology have grown in popularity and increased in precision and accuracy, it was discovered that many ncRNAs are translatable [13, 25–28]. At present, it is recognized that long ncRNAs (lncRNAs) and circular RNAs (circRNAs) contain sORFs that can be translated into functional small peptides.

lncRNAs are generally defined as long RNA transcripts (>200 nucleotides) that do not encode proteins [29, 30]. The number of lncRNAs may exceed that of

* Correspondence: xiongwei@csu.edu.cn

¹NHC Key Laboratory of Carcinogenesis and Hunan Key Laboratory of Translational Radiation Oncology, Hunan Cancer Hospital and The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, Hunan, China

²Key Laboratory of Carcinogenesis and Cancer Invasion of the Chinese Ministry of Education, Cancer Research Institute, Central South University, Changsha, Hunan, China

Full list of author information is available at the end of the article



protein-coding transcripts. lncRNAs participate in the epigenetic regulation of gene expression [31, 32]. Several lncRNAs resemble mRNAs and can be transcribed, spliced, capped, and polyadenylated by RNA polymerase II-like protein-encoding pathways. These lncRNAs have tissue modification profiles, splicing signals, and exon/intron lengths similar to those of mRNAs [33–37]. In most cases, lncRNAs do not biochemically differ from mRNAs except that they lack reading frames encoding proteins. However, mass spectrometry, deep RNA sequencing, and other advanced molecular techniques have revealed that certain lncRNAs have non-random long sORFs [38–40], their exons are more highly conserved than those in protein-coding genes [41], they can interact with ribosomes [42, 43], and they could encode proteins. Mature microRNAs (miRNAs) are produced by the cleavage of primary transcripts (pri-miRNA) via a series of nucleases [44, 45]. Pri-miRNA is a special type of lncRNA hundreds to thousands of nucleotides long that is transcribed by RNA polymerase II-like protein-coding genes [46–49]. Therefore, pri-miRNA may also be able to encode proteins or peptides.

CircRNAs were recently discovered as ncRNAs with covalently closed structures. They regulate disease development and occurrence [50–52]. CircRNA is transcribed by RNA polymerase II without 5'-3' polarity or polyadenine tails. It has the same transcriptional efficiency as linear RNA [53–55]. CircRNA is hundreds to thousands of bases long and mainly consists of exons. Studies have shown that mammalian circRNAs are endogenous, abundant, conserved, and stable [56, 57]. CircRNAs are miRNA sponges [58–60] that control gene transcription. Moreover, the highly conserved ORFs in circRNAs encode functional peptides both in vivo and in vitro in a manner independent of the 5' cap structure, such as internal ribosome entry site (IRES) induction [61], promoting adenosine methylation (N6-methyladenosine; m6A) [62], rolling cycle amplification [63, 64], and others [65, 66]. As circRNA has a unique covalently closed structure, the ORFs therein circulate across the splicing site and even beyond its length. For this reason, it can also produce proteins > 100 amino acids long [67, 68].

The present review discusses current research progress in lncRNA- and circRNA-encoded proteins. It focuses on the fact that certain cancer-related lncRNAs and circRNAs encode functional small peptides that regulate biological processes and influence tumorigenesis, invasion, metastasis, and so on. The review also predicts and identifies potential ncRNAs that can encode functional small peptides.

Prediction of ncRNA coding potential and identification of small peptides

In view of the increasing interest in ncRNA-encoded polypeptides, numerous prediction and experimental

identification methods have been developed to determine the coding ability of ncRNAs. These include reading frame prediction, translation initiation component prediction, conservation analysis, and translation omics and proteomics, and others.

Open reading frame prediction

Open reading frames (ORFs) are nucleic acid sequences starting with ATG (or AUG in RNA) and continuing in three-base sets to a stop codon [69]. The length of sORFs is usually < 300 nt. Calculations [70] and ribosome analyses [43] have disclosed that thousands of unannotated ORFs are translated in various species [14, 16, 18, 71–74]. Longer ORFs are the most likely to be encoded [75–77]. Regulatory elements (IRES, m6A-modified conserved sites, and so on) upstream in the open reading frame mediate translation [78, 79]. The positional relationship between the ORF and the cyclization site is significant in circRNAs. In general, an ORF spanning the splicing site is the distinctive feature of circRNA-encoded peptides [80]. Websites and software used to predict ORFs are listed in Table 1.

Predictions of translation starter elements: IRES

An IRES is an RNA regulatory element that recruits ribosomes, implements ribosomal assembly and reading frame protein translation, and initiates protein translation independent of the 5' cap structure and direct translation [96–98]. An earlier study found that a circRNA constructed in vitro with an IRES recruited ribosomes and underwent translation [61]. An IRES region has been found in a wide range of viral RNAs [99, 100]. The first was detected in the small RNA virus 5' untranslated region (5'-UTR). IRES have also observed in certain eukaryotic mRNAs. About 10% of mRNAs use IRES in the 5'-UTR to recruit ribosomes [101]. Moreover, the UTR of *circ-ZNF609* may serve as an IRES facilitating *circ-ZNF609* translation in a shear-dependent manner [78]. IRES may be difficult to detect in higher eukaryotes as these organisms have highly complex genomes and cellular regulatory networks. IRES appear mainly in the 5'-UTRs upstream of the ORFs they control. However, there are exceptions. Certain IRES may be seen between the ORFs while others reside within them [102–104]. IRES sequences in cells are generally less active and efficient than those in viruses. Nevertheless, the former have good characteristics and are reliable [105, 106]. Endogenous ncRNAs with IRES may translate long polypeptide chains on a continuous ORF [107, 108]. The selective regulation of IRES-mediated translation participates in physiological and pathological processes such as cell growth, proliferation, differentiation, stress response, and apoptosis [98, 109, 110].

Table 1 Classification of prediction methods for ORFs

Name	Characteristics	Website
CircRNADb [81]	It includes 32,914 circRNA records for human exons. Each of them has IRES sequence components, predicted ORFs, related references, and so on. The predicted ORF is usually the cross-splicing site.	http://202.195.183.4:8000/circnadb/circRNADb.php
sORFs.org [82, 83]	It comprises > 4,374,422 sORFs from six different species and derived from multiple ribosome profiling and sequencing datasets.	http://www.sorfs.org
ORF Finder [84]	It performs six-frame translations and returns the range of each ORF and its protein translation. These may be submitted directly for BLAST similarity or COGs database searches.	https://www.ncbi.nlm.nih.gov/orffinder/
ORF Predictor [85]	It was designed to predict expressed sequence tags (EST) or cDNA sequences. Its output file consists of predicted coding DNA sequences, the start and end of the coding region, and the protein peptide sequence.	http://bioinformatics.yzu.edu/tools/OrfPredictor.html
SMS:ORF Finder [86]	It searches for newly-sequenced DNA and returns the range of each ORF and its protein translation. SMS:ORF Finder supports the entire IUPAC alphabet and several genetic codes.	http://www.bioinformatics.org/sms2/orf_find.html
CSCD [87]	It contains cancer-associated circRNA alternative splicing, expression, and translation (ORF). It is linked to UCSC which predicts potential ORFs and highlight translatable circRNAs.	http://gb.whu.edu.cn/CSCD/
PhyloCSF [88–91]	It is based on the phylogenetic analysis of multi-species genomic sequence alignments and identifies conserved protein coding regions. It requires complete ORF sequences from different species to be able to evaluate their coding probabilities.	http://compbio.mit.edu/PhyloCSF/
CircPro [92]	Its data was derived from high-throughput sequencing (RNA-Seq and Ribo-Seq). It generates a list of circRNAs and reports genomic locations, ORF lengths, junction reads from Ribo-Seq, and so on.	http://bis.zju.edu.cn/CircPro/
cORF_pipeline [93, 94]	Its output predicts sORF sequences, start and stop positions, annotation data, and so on. The longest ORF spanning the circRNA splicing site is the one most likely to encode.	https://github.com/kadenerlab/cORF_pipeline
CircBank [95]	Its output contains the ORF size, the coding potential, and circRNA conservation.	http://www.circbank.cn/index.html

Websites currently used to predict IRES are listed in Table 2.

Prediction of m6A modification

The m6A modification is very common in the mRNAs and ncRNAs of higher organisms [115, 116]. The m6A modification regulates mammalian gene expression [117], as well as RNA stability, localization, shearing, and translation at the post-transcriptional level. It was recently discovered that m6A has various effects on translation [118–121]. Abnormalities in its regulatory mechanism are associated with

tumorigenesis [122, 123]. Using ribosome profiling, computational prediction, and mass spectrometry, m6A-driven endogenous ncRNA translation has been found to be widespread [62, 124]. Numerous translatable endogenous circular RNAs probably contain m6A sites. To examine the ability of m6A to drive circRNA translation, an m6A-modified circRNA was constructed in vitro. The m6A reading protein YTHDF3 was tightly bound to the translation initiation factor eIF4G2. The latter promoted circRNA translation in cells [62]. Commonly used tools for screening m6A motifs are listed in Table 3.

Table 2 IRES prediction methods

Name	Characteristics	Website
IRESite [111]	It is based on experimental data derived from 68 viruses and 115 eukaryotic cells. It furnishes information on experimental IRES fragments including their nature, function, origin, size, sequence, structure, relative position to the surrounding protein coding region, and so on.	www.iresite.org
IRESfinder [112]	It is a logit model-based forecasting tool based on 19 k-mer parameters. Its accuracy is ~80%. IRESfinder is a standalone script for Python that is applicable to high-throughput screening.	https://github.com/xiaofengsong/IRESfinder
IRESPred [113]	It predicts viral and cellular IRES via the Support Vector Machine (SVM). This predictive model integrates 35 features based on the sequence and structural properties of UTRs and their probabilities of interacting with small subunit ribosomal proteins (SSRPs). Its accuracy is ~75.75% accuracy and it had a 0.51 Matthews correlation coefficient (MCC) in blind testing.	http://bioinfo.net.in/IRESPred/
VIPS [114]	It consists of the RNAL folding, RNA Align, and pknotsRG programs. Evaluations of the UTR, IRES, and virus databases disclosed that it has superior accuracy and flexibility and can predict four different sets of IRES.	http://140.135.61.250/vips/

Table 3 M6A prediction methods

Name	Characteristics	Website
DeepM6ASeq [125]	It is based on miCLIP-Seq data at single-base resolution and detects m6A sites. It can recognize new reader FMR1.	https://github.com/rreybeyb/DeepM6ASeq
M6APred-EL [126]	It uses position-specific k-mer nucleotide propensity, physicochemical properties, and ring function hydrogen chemical properties to optimize m6A position recognition accuracy.	http://server.malab.cn/M6A-Pred-EL/
M6AMRFS [127]	It uses dinucleotide binary encoding and local position-specific dinucleotide frequencies to encode RNA sequences. It can identify m6A sites in multiple species.	http://server.malab.cn/M6AMRFS/
SRAMP [128]	It identifies mammalian m6A sites at single-nucleotide resolution and builds m6A site predictors. SRAMP = sequence-based RNA adenosine methylation site predictor.	http://www.cuilab.cn/sramp/
iRNA-Methyl [129]	Identifying m6A sites by incorporating the global and long-range sequence pattern information of RNA via the pseudo k-tupler nucleotide composition (PseKNC) approach.	http://lin.uestc.edu.cn/server/iRNA-Methyl
iRNA (m6A)-PseDNC [130]	It uses the Euclidean distance-based method and pseudodinucleotide composition to identify m6A sites in the <i>Saccharomyces cerevisiae</i> (yeast) genome.	http://lin-group.cn/server/iRNA(m6A)-PseDNC.php
m6Acomet [131]	It is based on the RNA co-methylation network comprising 339,158 putative gene ontology functions associated with 1,446 identified human m6A sites.	http://www.xjtlu.edu.cn/biologicalsciences/m6acomet
WHISTLE [132]	It integrates 35 genome-derived and conventional sequence-derived features. It enable direct queries of predicted RNA-methylation sites, their putative functions, and their associations with other methylation sites or genes.	http://whistle-epitranscriptome.com
pRNAm-PC [133]	It predicts m6A sites in RNA sequences based on physicochemical properties. RNA sequence samples are expressed by pseudodinucleotide composition (PseDNC).	http://www.jci-bioinfo.cn/pRNAm-PC
TargetM6A [134]	It identifies m6A sites from RNA sequences via position-specific nucleotide propensities (PSNP) and a support vector machine (SVM).	http://csbio.njust.edu.cn/bioinf/TargetM6A
AthMethPre [135]	It trains the SVM classifier using the positional flanking nucleotide sequence and the position-independent k-mer nucleotide spectrum to predict m6A sites in <i>Arabidopsis thaliana</i> .	http://bioinfo.tsinghua.edu.cn/AthMethPre/index.html
RNAMethPre [136]	It predicts m6A sites by integrating multiple mRNA features and training the SVM classifier in mammalian mRNA sequences.	http://bioinfo.tsinghua.edu.cn/RNAMethPre/index.html

Conservation analysis

Conserved sequences indicate potential functions and/or play important roles in cell development and regulation [137–139]. As a rule, coding region sequences are highly conserved. Evolutionary conservation (including ncRNA sequences, sORFs, and small peptide amino acid sequences) may serve as a predictor in the analysis of the coding functions of ncRNAs [71, 75, 140–142]. For conservation analysis, nucleotide- and protein-protein BLAST [23], UCSC [143, 144], and other websites may be consulted. Alternatively, software such as MegAlign [145], MEGA [146], and Clustal [147, 148] can be used.

Translational omics analysis

Most of the current research on ncRNA-encoded peptides is based on data analyses performed by ribosome display technology [43, 141]. The evolution of high-throughput sequencing has yielded four detection methods for translation omics analysis [149]. These include polysome profiling, ribosome immunoprecipitation/ribosome affinity purification, ribosome profiling (also known as ribo-seq), and ribosome-nascent chain complex (RNC)-seq.

Polysome profiling separates polyribosomes by sucrose density gradient centrifugation as ribosomes have high sedimentation coefficients. The rate of sedimentation during gradient centrifugation increases with the

number of ribosomes bound to the mRNA. Thus, mRNAs bound to different numbers of ribosomes may be separated in solution by centrifugation. mRNAs and their active translation ORFs in the separated components are then analyzed, and the output is used to evaluate the ncRNA coding potential [78]. However, the RNA recovery for translation is low in this method and may not suffice to meet the sample size requirement for full spectrum analysis [150, 151]. Ribosome profiling is a comprehensive quantitative method to sequence the mRNA segments in ribosomes [152]. It uses low-concentration RNase to digest the RNC, degrades the mRNA fragments without ribosome coverage, and sequences and analyzes RNA fragments ~22–30 bp long. These are known as ribosome footprints or ribosome-protected fragments. The ribosome distribution and density on each transcript can be determined, as well as the starting codon, ORF location, translation pause area, and other information [153–156]. The ribosomal characteristics of hundreds of ORFs in annotated non-coding genes, as well as new peptides, may also be identified from ribo-seq data [141, 157–161]. During translation, ribosomes bind and move along the mRNA chain and gradually synthesize a protein polypeptide chain based on the codon triplet information in the mRNA template. During this process, an RNC is formed. Ribosomes and tandem mRNA precipitates may be separated by sucrose

density gradient centrifugation and the mRNA further purified and separated for high-throughput sequencing, known as RNC-seq [162]. Using this method, ncRNA binding to ribosomes may also be analyzed. ncRNA can be translated into proteins in RNCs [79]. In ribosome immunoprecipitation/ribosome affinity purification, specific fusion marker proteins are used to bind ribosomal large subunits, and antibodies against these markers isolate the polymers. The mRNAs and ncRNAs are then isolated for microarray or sequencing analysis [163, 164].

Proteomics analysis

Proteomics can be used to discover and directly detect micropeptides encoded by ncRNAs, which in turn provides the most intuitive evidence that ncRNAs can encode small peptides. Among them, biological mass spectrometry is a common identification and analysis method for these micropeptides. Zhang et al. used immunoprecipitation combined with liquid chromatography tandem mass spectrometry (LC-MS/MS) to characterize the unique amino acid sequences encoded by *circ-FBXW7*, *circ-SHPRH*, and *circPINT*. The distinctive peptides identified in the mass spectrometry results also matched the ORF prediction results [67, 68, 79]. Commonly used software or databases that can be used for protein sequence alignment and peptide search of mass spectrometry data include UniProt [165] and Mascot daemon [166].

Experimental method identification

As more attention has been given to ncRNA-encoded proteins, several experimental identification methods have emerged to detect these proteins. To verify predicted reading frame expression, FLAG-labeled expression vectors constructed in vitro are imported into cells. Western blots identify distinct bands at the expected molecular weight, indicating that the artificially constructed ncRNA with the FLAG label was translated. CRISPR has also been used to knock FLAG labels into endogenous ncRNA coding regions and detect endogenous protein expression. Sucrose density gradient centrifugation, puromycin treatment, and other techniques determine the extent to which the target ncRNA recruits and binds the ribosomes in the translation machinery. Dual luciferase and other reporting assays elucidate IRES activity and predict ncRNA encoding ability [78, 167]. Overexpression and mutation experiments demonstrate the functions of each regulatory sequence and site. Use of vectors with manipulation of translational elements, such as mutated forms of the predicted IRES, m6A modification sites, or ATG start codons, may confirm whether the translation occurs as normal and the phenotype is consistent. Endogenous translation products may be identified by western blot or with specific antibodies

such as those designed for unique amino acid sequences across the circRNA splicing site [67, 68] or common amino acid sequences encoded by lncRNA- and circRNA-derived transcripts [168]. In this way, the translational functions of the endogenous circRNAs may be verified, and the overexpression and knockdown of the translation products can be simulated. The proteins and polypeptides in the samples are isolated and determined by LC-MS/MS.

Tumor-related functional peptides

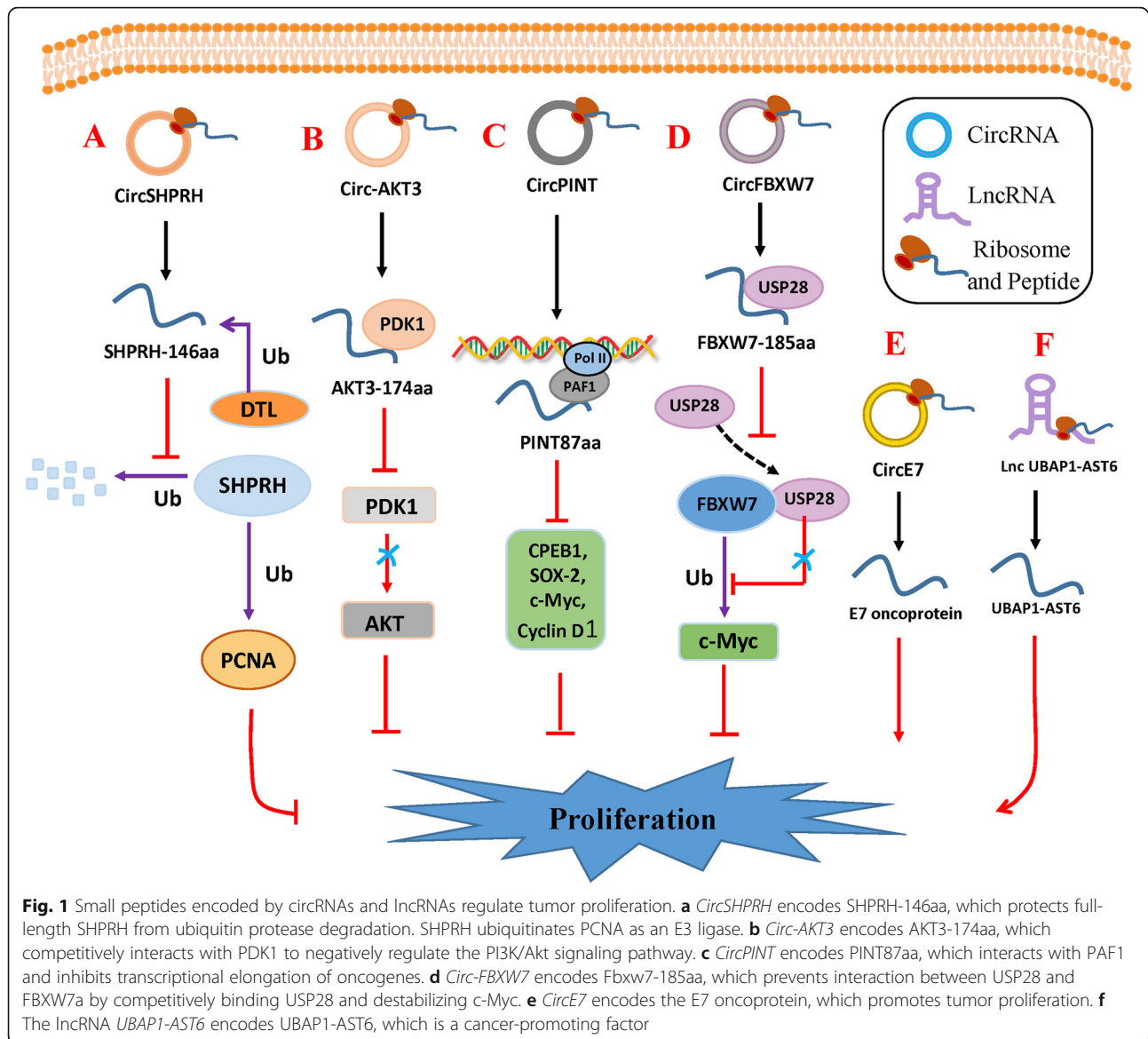
Research on ncRNA-encoding proteins has been increasing in recent years. Multiple ncRNAs encode small peptides and regulate various malignant tumor phenotypes, such as cell proliferation, invasion, and metastasis. Below are certain tumor-related functional peptides known to be encoded by circRNAs and lncRNAs.

SHPRH-146aa

The circular form of the SNF2 histone linker PHD RING helicase (*SHPRH*) gene encodes the protein SHPRH-146aa. *Circ-SHPRH* and SHPRH-146aa are highly expressed in normal human brain tissue and downregulated in glioblastoma. Cyclization in *circ-SHPRH* results in the tandem stop codon UGAUGA. The entire *circ-SHPRH* is translated into a 146-aa protein by starting and stopping translation with overlapping genetic codes. An antibody against the unique amino acid sequence generated by the ORF spanning the splicing site and identification of the SHPRH-146aa amino acid sequence by LC-MS/MS confirmed that *circ-SHPRH* was translated into SHPRH-146aa. The latter participates in the development of central nervous system cancer through regulation of protein ubiquitination pathways. SHPRH-146aa overexpression in U251 and U373 glioblastoma cells reduces their malignancy and tumorigenicity in vitro and in vivo. SHPRH-146aa protects full-length SHPRH from degradation by ubiquitin proteases. It also stabilizes SHPRH as an E3 ligase by ubiquitinating proliferating cell nuclear antigen. In this manner, it inhibits cell proliferation and tumorigenicity [68, 169] (Fig. 1a).

AKT3-174aa

Circ-AKT3 is formed by the cyclization of the third to seventh exons of *AKT3*. It is 524-nt long and localized mainly to the cytoplasm. When it is driven by an active IRES, *circ-AKT3* encodes a 174-aa protein, AKT3-174aa, via the overlapping start-stop codon UAAUGA. AKT3-174aa has the same amino acid sequence as residues 62–232 of *AKT3*. Compared with normal brain tissue, AKT3-174aa is downregulated in glioblastoma tissue. AKT3-174aa, but not *circ-AKT3*, acts as a tumor suppressor. AKT3-174aa overexpression inhibits glioblastoma cell proliferation, radiation resistance, and tumorigenicity. The PI3K/Akt



pathway plays central roles in various oncogenic signaling pathways promoting glioblastoma development and progression [170, 171]. After PI3K activation, Akt is recruited to the membrane via the PH-domain and is fully activated after Thr308 and Ser473 are sequentially phosphorylated. PDK1 directly phosphorylates Akt at Thr308. This initial step is the most critical in Akt activation. The amino acid sequence of AKT3-174aa is partially identical to that of AKT3. Thus, AKT3-174aa competitively interacts with activated PDK1, inhibits Akt phosphorylation at Thr308, and negatively regulates the PI3K/Akt signaling pathway [172] (Fig. 1b).

PINT87aa

Zhang et al. identified certain circRNAs by performing circRNA transcriptome and RNC-RNA sequencing and

bioinformatics integration analysis on normal human astrocytes and U251 glioblastoma cells. The second exon of the lncRNA *LINC-PINT* formed the circular molecule *circPINT* by self-cyclization. The latter contained an sORF and a natural IRES encoding an 87-aa polypeptide translated from endogenous *circPINT* exon 2 rather than linear *LINC-PINT*, termed PINT87aa. It is localized mainly to the nucleus, directly interacts with PAF1, regulates the PAF1/POLII complex, inhibits the transcriptional elongation of the downstream oncogenes *cpeb1*, *sox-2*, *c-Myc*, *cyclin D1*, and others, and inhibits the proliferation and tumorigenesis of glioblastoma and other cancer cell types [79] (Fig. 1c).

FBXW7-185aa

Circ-FBXW7 may have an ORF spanning the splice site. It is highly conserved among different species and

encodes a 185-aa protein driven by an IRES independently of the 5' cap translational machinery. *Circ-FBXW7* was able to be translated in human cells using a construct harboring a FLAG sequence before the ORF stop codon. The *circ-FBXW7* IRES-mut vector, which has a mutation in the IRES sequence, was transfected into U251 and U373 cells. However, cells transfected with the vector formed a circular RNA similar to *circ-FBXW7*. Therefore, it is FBXW7-185aa rather than *circ-FBXW7* that induces cell cycle arrest and hinders glioma cell proliferation. FBXW7a is the most abundant isoform of FBXW7. It uses c-Myc as a tumorigenesis regulator for ubiquitination-induced degradation. The de-ubiquitinating enzyme USP28 stabilizes c-Myc by binding it via interaction with the N-terminus of FBXW7a. The protein FBXW7-185aa translated by *circ-FBXW7* has a relatively higher affinity for USP28. It functions as bait and competitively inhibits USP28 from binding FBXW7a. In this way, it perturbs c-Myc stabilization induced by USP28 and shortens its half-life. FBXW7-185aa is a synergistic parental gene that encodes FBXW7a, stabilizes c-Myc, inhibits tumor cell proliferation and malignant phenotypes, and impedes malignant glioma progression [67] (Fig. 1d).

E7 protein

Human papillomaviruses produce the 472-nt oncogene called *CircE7* containing the entire E7 ORF. *CircE7* is modified by m6A and is localized mainly to the cytoplasm. It is closely associated with polyribosomes and may be translated to the E7 oncoprotein. The latter process is upregulated by cellular stressors such as heat shock. E7 translation was increased two- to four-fold under 42 °C heat shock. *CircE7* knockdown in CaSki cervical cancer cells reduces E7 protein levels, inhibits cancer cell proliferation and colony formation, and suppresses tumor growth and malignancy. *CircE7* is essential for E7 protein expression and transformation in CaSki cervical cancer cells both in vitro and in transplanted tumors [124] (Fig. 1e).

UBAP1-AST6

The protein UBAP1-AST6 is translated from a lncRNA, is localized mainly to the nucleus, and is expressed in A549 lung cancer cells. UBAP1-AST6 promotes cancer, and its overexpression significantly induces cancer cell proliferation and colony formation [173] (Fig. 1f).

circPPP1R12A-73aa

CircPPP1R12A is highly expressed in colon cancer tissues and serves as a prognostic marker of survival. Patients with increased *circPPP1R12A* have comparatively poorer overall survival. *CircPPP1R12A* contains a short 216-nt ORF encoding the conserved 73-aa peptide *circPPP1R12A-73aa*. Silencing *CircPPP1R12A* markedly

inhibits colon cancer cell proliferation, migration, and invasion. Construction of the FLAG-*circPPP1R12A* overexpression vector with an initiation codon mutation ATG/ACG confirmed that it is *circPPP1R12A-73aa* rather than *circPPP1R12A* that plays key roles in colon cancer cell proliferation, invasion, and metastasis. The YAP1-specific inhibitor peptide 17 dramatically attenuated colon cancer cell proliferation, migration, and invasion promoted by *circPPP1R12A-73aa* overexpression. Induction of colon cancer growth and metastasis by *circPPP1R12A-73aa* was validated in vitro and in vivo by activating the Hippo-YAP signaling pathway [174] (Fig. 2a).

HOXB-AS3 peptide

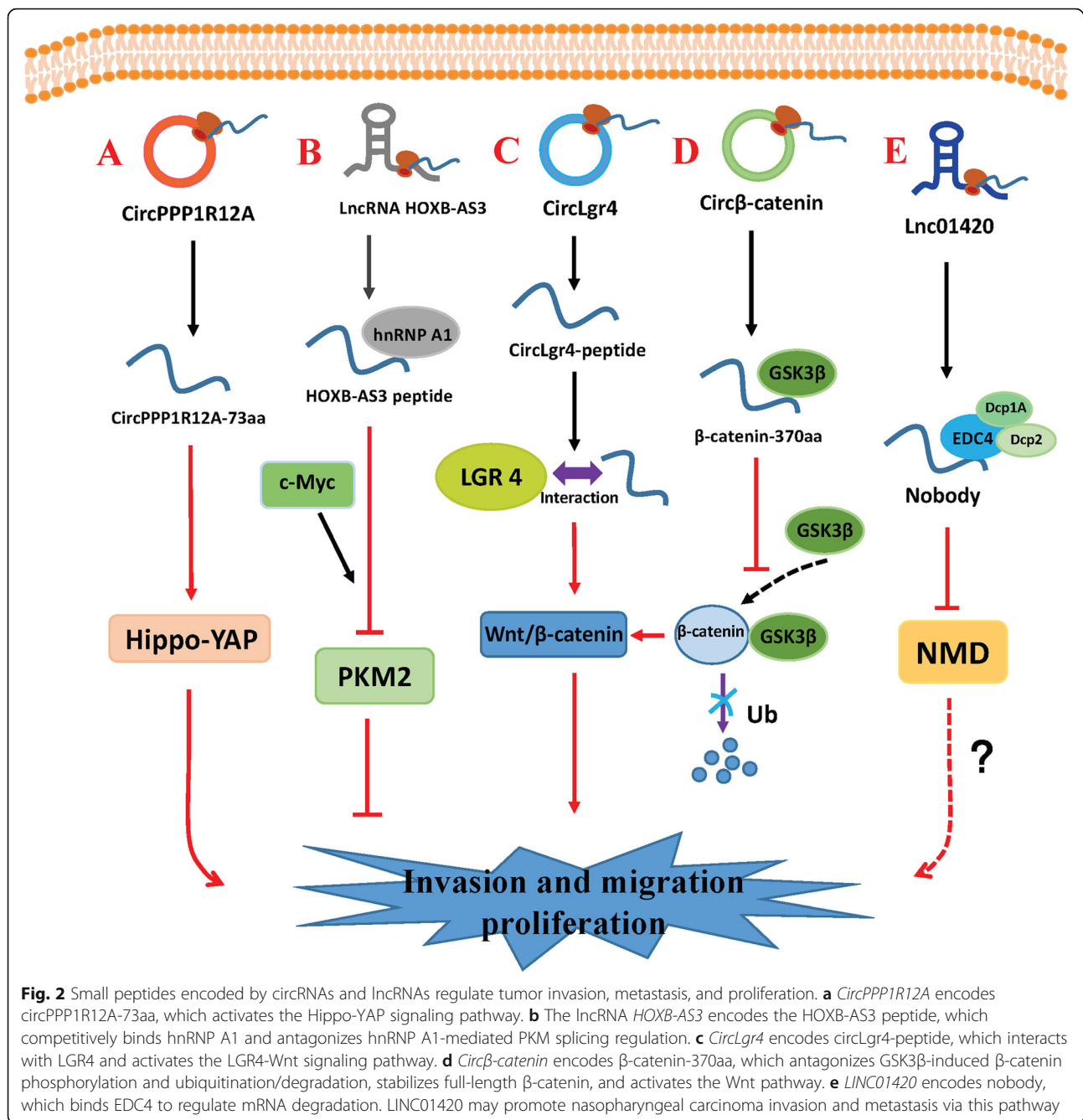
The lncRNA *HOXB-AS3* is a tumor suppressor that is substantially downregulated in highly metastatic and primary colorectal cancer (CRC) tissues. *HOXB-AS3* binds ribosomes and encodes a highly conserved 53-aa peptide called HOXB-AS3. It is endogenous, naturally occurring, and widely expressed in various tumor tissues. HOXB-AS3 inhibits cancer cell proliferation, invasion, and metastasis and suppresses tumor growth. Colon cancer patients with low HOXB-AS3 levels generally have poor prognoses. HOXB-AS3 competitively binds arginine in the RGG motif of hnRNP A1. In this manner, it blocks hnRNP A1 binding to the pyruvate kinase M (PKM) EI9 sequence, antagonizes hnRNP A1-mediated PKM splicing regulation, and inhibits PKM 2 subtype formation and miR-18a production. HOXB-AS3 downregulates PKM2 but upregulates PKM1. PKM2 is a key regulator of aerobic glycolysis and increases lactic acid production. Therefore, HOXB-AS3 inhibits aerobic glycolysis in CRC cells. The loss of HOXB-AS3 is a key oncogenic event in CRC metabolic reprogramming [175] (Fig. 2b).

CircLgr4-peptide

CircLgr4 is highly expressed in advanced CRC and is associated with poor prognosis. LGR4 is also highly expressed in colorectal tumors and activates Wnt/ β -catenin signaling via ubiquitination and FZD receptor stabilization. Thus, it drives colorectal stem cell self-renewal and invasion. *CircLgr4* encodes the *circLgr4*-peptide, which interacts with LGR4 to activate the LGR4-Wnt signaling pathway. *CircLgr4* drives colorectal stem cell self-renewal and invasion in a manner dependent on LGR4. The *circLgr4*-peptide-Lgr4 axis may be used in targeted CRC therapy [176] (Fig. 2c).

β -catenin-370aa

Circ β -catenin is derived from *CTTNB1*, which encodes β -catenin, a major regulator of the Wnt pathway in liver cancer. *Circ β -catenin* is upregulated in hepatocarcinoma tissues and is localized mainly to the cytoplasm. It has



an ORF and an active IRES encoding the 370-aa β -catenin isomer β -catenin-370aa. *Circ β -catenin* knock-down inhibits hepatoma cell growth and migration in vitro and in vivo, impedes tumorigenesis and metastasis, and suppresses the Wnt/ β -catenin pathway. Construction of the *circ β -catenin* expression vector with an initiation codon mutation disclosed that its functionality could be attributed to its protein-coding ability rather than its non-coding property. *Circ β -catenin* knockdown had no effect on the *CTTNB1* mRNA level but significantly reduced the β -

catenin stability is closely linked to its phosphorylation state. After β -catenin is phosphorylated by GSK3 β , it is ubiquitinated by the ubiquitin ligase β -TrCP and degraded by the proteasome. β -catenin-370aa, encoded by *circ β -catenin*, interacts with GSK3 β and acts as a bait to block it from binding the full-length β -catenin protein. In this manner, it represses GSK3 β -induced β -catenin degradation. In liver cancer, β -catenin-370aa stabilizes β -catenin by reducing its ubiquitination, activating the Wnt/ β -catenin pathway, and promoting tumor growth [168] (Fig. 2d).

Nobody

LINC01420 is a lncRNA that is highly expressed in nasopharyngeal carcinoma. The overall survival rate is low in patients with nasopharyngeal carcinoma presenting with elevated *LINC01420* expression. *LINC01420* knockdown significantly inhibits nasopharyngeal carcinoma cell invasion [177]. The sORF of *LINC01420/LOC550643* encodes a highly sequence-conserved microprotein named nobody. It interacts with an mRNA capping protein, directly binds EDC4, removes the 5' cap from mRNA, promotes 5'-to-3' decay, and regulates the degradation of normal and aberrant transcripts. Nobody is localized mainly to P-bodies. Its level decreases with increasing P-body number. The latter perturbs the homeostasis of endogenous cellular nonsense-mediated decay substrates. Nevertheless, the effects of this process on tumor growth, development, and metabolism are unclear [178] (Fig. 2e).

Other functional peptides

LINC00961 is substantially downregulated in human non-small cell lung cancer (NSCLC). Low tissue *LINC00961* levels are associated with clinical stage, lymph node metastasis, and shorter survival time in NSCLC patients [179, 180]. *LINC00961* may also inhibit tumor progression in oral squamous and renal cell carcinoma, glioma, and other cancers [181–184]. Matsu-moto et al. reported that *LINC00961* is translatable. Its encoded small peptide SPAR is localized to the late lysosome and interacts with lysosomal V-ATPase. SPAR functions upstream of Rags and the Ragulator complex and at the v-ATPase level. It induces interactions of the v-ATPase-Ragulator-Rags supercomplex. SPAR impedes lysosomal mTORC1 reuptake, inhibits mTORC1 activation by amino acid stimulation, and affects muscle regeneration [185–187]. *Circ-ZNF609* is formed from the cyclization of the second exon of *ZNF609*. It is up-regulated in nasopharyngeal carcinoma, renal and breast cancer, and other cancers. *Circ-ZNF609* knockdown dramatically inhibits cancer cell proliferation, invasion, and metastasis [188–190]. Bozzoni et al. reported that *circ-ZNF609* was strongly expressed in muscle cells, highly conserved evolutionarily, and contained a 753-nt ORF. Its UTR had IRES-like activity and encoded a protein in a splicing-dependent manner. This peptide regulated myoblast proliferation [78]. MiPEP-200a and miPEP-200b, encoded by primary miRNAs (*miR-200a* and *miR-200b*), can inhibit the migration of prostate cancer cells by regulating the epithelial to mesenchymal transition of tumor cells [191].

CircRNAs, lncRNAs, and the small peptides they encode may regulate tumorigenesis. Moreover, certain ncRNAs in various species encode proteins that regulate various biological and disease processes in vivo. For

example, peptides 11–32 aa long encoded by sORFs from polished rice control epidermal differentiation in *Drosophila* by modifying the transcription factor Shaven-baby [14]. Myomodulin (MLN) is a highly conserved micropeptide encoded by a 138-nt ORF in a lncRNA. MLN structurally and functionally resembles phospholipids and phosphatidylcholine and inhibits SERCA in a similar manner. In this way, MLN regulates muscle motility [15]. A muscle-specific lncRNA encodes a small 34-aa peptide, DWORF. It increases the activity of SERCA pumps which, in turn, enhance cardiac contractility during a heart attack [192]. In the *Drosophila* heart, a lncRNA (*pncr003:2L*) encodes two peptides ≤ 30 -aa long that regulate calcium transport and affect muscle contraction [23]. Pauli et al. found that the short, conserved polypeptide Toddler encoded by a lncRNA in zebrafish promotes cell movement during gastrulation by activating APJ/apelin receptor signaling [193]. *Pri-miR171b* from alfalfa and *pri-miR165a* from *Arabidopsis* produce peptides that promote the accumulation of mature miRNAs and downregulate target genes regulating root development [194]. Kadener et al. reported that *circMbl*-encoded proteins are enriched in synaptosomes and modulated by starvation and FOXO [94]. Abou-Haidar et al. found a covalently closed, 220-nt circular RNA in a viroid. The translated protein was rich in basic amino acids, expressed only in RYMV-infected rice plants, and bound homologous (scRYMV) and heterologous [potato virus X] RNA [195].

Conclusions and future perspectives

ncRNA-encoded proteins have attracted a great deal of scientific curiosity. Research has established the existence and confirmed the importance of ncRNA-encoded functional peptides. However, the assessment of ncRNA coding potential is difficult [79]. The database used to predict interspecies conservation of ORFs, IRES, and m6A in ncRNAs is incomplete, and experimental validation protocols are still under development [196]. Most circRNAs are produced by protein-encoded exons, which may overlap with their associated mRNAs and render it difficult to distinguish the source of the translation product. High-throughput analytical and detection methods such as ribosome profiling have technical challenges [149, 152]. The identification of small peptides requires specific biochemical and bioinformatics methods seldom applied in genome-wide characterization. Moreover, cell- and tissue-specific expression complicate these assays. Therefore, the actual number of translatable sORFs and their biological functions remain unknown.

Here, we reviewed the recent advances in ncRNA-encoded small peptides regulating human cancer behavior. This investigation provided new perspectives on

ncRNA functions and mechanisms. Therefore, it also suggests that future research on ncRNA may be conducted in depth in several areas, including whether there are more functional peptides or proteins encoded by ncRNA, were the ncRNAs of earlier studies analyzed as RNA or were they examined for their potential coding functions, what is the mechanism of the dynamic translation of ncRNAs encoding functional peptides, do ncRNAs encoding small peptides undergo post-translational modification in a manner similar to that for mRNA, and which factors and conditions affect ncRNA translation.

In the future, functional peptides encoded by ncRNAs may be routinely applied in cancer research, therapy, diagnostics, and prognostics, due to their potential developmental value and clinical utility. NcRNAs can encode some cancer-suppressive peptides/proteins (e.g., FBXW7-185aa, SHPRH-146aa, AKT3-174aa, and PINT87aa). Researchers can deliver these peptides/proteins to tumor cells through nanoparticles or recombine them with adenovirus and inject them into patients as anti-cancer therapy [197]. Moreover, these peptides/proteins can be used with classical anticancer drugs or in combination with traditional radiotherapy and chemotherapy to enhance the effectiveness of cancer therapy. These functional peptides encoded by ncRNAs can also play an important role in tumorigenesis, which makes them potential new targets for drug development. Researchers are also attempting to rescue or strengthen the function of tumor suppressor peptides/proteins by vaccination with synthetic peptides or viral vector vaccines encoding relevant peptides sequences for cancer therapy [198]. The application of these hidden peptides/proteins encoded by ncRNAs as therapy targets in cancer is increasingly promising. Additionally, ncRNA itself can perform biological functions and act as a molecular marker or potential target. Therefore, both functional peptides and ncRNAs can be used as cancer biomarkers for clinical applications at the dual levels of transcription and translation, helping to improve the accuracy and specificity of diagnosis and treatment. In the future, the differential expression and prognostic correlation of these peptides/proteins in cancer may also be determined through more experimental analysis and clinical examination, such as the immunohistochemical analysis of paraffin sections of tumor tissues and body fluid examination.

Here, we discussed how genetic information may also be transferred from ncRNAs to proteins and that this mechanism may participate considerably in the regulation of certain biological and oncological processes. This may help us further clarify biological operating mechanisms and regularity. As functional peptides encoded by ncRNAs is a comparatively new experimental and research field, its mechanisms, functions, regulatory factors, and prospective clinical and scientific applications require and merit further investigation.

Abbreviations

BLAST: Basic local alignment search tool; circRNA: Circular RNA; COGs: Cluster of orthologous groups; CRC: Colorectal cancer; EST: Expressed sequence tag; GSK3 β : Glycogen synthase kinase 3 β ; HPV: Human papillomavirus; IRES: Internal ribosome entry site; LC-MS/MS: Liquid chromatography/tandem mass spectrometry; lncRNA: long non-coding RNA; m6A: N6-methyladenosine; MCC: Matthews correlation coefficient; miRNA: microRNA; MLN: Myomodulin; ncRNA: non-coding RNA; NMD: Nonsense-mediated decay; NSCLC: Non-small cell lung cancer; ORF: Open reading frame; PCNA: Proliferating cell nuclear antigen; PDK1: Phosphoinositide-dependent kinase-1; PI3K: Phosphoinositide 3-kinase; PKM: Pyruvate kinase M; pri-miRNA: Primary microRNA transcript; PseDNC: Pseudonucleotide composition; PseKNC: Pseudo k-tupler nucleotide composition; RFP: Ribosome footprints; RNC: Ribosome-nascent chain complex; RPF: Ribosome-protected fragments; SERCA: Sarco-endoplasmic reticulum Ca²⁺ adenosine triphosphatase; SHPRH: SNF2 histone linker PHD RING helicase; sORF: Short open reading frame; SPAR: Small regulatory polypeptide of amino acid response; SRAMP: Sequence-based RNA adenosine methylation site predictor; SSRP: Small subunit ribosomal proteins; SVM: Support Vector Machine; UTR: Untranslated region

Acknowledgements

Not applicable

Authors' contributions

PW, YM, MP, TT, YZ and XD collected the related paper and finished the manuscript and figures. WX, ZZ gave constructive guidance and made critical revisions. FX, CG, XW, YL, XL, GL participated in the design of this review. All authors read and approved the final manuscript.

Funding

This work was supported partially by grants from the National Natural Science Foundation of China (81972776, 81872278, 81772928, 81702907, and 81672683), the Natural Science Foundation of Hunan Province (2018SK21210, 2018SK21211, 2018JJ3704, 2018JJ3815 and 2017SK2105), the Hunan Provincial Innovation Foundation for Postgraduate (CX20190234, CX20190235), and the Fundamental Research Funds for the Central South University (2019zzts179, 2019zzts178, 2019zzts325, 2019zzts730, 2019zzts890).

Availability of data and materials

Not applicable

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹NHC Key Laboratory of Carcinogenesis and Hunan Key Laboratory of Translational Radiation Oncology, Hunan Cancer Hospital and The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, Hunan, China. ²Key Laboratory of Carcinogenesis and Cancer Invasion of the Chinese Ministry of Education, Cancer Research Institute, Central South University, Changsha, Hunan, China. ³Hunan Key Laboratory of Nonresolving Inflammation and Cancer, Disease Genome Research Center, the Third Xiangya Hospital, Central South University, Changsha, Hunan, China. ⁴Department of Medicine, Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas, USA.

Received: 19 November 2019 Accepted: 28 January 2020

Published online: 04 February 2020

References

1. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447:799–816.

2. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
3. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature*. 2012;489:101–8.
4. Bo H, Fan L, Li J, Liu Z, Zhang S, Shi L, et al. High Expression of lncRNA AFAP1-AS1 Promotes the Progression of Colon Cancer and Predicts Poor Prognosis. *J Cancer*. 2018;9:4677–83.
5. Lian Y, Xiong F, Yang L, Bo H, Gong Z, Wang Y, et al. Long noncoding RNA AFAP1-AS1 acts as a competing endogenous RNA of miR-423-5p to facilitate nasopharyngeal carcinoma metastasis through regulating the Rho/Rac pathway. *J Exp Clin Cancer Res*. 2018;37:253.
6. Wang W, Zhou R, Wu Y, Liu Y, Su W, Xiong W, et al. PVT1 Promotes Cancer Progression via MicroRNAs. *Front Oncol*. 2019;9:609.
7. Jin K, Wang S, Zhang Y, Xia M, Mo Y, Li X, et al. Long non-coding RNA PVT1 interacts with MYC and its downstream molecules to synergistically promote tumorigenesis. *Cell Mol Life Sci*. 2019;76:4275–89.
8. Fan C, Tang Y, Wang J, Wang Y, Xiong F, Zhang S, et al. Long non-coding RNA LOC284454 promotes migration and invasion of nasopharyngeal carcinoma via modulating the Rho/Rac signaling pathway. *Carcinogenesis*. 2019;40:380–91.
9. Bo H, Fan L, Gong Z, Liu Z, Shi L, Guo C, et al. Upregulation and hypomethylation of lncRNA AFAP1AS1 predicts a poor prognosis and promotes the migration and invasion of cervical cancer. *Oncol Rep*. 2019;41:2431–9.
10. Fan CM, Wang JP, Tang YY, Zhao J, He SY, Xiong F, et al. circMAN1A2 could serve as a novel serum biomarker for malignant tumors. *Cancer Sci*. 2019; 110:2180–8.
11. 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;9:59–64.
12. Li LJ, Leng RX, Fan YG, Pan HF, Ye DQ. Translation of noncoding RNAs: Focus on lncRNAs, pri-miRNAs, and circRNAs. *Exp Cell Res*. 2017;361:1–8.
13. Choi SW, Kim HW, Nam JW. The small peptide world in long noncoding RNAs. *Brief Bioinform*. 2019;20:1853–64.
14. Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, et al. Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science*. 2010;329:336–9.
15. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*. 2015;160:595–606.
16. Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M. Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc Natl Acad Sci U S A*. 2002;99:1915–20.
17. Savard J, Marques-Souza H, Aranda M, Tautz D. A segmentation gene in *tribolium* produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell*. 2006;126:559–69.
18. Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol*. 2007;5:e106.
19. 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;17:632–40.
20. Hanada K, Akiyama K, Sakurai T, Toyoda T, Shinozaki K, Shiu SH. sORF finder: a program package to identify small open reading frames with high coding potential. *Bioinformatics*. 2010;26:399–400.
21. Ladoukakis E, Pereira V, Magny EG, Eyre-Walker A, Couso JP. Hundreds of putatively functional small open reading frames in *Drosophila*. *Genome Biol*. 2011;12:R118.
22. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet*. 2014;15:193–204.
23. Magny EG, Pueyo JI, Pearl FM, Cespedes MA, Niven JE, Bishop SA, et al. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science*. 2013;341:1116–20.
24. Tonkin J, Rosenthal N. One small step for muscle: a new micropeptide regulates performance. *Cell Metab*. 2015;21:515–6.
25. Matsumoto A, Nakayama KI. Hidden Peptides Encoded by Putative Noncoding RNAs. *Cell Struct Funct*. 2018;43:75–83.
26. Pan J, Meng X, Jiang N, Jin X, Zhou C, Xu D, et al. Insights into the Noncoding RNA-encoded Peptides. *Protein Pept Lett*. 2018;25:720–7.
27. Zhu S, Wang J, He Y, Meng N, Yan GR. Peptides/Proteins Encoded by Non-coding RNA: A Novel Resource Bank for Drug Targets and Biomarkers. *Front Pharmacol*. 2018;9:1295.
28. 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;178:242–60 e29.
29. Necsculea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*. 2014;505:635–40.
30. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*. 2016;17:47–62.
31. Peng M, Mo Y, Wang Y, Wu P, Zhang Y, Xiong F, et al. Neoantigen vaccine: an emerging tumor immunotherapy. *Mol Cancer*. 2019;18:128.
32. Ren D, Hua Y, Yu B, Ye X, He Z, Li C, et al. Predictive biomarkers and mechanisms underlying resistance to PD1/PD-L1 blockade cancer immunotherapy. *Mol Cancer*. 2020;19:19.
33. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. *Science*. 2005;309: 1559–63.
34. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458:223–7.
35. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22:1775–89.
36. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature*. 2012;482:339–46.
37. van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, et al. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol*. 2014;15:R6.
38. Dinger ME, Gascoigne DK, Mattick JS. The evolution of RNAs with multiple functions. *Biochimie*. 2011;93:2013–8.
39. Banfai B, Jia H, Khatun J, Wood E, Risk B, Gundling WE Jr, et al. Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res*. 2012;22:1646–57.
40. Gascoigne DK, Cheetham SW, Cattenoz PB, Clark MB, Amaral PP, Taft RJ, et al. Pinstripe: a suite of programs for integrating transcriptomic and proteomic datasets identifies novel proteins and improves differentiation of protein-coding and non-coding genes. *Bioinformatics*. 2012;28:3042–50.
41. Ponjavic J, Ponting CP, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res*. 2007;17: 556–65.
42. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 2011;477:295–300.
43. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011;147:789–802.
44. Voinnet O. Origin, biogenesis, and activity of plant microRNAs. *Cell*. 2009; 136:669–87.
45. Stepien A, Knop K, Dolata J, Taube M, Bajczyk M, Barciszewska-Pacak M, et al. Posttranscriptional coordination of splicing and miRNA biogenesis in plants. *Wiley Interdiscip Rev RNA*. 2017;8:e1403.
46. Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*. 2004;10:1957–66.
47. Cullen BR. Transcription and processing of human microRNA precursors. *Mol Cell*. 2004;16:861–5.
48. Waterhouse PM, Hellens RP. Plant biology: Coding in non-coding RNAs. *Nature*. 2015;520:41–2.
49. Church VA, Pressman S, Isaji M, Truscott M, Cizmecioglu NT, Buratowski S, et al. Microprocessor Recruitment to Elongating RNA Polymerase II Is Required for Differential Expression of MicroRNAs. *Cell Rep*. 2017;20: 3123–34.
50. He R, Liu P, Xie X, Zhou Y, Liao Q, Xiong W, et al. circGFRA1 and GFRA1 act as ceRNAs in triple negative breast cancer by regulating miR-34a. *J Exp Clin Cancer Res*. 2017;36:145.
51. Zhong Y, Du Y, Yang X, Mo Y, Fan C, Xiong F, et al. Circular RNAs function as ceRNAs to regulate and control human cancer progression. *Mol Cancer*. 2018;17:79.
52. Zhou R, Wu Y, Wang W, Su W, Liu Y, Wang Y, et al. Circular RNAs (circRNAs) in cancer. *Cancer Lett*. 2018;425:134–42.
53. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell*. 2014;159:134–47.

54. Chen LL, Yang L. Regulation of circRNA biogenesis. *RNA Biol.* 2015;12:381–8.
55. Yang Z, Xie L, Han L, Qu X, Yang Y, Zhang Y, et al. Circular RNAs: Regulators of Cancer-Related Signaling Pathways and Potential Diagnostic Biomarkers for Human Cancers. *Theranostics.* 2017;7:3106–17.
56. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature.* 2013;495:333–8.
57. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PLoS Genet.* 2013;9:e1003777.
58. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. *Nature.* 2013;495:384–8.
59. Piwecka M, Glazar P, Hernandez-Miranda LR, Memczak S, Wolf SA, Rybak-Wolf A, et al. Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science.* 2017;357:eaam8526.
60. Xu L, Feng X, Hao X, Wang P, Zhang Y, Zheng X, et al. CircSETD3 (Hsa_circ_0000567) acts as a sponge for microRNA-421 inhibiting hepatocellular carcinoma growth. *J Exp Clin Cancer Res.* 2019;38:98.
61. Chen CY, Sarnow P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science.* 1995;268:415–7.
62. Yang Y, Fan X, Mao M, Song X, Wu P, Zhang Y, et al. Extensive translation of circular RNAs driven by N(6)-methyladenosine. *Cell Res.* 2017;27:626–41.
63. Abe N, Hiroshima M, Maruyama H, Nakashima Y, Nakano Y, Matsuda A, et al. Rolling circle amplification in a prokaryotic translation system using small circular RNA. *Angew Chem Int Ed Engl.* 2013;52:7004–8.
64. Abe N, Matsumoto K, Nishihara M, Nakano Y, Shibata A, Maruyama H, et al. Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci Rep.* 2015;5:16435.
65. Wesselhoeft RA, Kowalski PS, Anderson DG. Engineering circular RNA for potent and stable translation in eukaryotic cells. *Nat Commun.* 2018;9:2629.
66. Wesselhoeft RA, Kowalski PS, Parker-Hale FC, Huang Y, Bisaria N, Anderson DG. RNA Circularization Diminishes Immunogenicity and Can Extend Translation Duration In Vivo. *Mol Cell.* 2019;74:508–20 e4.
67. Yang Y, Gao X, Zhang M, Yan S, Sun C, Xiao F, et al. Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J Natl Cancer Inst.* 2018;110:304–15.
68. Zhang M, Huang N, Yang X, Luo J, Yan S, Xiao F, et al. A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene.* 2018;37:1805–14.
69. Mo Y, Wang Y, Xiong F, Ge X, Li Z, Li X, et al. Proteomic Analysis of the Molecular Mechanism of Lovastatin Inhibiting the Growth of Nasopharyngeal Carcinoma Cells. *J Cancer.* 2019;10:2342–9.
70. Frith MC, Forrest AR, Nourbakhsh E, Pang KC, Kai C, Kawai J, et al. The abundance of short proteins in the mammalian proteome. *PLoS Genet.* 2006;2:e52.
71. Kastenmayer JP, Ni L, Chu A, Kitchen LE, Au WC, Yang H, et al. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res.* 2006;16:365–73.
72. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol.* 2007;9:660–5.
73. Wadler CS, Vanderpool CK. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc Natl Acad Sci U S A.* 2007;104:20454–9.
74. Jackson R, Kroehling L, Khitun A, Bailis W, Jarret A, York AG, et al. The translation of non-canonical open reading frames controls mucosal immunity. *Nature.* 2018;564:434–8.
75. Liu J, Gough J, Rost B. Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genet.* 2006;2:e29.
76. Fan C, Tu C, Qi P, Guo C, Xiang B, Zhou M, et al. GPC6 Promotes Cell Proliferation, Migration, and Invasion in Nasopharyngeal Carcinoma. *J Cancer.* 2019;10:3926–32.
77. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 2013;41:e166.
78. Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, et al. Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol Cell.* 2017;66:22–37 e9.
79. Zhang M, Zhao K, Xu X, Yang Y, Yan S, Wei P, et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat Commun.* 2018;9:4475.
80. Wang Y, Wang Z. Efficient backsplicing produces translatable circular mRNAs. *RNA.* 2015;21:172–9.
81. Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep.* 2016;6:34985.
82. 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:D324–9.
83. Olexiouk V, Van Criekinge W, Menschaert G. An update on sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* 2018;46:D497–502.
84. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, et al. Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* 2003;31:28–33.
85. Min XJ, Butler G, Storms R, Tsang A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* 2005;33:W677–80.
86. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques.* 2000;28:1102–4.
87. Xia S, Feng J, Chen K, Ma Y, Gong J, Cai F, et al. CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res.* 2018;46:D925–9.
88. Lin MF, Carlson JW, Crosby MA, Matthews BB, Yu C, Park S, et al. Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res.* 2007;17:1823–36.
89. Lin MF, Deoras AN, Rasmussen MD, Kellis M. Performance and scalability of discriminative metrics for comparative gene identification in 12 *Drosophila* genomes. *PLoS Comput Biol.* 2008;4:e1000067.
90. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol.* 2010;28:503–10.
91. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics.* 2011;27:i275–82.
92. Meng X, Chen Q, Zhang P, Chen M. CircPro: an integrated tool for the identification of circRNAs with protein-coding potential. *Bioinformatics.* 2017;33:3314–6.
93. Chen L, Ding X, Zhang H, He T, Li Y, Wang T, et al. Comparative analysis of circular RNAs between soybean cytoplasmic male-sterile line NJCMS1A and its maintainer NJCMS1B by high-throughput sequencing. *BMC Genomics.* 2018;19:663.
94. Pamudurti NR, Bartok O, Jens M, Ashwal-Fluss R, Stottmeister C, Ruhe L, et al. Translation of CircRNAs. *Mol Cell.* 2017;66:9–21 e7.
95. Liu M, Wang Q, Shen J, Yang BB, Ding X. Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol.* 2019;16:899–905.
96. Holcik M, Sonenberg N. Translational control in stress and apoptosis. *Nat Rev Mol Cell Biol.* 2005;6:318–27.
97. King HA, Cobbold LC, Willis AE. The role of IRES trans-acting factors in regulating translation initiation. *Biochem Soc Trans.* 2010;38:1581–6.
98. Stoneley M, Willis AE. Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene.* 2004;23:3200–7.
99. Pelletier J, Sonenberg N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature.* 1988;334:320–5.
100. Cevallos RC, Sarnow P. Factor-independent assembly of elongation-competent ribosomes by an internal ribosome entry site located in an RNA virus that infects penaeid shrimp. *J Virol.* 2005;79:677–83.
101. Bonnal S, Boutonnet C, Prado-Lourenco L, Vagner S. IRESdb: the Internal Ribosome Entry Site database. *Nucleic Acids Res.* 2003;31:427–8.
102. Bielecki L, Talbot SJ. Kaposi's sarcoma-associated herpesvirus vCyclin open reading frame contains an internal ribosome entry site. *J Virol.* 2001;75:1864–9.
103. Wu Y, Wei F, Tang L, Liao Q, Wang H, Shi L, et al. Herpesvirus acts with the cytoskeleton and promotes cancer progression. *J Cancer.* 2019;10:2185–93.
104. Hertz MI, Thompson SR. Mechanism of translation initiation by Dicistroviridae IGR IRESs. *Virology.* 2011;411:355–61.
105. Mauro VP, Edelman GM, Zhou W. Reevaluation of the conclusion that IRES-activity reported within the 5' leader of the TIF4631 gene is due to promoter activity. *RNA.* 2004;10:895–7 discussion 898.

106. Baranick BT, Lemp NA, Nagashima J, Hiraoka K, Kasahara N, Logg CR. Splicing mediates the activity of four putative cellular internal ribosome entry sites. *Proc Natl Acad Sci U S A*. 2008;105:4733–8.
107. Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. CirInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol*. 2016;13:34–42.
108. Meganck RM, Borchardt EK, Castellanos Rivera RM, Scalabrino ML, Wilusz JE, Marzluff WF, et al. Tissue-Dependent Expression and Translation of Circular RNAs with Recombinant AAV Vectors In Vivo. *Mol Ther Nucleic Acids*. 2018;13:89–98.
109. Hellen CU, Sarnow P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev*. 2001;15:1593–612.
110. Komar AA, Hatzoglou M. Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell Cycle*. 2011;10:229–40.
111. Mokrejs M, Vopalensky V, Kolenaty O, Masek T, Feketova Z, Sekyrova P, et al. IRESite: the database of experimentally verified IRES structures (www.iresite.org). *Nucleic Acids Res*. 2006;34:D125–30.
112. Zhao J, Wu J, Xu T, Yang Q, He J, Song X. IRESfinder: Identifying RNA internal ribosome entry site in eukaryotic cell using framed k-mer features. *J Genet Genomics*. 2018;45:403–6.
113. Kolekar P, Pataskar A, Kulkarni-Kale U, Pal J, Kulkarni A. IRESPred: Web Server for Prediction of Cellular and Viral Internal Ribosome Entry Site (IRES). *Sci Rep*. 2016;6:27436.
114. Hong JJ, Wu TY, Chang TY, Chen CY. Viral IRES prediction system - a web server for prediction of the IRES secondary structure in silico. *PLoS One*. 2013;8:e79288.
115. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell*. 2017;169:1187–200.
116. Ries RJ, Zaccara S, Klein P, Olerariu-George A, Namkoong S, Pickering BF, et al. m(6)A enhances the phase separation potential of mRNA. *Nature*. 2019;571:424–8.
117. Meyer KD, Jaffrey SR. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat Rev Mol Cell Biol*. 2014;15:313–26.
118. Meyer KD, Patil DP, Zhou J, Zinoviev A, Skabkin MA, Elemento O, et al. 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell*. 2015;163:999–1010.
119. Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, et al. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell*. 2015;161:1388–99.
120. Li A, Chen YS, Ping XL, Yang X, Xiao W, Yang Y, et al. Cytoplasmic m(6)A reader YTHDF3 promotes mRNA translation. *Cell Res*. 2017;27:444–7.
121. Shi H, Wang X, Lu Z, Zhao BS, Ma H, Hsu PJ, et al. YTHDF3 facilitates translation and decay of N(6)-methyladenosine-modified RNA. *Cell Res*. 2017;27:315–28.
122. Wang S, Chai P, Jia R, Jia R. Novel insights on m(6)A RNA methylation in tumorigenesis: a double-edged sword. *Mol Cancer*. 2018;17:101.
123. Liu J, Harada BT, He C. Regulation of Gene Expression by N(6)-methyladenosine in Cancer. *Trends Cell Biol*. 2019;29:487–99.
124. Zhao J, Lee EE, Kim J, Yang R, Chamseddin B, Ni C, et al. Transforming activity of an oncoprotein-encoding circular RNA from human papillomavirus. *Nat Commun*. 2019;10:2300.
125. Zhang Y, Hamada M. DeepM6ASeq: prediction and characterization of m6A-containing sequences using deep learning. *BMC Bioinformatics*. 2018;19:524.
126. Wei L, Chen H, Su R. M6APred-EL: A Sequence-Based Predictor for Identifying N6-methyladenosine Sites Using Ensemble Learning. *Mol Ther Nucleic Acids*. 2018;12:635–44.
127. Qiang X, Chen H, Ye X, Su R, Wei L. M6AMRFS: Robust Prediction of N6-Methyladenosine Sites With Sequence-Based Features in Multiple Species. *Front Genet*. 2018;9:495.
128. Zhou Y, Zeng P, Li YH, Zhang Z, Cui Q. SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res*. 2016;44:e91.
129. Chen W, Feng P, Ding H, Lin H, Chou KC. iRNA-Methyl: Identifying N(6)-methyladenosine sites using pseudo nucleotide composition. *Anal Biochem*. 2015;490:26–33.
130. Chen W, Ding H, Zhou X, Lin H, Chou KC. iRNA(m6A)-PseDNC: Identifying N(6)-methyladenosine sites using pseudo dinucleotide composition. *Anal Biochem*. 2018;561-562:59–65.
131. Wu X, Wei Z, Chen K, Zhang Q, Su J, Liu H, et al. m6Acomet: large-scale functional prediction of individual m(6)A RNA methylation sites from an RNA co-methylation network. *BMC Bioinformatics*. 2019;20:223.
132. Chen K, Wei Z, Zhang Q, Wu X, Rong R, Lu Z, et al. WHISTLE: a high-accuracy map of the human N6-methyladenosine (m6A) epitranscriptome predicted using a machine learning approach. *Nucleic Acids Res*. 2019;47:e41.
133. Liu Z, Xiao X, Yu DJ, Jia J, Qiu WR, Chou KC. pRNAm-PC: Predicting N(6)-methyladenosine sites in RNA sequences via physical-chemical properties. *Anal Biochem*. 2016;497:60–7.
134. Li GQ, Liu Z, Shen HB, Yu DJ. TargetM6A: Identifying N(6)-Methyladenosine Sites From RNA Sequences via Position-Specific Nucleotide Propensities and a Support Vector Machine. *IEEE Trans Nanobioscience*. 2016;15:674–82.
135. Xiang S, Yan Z, Liu K, Zhang Y, Sun Z. AthMethPre: a web server for the prediction and query of mRNA m(6)A sites in *Arabidopsis thaliana*. *Mol Biosyst*. 2016;12:3333–7.
136. Xiang S, Liu K, Yan Z, Zhang Y, Sun Z. RNAMethPre: A Web Server for the Prediction and Query of mRNA m6A Sites. *PLoS One*. 2016;11:e0162707.
137. Panek J, Kolar M, Vohradsky J, Shivaya VL. An evolutionary conserved pattern of 18S rRNA sequence complementarity to mRNA 5' UTRs and its implications for eukaryotic gene translation regulation. *Nucleic Acids Res*. 2013;41:7625–34.
138. Hoepfner MP, Denisenko E, Gardner PP, Schmeier S, Poole AM. An Evaluation of Function of Multicopy Noncoding RNAs in Mammals Using ENCODE/FANTOM Data and Comparative Genomics. *Mol Biol Evol*. 2018;35:1451–62.
139. Herpin A, Schmidt C, Kneitz S, Gobe C, Regensburger M, Le Cam A, et al. A novel evolutionary conserved mechanism of RNA stability regulates synexpression of primordial germ cell-specific genes prior to the sex-determination stage in medaka. *PLoS Biol*. 2019;17:e3000185.
140. Sun K, Chen X, Jiang P, Song X, Wang H, Sun H. iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data. *BMC Genomics*. 2013;14 Suppl 2:S7.
141. Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J*. 2014;33:981–93.
142. Hu L, Xu Z, Hu B, Lu ZJ. COME: a robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. *Nucleic Acids Res*. 2017;45:e2.
143. Tyner C, Barber GP, Casper J, Clawson H, Diekhans M, Eisenhart C, et al. The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res*. 2017;45:D626–34.
144. Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, Rosenbloom KR, et al. The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res*. 2018;46:D762–9.
145. Houshmand M, Mahmoudi T, Panahi MS, Seyedena Y, Saber S, Ataei M. Identification of a new human mtDNA polymorphism (A14290G) in the NADH dehydrogenase subunit 6 gene. *Braz J Med Biol Res*. 2006;39:725–30.
146. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016;33:1870–4.
147. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23:2947–8.
148. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci*. 2018;27:135–45.
149. King HA, Gerber AP. Translatome profiling: methods for genome-scale analysis of mRNA translation. *Brief Funct Genomics*. 2016;15:22–31.
150. Chasse H, Boulben S, Costache V, Cormier P, Morales J. Analysis of translation using polysome profiling. *Nucleic Acids Res*. 2017;45:e15.
151. Thermann R, Hentze MW. Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation. *Nature*. 2007;447:875–8.
152. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009;324:218–23.
153. Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*. 2003;100:3889–94.
154. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc*. 2012;7:1534–50.
155. Ingolia NT. Ribosome Footprint Profiling of Translation throughout the Genome. *Cell*. 2016;165:22–33.
156. Gobet C, Naef F. Ribosome profiling and dynamic regulation of translation in mammals. *Curr Opin Genet Dev*. 2017;43:120–7.

157. Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJ, Jackson SE, et al. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep*. 2014;8:1365–79.
158. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, et al. Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods*. 2016;13:165–70.
159. 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.
160. Wang H, Wang Y, Xie S, Liu Y, Xie Z. Global and cell-type specific properties of lincRNAs with ribosome occupancy. *Nucleic Acids Res*. 2017;45:2786–96.
161. Li Q, Ahsan MA, Chen H, Xue J, Chen M. Discovering Putative Peptides Encoded from Noncoding RNAs in Ribosome Profiling Data of Arabidopsis thaliana. *ACS Synth Biol*. 2018;7:655–63.
162. Wang T, Cui Y, Jin J, Guo J, Wang G, Yin X, et al. Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. *Nucleic Acids Res*. 2013;41:4743–54.
163. Deng X, Xiong F, Li X, Xiang B, Li Z, Wu X, et al. Application of atomic force microscopy in cancer research. *J Nanobiotechnology*. 2018;16:102.
164. Jiao Y, Meyerowitz EM. Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol Syst Biol*. 2010;6:419.
165. The UPC. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45:D158–69.
166. Bollineni RC, Koehler CJ, Gislefoss RE, Anonsen JH, Thiede B. Large-scale intact glycopeptide identification by Mascot database search. *Sci Rep*. 2018; 8:2117.
167. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*. 2013;154:240–51.
168. Liang WC, Wong CW, Liang PP, Shi M, Cao Y, Rao ST, et al. Translation of the circular RNA circbeta-catenin promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biol*. 2019;20:84.
169. Begum S, Yiu A, Stebbing J, Castellano L. Novel tumour suppressive protein encoded by circular RNA, circ-SHPRH, in glioblastomas. *Oncogene*. 2018;37: 4055–7.
170. Wang YA, Li XL, Mo YZ, Fan CM, Tang L, Xiong F, et al. Effects of tumor metabolic microenvironment on regulatory T cells. *Mol Cancer*. 2018;17:168.
171. Xiao L, Wei F, Liang F, Li Q, Deng H, Tan S, et al. TSC2D2 identified as a candidate susceptibility gene of multi-cancer pedigree using genome-wide linkage analysis and whole-exome sequencing. *Carcinogenesis*. 2019;40: 819–27.
172. Xia X, Li X, Li F, Wu X, Zhang M, Zhou H, et al. A novel tumor suppressor protein encoded by circular AKT3 RNA inhibits glioblastoma tumorigenicity by competing with active phosphoinositide-dependent Kinase-1. *Mol Cancer*. 2019;18:131.
173. 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;47:8111–25.
174. Zheng X, Chen L, Zhou Y, Wang Q, Zheng Z, Xu B, et al. A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling. *Mol Cancer*. 2019; 18:47.
175. Huang JZ, Chen M, Chen GXC, Zhu S, Huang H, et al. A Peptide Encoded by a Putative lincRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol Cell*. 2017;68:171–84 e6.
176. Zhi X, Zhang J, Cheng Z, Bian L, Qin J. circLgr4 drives colorectal tumorigenesis and invasion through Lgr4-targeting peptide. *Int J Cancer*. 2019. <https://doi.org/10.1002/ijc.32549>.
177. Yang L, Tang Y, He Y, Wang Y, Lian Y, Xiong F, et al. High Expression of LINC01420 indicates an unfavorable prognosis and modulates cell migration and invasion in nasopharyngeal carcinoma. *J Cancer*. 2017;8:97–103.
178. D'Lima NG, Ma J, Winkler L, Chu Q, Loh KH, Corpuz EO, et al. A human microprotein that interacts with the mRNA decapping complex. *Nat Chem Biol*. 2017;13:174–80.
179. Huang Z, Lei W, Tan J, Hu HB. Long noncoding RNA LINC00961 inhibits cell proliferation and induces cell apoptosis in human non-small cell lung cancer. *J Cell Biochem*. 2018;119:9072–80.
180. Jiang B, Liu J, Zhang YH, Shen D, Liu S, Lin F, et al. Long noncoding RNA LINC00961 inhibits cell invasion and metastasis in human non-small cell lung cancer. *Biomed Pharmacother*. 2018;97:1311–8.
181. Lu XW, Xu N, Zheng YG, Li QX, Shi JS. Increased expression of long noncoding RNA LINC00961 suppresses glioma metastasis and correlates with favorable prognosis. *Eur Rev Med Pharmacol Sci*. 2018;22:4917–24.
182. Chen D, Zhu M, Su H, Chen J, Xu X, Cao C. LINC00961 restrains cancer progression via modulating epithelial-mesenchymal transition in renal cell carcinoma. *J Cell Physiol*. 2019;234:7257–65.
183. Mo Y, Wang Y, Zhang L, Yang L, Zhou M, Li X, et al. The role of Wnt signaling pathway in tumor metabolic reprogramming. *J Cancer*. 2019;10: 3789–97.
184. Zhang L, Shao L, Hu Y. Long noncoding RNA LINC00961 inhibited cell proliferation and invasion through regulating the Wnt/beta-catenin signaling pathway in tongue squamous cell carcinoma. *J Cell Biochem*. 2019;120:12429–35.
185. Rion N, Ruegg MA. LncRNA-encoded peptides: More than translational noise? *Cell Res*. 2017;27:604–5.
186. Tajbaksh S. lncRNA-Encoded Polypeptide SPAR(s) with mTORC1 to Regulate Skeletal Muscle Regeneration. *Cell Stem Cell*. 2017;20:428–30.
187. 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;541:228–32.
188. Wang S, Xue X, Wang R, Li X, Li Q, Wang Y, et al. CircZNF609 promotes breast cancer cell growth, migration, and invasion by elevating p70S6K1 via sponging miR-145-5p. *Cancer Manag Res*. 2018;10:3881–90.
189. Xiong Y, Zhang J, Song C. CircRNA ZNF609 functions as a competitive endogenous RNA to regulate FOXp4 expression by sponging miR-138-5p in renal carcinoma. *J Cell Physiol*. 2019;234:10646–54.
190. Zhu L, Liu Y, Yang Y, Mao XM, Yin ZD. CircRNA ZNF609 promotes growth and metastasis of nasopharyngeal carcinoma by competing with microRNA-150-5p. *Eur Rev Med Pharmacol Sci*. 2019;23:2817–26.
191. Fang J, Morsalin S, Rao V, Reddy E. Decoding of Non-Coding DNA and Non-Coding RNA: Pri-Micro RNA-Encoded Novel Peptides Regulate Migration of Cancer Cells. *J Pharm Sci Pharmacol*. 2017;3:23–7.
192. 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;351:271–5.
193. Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, et al. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science*. 2014;343:1248636.
194. Lauresergues D, Couzigou JM, Clemente HS, Martinez Y, Dunand C, Becard G, et al. Primary transcripts of microRNAs encode regulatory peptides. *Nature*. 2015;520:90–3.
195. AbouHaidar MG, Venkataraman S, Golshani A, Liu B, Ahmad T. Novel coding, translation, and gene expression of a replicating covalently closed circular RNA of 220 nt. *Proc Natl Acad Sci U S A*. 2014;111:14542–7.
196. Saghatelian A, Couso JP. Discovery and characterization of smORF-encoded bioactive polypeptides. *Nat Chem Biol*. 2015;11:909–16.
197. Xiong F, Deng S, Huang HB, Li XY, Zhang WL, Liao QJ, et al. Effects and mechanisms of innate immune molecules on inhibiting nasopharyngeal carcinoma. *Chin Med J (Engl)*. 2019;132:749–52.
198. Ge J, Wang J, Wang H, Jiang X, Liao Q, Gong Q, et al. The BRAF V600E mutation is a predictor of the effect of radioiodine therapy in papillary thyroid cancer. *J Cancer*. 2020;11:932–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

