

Review

Not peer-reviewed version

---

# Exploring the Dark Matter of Human Proteome: Emerging Role of Non-canonical Open Reading Frame (ncORF) in Cancer Diagnosis, Biology, and Therapy

---

[Anni Ge](#) , [Curtis Chan](#) , [Xiaolong Yang](#) \*

Posted Date: 28 June 2024

doi: 10.20944/preprints202406.2031.v1

Keywords: human proteome; ncORF; micropeptide; microprotein; cancer biology, therapy



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

# Exploring the Dark Matter of Human Proteome: Emerging Role of Non-Canonical Open Reading Frame (ncORF) in Cancer Diagnosis, Biology, and Therapy

Anni Ge, Curtis Chan and Xiaolong Yang \*

Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON K7L 3N6, Canada; 21ag87@queensu.ca; 18cbc4@queensu.ca

\* Correspondence: yangx@queensu.ca; Tel.: +1-613533-6000 (ext. 75998)

**Simple Summary:** With advanced bioinformatic and sequencing technologies, many non-coding RNAs have been found to harbor non-canonical open reading frame (ncORF). Some studies suggest that many ncORF-encoded microproteins possess important roles in cancer biology, yet the functional roles of these ncORF-encoded peptides are mostly unknown. The purpose of this review is to provide a comprehensive overview of existing knowledge on ncORF-encoded microproteins in cancer biology and highlight their role as potential prognostic and therapy targets in cancer.

**Abstract:** Cancer develops from abnormal cell growth in the body, causing significant mortalities every year. To date, potent therapeutic approaches have been developed to eradicate tumor cells, but intolerable toxicity and drug resistance can occur in treated patients, limiting the efficiency of existing treatment strategies. Therefore, searching for novel genes critical for cancer progression and therapeutic response is urgently needed for successful cancer therapy. Recent advances in bioinformatic and proteomic techniques have allowed the identification of a novel category of peptides encoded by non-canonical open reading frames (ncORFs) from historically non-coding genomic regions. Surprisingly, many of the ncORFs express functional microproteins that play a vital role in human cancers. In this review, we provide a comprehensive description of different ncORF types with coding capacity and technological methods in discovering ncORFs among human genomes. We also summarize the diagnostic and prognostic value of these ncORF as well as their carcinogenic role in cancer progression and underestimated potential as therapeutic targets.

**Keywords:** human proteome; ncORF; micropeptide; microprotein; cancer biology; therapy

## 1. Introduction

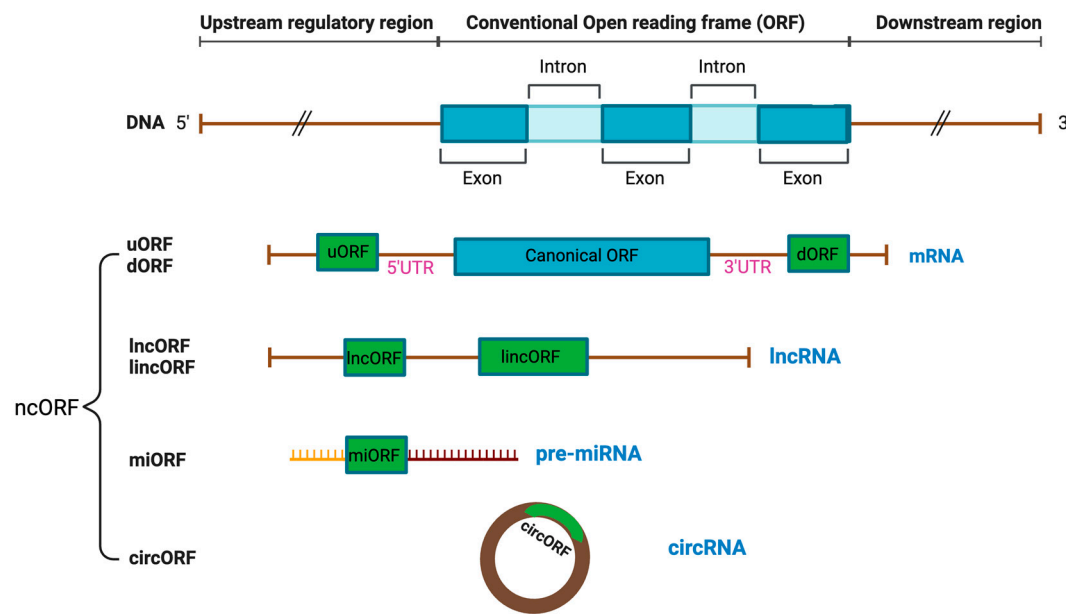
### 1.1. Definition of ncORF

Non-canonical open reading frames (ncORFs) are regions of translation that occur separately from the canonical protein-coding sequences [1]. In 2004, the Human Genome Project annotated the entire human genome and identified the location and sequences of around 20,000 protein-coding genes [2]. Historically, stringent rules were placed on the prediction of protein-coding open reading frames (ORFs) to minimize false-positives [3]. For example, coding sequences (CDS) must meet a size cut-off of 100 amino acids, initiate translation from the conventional AUG start codon, and have monocistronic transcription, meaning each mRNA molecule could not encode more than one polypeptide [3]. However, these rules created a blind spot in the human proteome for ncORFs that did not fit these criteria. Over the past two decades, improved understanding of gene expression and modern detection methods have rendered these criteria obsolete. To date, numerous ncORFs have

been identified and are believed to have the potential to regulate gene expression or produce proteins with unique biological roles [4].

### 1.2. Classification of ncORF

Technological advances have enabled the discovery of many neglected ncORFs, which are classified into different categories, such as long non-coding ORFs (lncORFs), circular ORFs (circORFs), intronic ORFs, primary microRNA-derived ORFs (pri-miORFs), and small ORFs (smORFs) (Figure 1) [1,3]. This section reviews ncORFs based on each category.



**Figure 1.** Different types of RNAs with coding capacity. Distinct categories of non-canonical open reading frames (ncORFs) have been identified from non-coding RNAs, which include 5'UTR upstream ORF (uORF), 3' UTR downstream ORF (dORF), long non-coding ORF (lncORF) and long intergenic non-coding ORF (lincORF) from long non-coding RNAs (lncRNAs), miRNA ORF (miORF) from miRNA and ORF derived from circular RNAs (circORF).

ncORFs can be found within the 5' and 3' untranslated regions (UTR) of the annotated CDS in an alternate reading frame from the CDS, including upstream (uORFs) and downstream ORFs (dORFs), respectively (Figure 1)[3]. uORFs regulate translation of the main CDS through interactions with the 43S ribosome pre-initiation complex in the "leaky scanning" mechanism. As the 43S ribosome pre-initiation complex scans the mRNA and encounters the uORF, which often has a near-cognate non-AUG start codon, it has three options: (1) initiate translation at the uORF and disassemble upon reaching the stop codon, resulting in no translation of the downstream CDS, (2) fail to recognize the uORF and continues scanning until it reaches the downstream CDS, resulting in CDS translation, and (3) initiate translation at the uORF, remain bound to the mRNA, and resume scanning, resulting in CDS translation [3,5]. uORFs can encode proteins that interact or functionally cooperate with the downstream CDS protein [6]. A ncORF positioned at the upstream of a protein kinase C (PKC) isoform encodes uORF2, which significantly impairs viability of breast cancer and leukemia cells through inhibition of the PKC family [7]. Compared to uORFs, the function of dORFs is less well-defined, but it has been hypothesized that translation of dORFs can enhance main CDS translation by recruiting ribosomes or translation initiation factors to the main CDS [8]. Unlike uORFs and dORFs, alternative ORFs (altORFs) overlap with the main CDS, but in a shifted or alternate ORF [3]. The altFUS protein is a prime example of an altORF, where it overlaps its main CDS, FUS, in an open reading frame that is shifted by a single nucleotide [9]. ncORFs can also be in-frame with the main CDS, expressing extended or truncated isoforms of the annotated proteins based on the locations and presence of the ncORF start and stop codons [3,10]. For example, the MYC gene has an

alternate CUG start codon upstream of the main AUG start codon, which can translate into N-terminally extended variants of the MYC protein [11].

lncORFs are encoded within long non-coding RNA (lncRNA), which are RNA transcripts longer than 200 nucleotides annotated as non-coding (Figure 1)[3,12]. lncRNAs are pervasively expressed and play important roles in gene regulation as untranslated RNA molecules [13]. Many RNA sequences that are previously considered as lncRNA have now been discovered to contain ORFs with protein-coding potential [14–20]. An example of a lncORF is the steroid receptor RNA activator (SRA) gene. In its RNA form, SRA functions as a nuclear receptor coactivator[21,22], while it also encodes a protein, SRAP, associated with breast cancer cell motility [21,22]. lncRNAs can originate from pseudogenes, referring to DNA regions that contain presumably non-functional, untranslatable copies annotated genes derived from retrotransposition of processed mRNAs and segmental duplication [12,23]. Increasing evidence has revealed peptides translated from pseudogenic ncORFs [24]. In the case of lncRNAs located in between genes with no overlap with the CDS, they are recognized as intergenic, and ncORFs derived from these regions are called long intergenic non-coding ORFs (lincORFs) or intergenic ORFs [25]. Though rare, intronic ORFs with translational capabilities have also been discovered in the introns of pre-sliced mRNAs [26]. Another type of non-coding RNAs (ncRNAs) is micro-RNAs (miRNAs) with 18-24 nucleotides, which originate as primary transcripts of miRNAs (pri-miRNAs) and become mature miRNAs through specific cleavage and progressing [27,28]. ncORFs derived from these unprocessed transcripts of miRNAs are referred to as pri-miORFs (Figure 1).

Circular RNAs (circRNAs) are single-stranded RNA molecules covalently linked at the 5' and 3' ends as a product of back-splicing and display translation potential, resulting in the discovery of a new class of ncORFs, termed circORFs [3,29]. circORFs can regulate gene expression through their interactions with micro-RNA and circRNA binding proteins (cRBPs) [30]. Due to the lack of the 5' end of linear messenger transcripts (5' cap), circRNAs were historically considered to be a type of lncRNA [12]. However, through cap-independent initiation mechanisms, circRNAs have been demonstrated to encode functionally significant circORF proteins (Figure 1)[31]. An example of a circORF is circMAPK1, encoding MAPK1–109aa, a microprotein with tumor suppressive functions in gastric cancer cells through interactions with MEK1 of the MAPK signaling pathway [32].

Finally, smORFs are protein-coding sequences that are 100 amino acids or smaller in size [3]. This arbitrary size limit is due to the historical assumption that proteins smaller than 100 amino acids are statistically unlikely to be functional and the difficulty of detecting DNA sequences less than 300 nucleotides long before the emergence of modern omics-based technology, such as ribosome sequencing [3]. Regardless, all other classifications of ncORFs are also classified as smORF if they meet the size cut-off. Outside of the nuclear genome, 8 instances of smORFs have been identified in mitochondrial DNA, which encodes mitochondrial-derived peptides (MDPs) [33].

### 1.3. Identification of ncORF

Computational analyses have been the most common method in annotating canonical genes, where they are defined as the longest evolutionarily conserved AUG-containing ORF in an mRNA [34]. However, ncORFs/smORFs are trickier to predict and produce more noise, thus alternative methods are required for their annotation [35]. The detection of ncORFs in the early stages was challenging due to the conservativeness of existing gene identification algorithms, which were not designed for ncORFs [36]. sORF finder is one of the first successful ncORF prediction tools, resulting in the identification of over 2,000 intergenic ORFs in a plant species called *Arabidopsis thaliana* [37]. Technological improvements led to the establishment of various ncORF/smORF prediction tools and databases, such as PhyloCSF, MiPepid, uPEPperoni, and PhyloCSF. [36,38–41]. While bioinformatic tools provide great value in reduced cost and convenience compared to experimental validations of ncORFs, they are inadequate in confirming the translation of novel ncORFs [36]. Investigation and characterizations of these translatable ncORFs require further study.

During translation, ribosomes enable protein synthesis by reading mRNA transcripts. As the ribosome reads the mRNA, it protects 30-31 nucleotides of the mRNA from nuclease degradation,

creating a ribosome footprint. In 2009, Ingolia et al. exploited this phenomenon to develop the technique known as Ribo-Seq [42]. By converting these footprints to DNA sequences and utilizing deep sequencing, they were able to map the precise positions of ribosomes and quantify translational activity. Actively translated ORFs are characterized by continuous 3-nucleotide periodicity, which results from the 80S ribosomes reading the mRNA template one codon at a time [34]. Furthermore, ribosome footprint density can be used to deduce the rate of translation for a particular polypeptide, with higher ribosome density meaning slower elongation, and vice versa [43]. Another advantage of Ribo-Seq is its ability to identify ncORFs regardless of their start codon, which is invaluable since a large portion of ncORFs do not initiate translation with an AUG start codon [44,45]. Ribo-Seq data can be analyzed using computational methods, such as RibORF, which calculates the overall probability of translation [24,35]. However, validation of ncORFs at the protein level is required, as ribosome occupancy alone cannot distinguish between coding and non-coding RNA transcripts, limiting the reliability of Ribo-seq [36,46,47]. To reduce the number of false positives detected, a method called polysome profiling or Poly-Ribo-Seq was developed [48]. Scanning of the 40S ribosome subunit alone can result in ribosome footprints and the false detection of translation. Poly-Ribo-Seq takes advantage of the fact that polysomes, which consists of multiple ribosomes, bind to mRNAs collectively during active translation. By isolating polysomal fractions, actively translation regions can be differentiated from false positives derived from single ribosomes or ribosome subunits. Yet, this is not a perfect solution as it leaves a blind spot for particularly short smORFs, which are not long enough to bind multiple ribosomes [48].

MS-based proteomic techniques are often used to compliment Ribo-Seq in the direct identification and quantification of ncORF-encoded proteins [36,49]. MS-based approaches can also aid in determining post-translation modifications, furthering our understanding of the protein’s functional mechanisms [50].

Finally, the biological functions of ncORF-encoded protein can be validated using CRISPR/Cas9-based approaches [36,51,52]. Knockout (loss-of-function) or overexpression (gain-of-function) assays can be performed to elucidate the functions of specific ncORF-encoded proteins[36,51,52]. The CRISPR/Cas9 system can also be used to observe the expression and localization of ncORF-encoded proteins by knocking-in epitope tags into DNA sequences of ncORFs, which can be detected by the corresponding antibodies [53]. Genome-wide CRISPR-based mutagenesis screens have successfully identified high-priority ncORFs that may encode functionally significant proteins such as GREP1, ASNSD1-uORF or ASDURF [51,52,54]. CRISPR screening can also be combined with RNA sequencing (RNA-Seq) to confirm the biological function of ncORFs. By perturbing expression of candidate ncORFs using CRISPR/Cas9 and observing the changes in RNA-seq profiles within a single cell, the molecular mechanisms of the ncORF-encoded peptides can be readily elucidated [51].

2. ncORFs in Cancer Diagnosis and Prognosis

Selective expression of certain ncORFs have been observed in specific cancers and is associated with worse disease outcomes, enabling early detection and progression prediction of certain diseases [55–59]. Here, we discuss known non-canonical proteins and their involvements as diagnostic and prognostic factors (Table 1).

Table 1. ncORFs as prognosis and diagnosis biomarkers.

ncORFs	Cancer types	Biomarker	References
HOXB-AS3, SRSP, RGRP, ASAP	CRC	Prognosis	[15,6–62]
ASRPS, CIP2A- BP	BC	Prognosis	[16,18]
pTINCR	BC, C5CC, LAC	Prognosis	

SHPRH-146aa, PINT87aa, FBXW7-185aa	GBM	Prognosis	[63–65]
SMIM30, MPM	HCC	Prognosis	[66,67]
DDUP	OC	Prognosis	[68,69]
SHLP2	PC	Diagnosis	[70]
miPEP133	NPC	Prognosis	[71]
APPLE/ASH1L- AS1	AML	Prognosis Diagnosis	[17]
circMAPK1	GC	Prognosis	[32]
MIAC	RCC	Diagnosis Prognosis	[72]

AML, acute myeloid leukemia; BC, breast cancer; ; CRC, colorectal carcinoma; CSCC, cutaneous squamous cell carcinoma; GBM, glioblastoma; GC, gastric cancer; HCC, hepatocellular carcinoma; LAC: lung adenocarcinoma; NPC, nasopharyngeal carcinoma OC, ovarian cancer; PC, prostate cancer; RCC, renal cell carcinoma.

2.1. Colorectal Cancer

HOXB-AS3 is a lncRNA that plays an oncogenic role in human cancers and displays an elevated RNA expression in the cancer cells [73–76]. Interestingly, Huang *et al.* (2017) discovered that HOXB-AS3 encodes a small peptide, and they found that the expression of HOXB-AS3 RNA and peptide was downregulated in colorectal cancer (CRC) tissues compared to paired adjacent normal tissues [60]. Furthermore, the tumor suppressive effects of the HOXB-AS3 peptide were revealed, and low level of HOXB-AS3 peptide was associated with worse prognosis of CRC patients [60]. Specifically, Kaplan-Meier survival analyses indicated that CRC patients with high HOXB-AS3 peptide expression had a 1.6-fold increase in mean overall survival time compared to those with low expression, suggesting that high expression of HOXB-AS3 peptide might reduce CRC-related mortalities [60]. The expression of SRSP, another lncORF-encoded peptide, was significantly elevated in CRC tissues compared to normal colorectal tissues [61]. Further investigation revealed that SRSP expression was positively associated with clinicopathological features of CRC patients, such as histological grade, presence of lymph node metastasis, clinical stage, and risk of death [61]. CRC patients with upregulated SRSP expression had 3.3-fold reduction in median survival time compared to those with low expression [61]. Other examples of potentially prognostic or diagnostic ncORF-encoded peptides in CRC include RBRP and ASAP, derived from lncRNAs, and PPP1R12A-C, derived from circRNA [15,62,77].

2.2. Breast Cancer

ASRPS, a lncORF-encoded peptide, is significantly downregulated in triple negative breast cancer (TNBC) compared to other breast cancer subtypes, and low ASRPS expression is correlated with worse prognosis and overall survival in TNBC patients [16]. Another lncORF-encoded peptide, CIP2A-BP, showed a similar association to poorer prognosis at low expression in TNBC [18]. Boix and colleagues (2022) demonstrated that the lncORF-encoded peptide pTINCR is upregulated in epithelial tissue upon cellular stress and promotes epithelial differentiation in several cancers of epithelial origin, including luminal breast cancer [71], cutaneous squamous cell carcinoma, and lung adenocarcinoma. Although pTINCR peptide expression was not analyzed, correlation studies using public data revealed that high TINCR lncRNA expression is associated with more favorable prognoses in various epithelial cancers including breast cancer, suggesting the pTINCR peptide may be prognostically significant [71].

### 2.3. Glioblastoma

SHPRH-146aa is a circORF-encoded peptide that has the potential to serve as a prognostic or diagnostic biomarker for glioblastoma (GBM) [63]. The expression of SHPRH-146aa was downregulated in GBM tissue samples, and longer patient survival times were correlated with higher SHPRH-146aa expression [63]. PINT87aa, another circORF-encoded peptide, displayed similar association with clinical prognosis of GBM [64]. When comparing GBM of varying WHO grades and normal brain tissues, PINT87aa expression was significantly reduced in GBM tissues and negatively correlated with tumor grade, with WHO grade IV glioblastomas displaying the lowest PINT87aa expression [64]. Another circORF-encoded peptide, FBXW7-185aa, also exhibits reduced expression and positive correlations with survival times of GBM patients [65].

### 2.4. Hepatocellular Carcinoma (HCC)

In HCC, the lncORF-encoded peptide SMIM30 has been associated with poor patient survival rates [66]. MPM, also a lncORF-encoded peptide, is significantly downregulated in HCC tissues, with low MPM expression associated with increased HCC metastasis, recurrence, and mortality [67].

### 2.5. Ovarian Cancer (OC)

DDUP is a peptide encoded within the lncRNA CTBP1-DT and is involved in DNA damage response (DDR) signaling [69]. Ren and colleagues (2023) demonstrated that high CTBP1-DT expression confers markedly shorter overall and progression-free survival in ovarian, lung, and gastric cancer patients [69]. Notably, it has been showed that the DDUP peptide, not the CTBP1-DT lncRNA, promotes chemoresistance in patient-derived ovarian cancer cells [69].

### 2.6. Prostate Cancer

Although these results emphasize that ncORFs and their encoded peptides have potential clinical diagnostic and prognostic value in various human cancers, much of their detection occurs in the context of biopsied tissue samples. If these peptides can be detected non-invasively, such as in body fluids like blood, serum, or urine, it would immensely benefit their adoption as cancer biomarkers in clinical settings [55]. An example of such an ncORF is SHLP2, a mitochondrial-derived peptide which has demonstrated potential diagnostic significance in prostate cancer (PC) [70]. Xiao and colleagues (2017) analyzed serum SHLP2 concentrations in black and white PC patients and found SHLP2 expression to be significantly reduced in white confirmed PC cases compared to white healthy controls, while no significant difference in SHLP2 expression can be observed in black cases vs controls. A serum concentration of SHLP2 greater than 350 pg/ml is a strong predictor of negative histopathology results for prostate biopsies from men of both races. For this study sample of 100 PC patients, using a SHLP2 cut-off of 350 pg/ml or greater for PC diagnosis would have avoided 57 negative biopsies, while only missing one positive case of Gleason 7. Inclusion of this cut-off value of 350 pg/ml serum SHLP2 substantially improved accuracy of predicted PC risk in white males, while only slightly improving in black males. Overall, downregulation of SHLP2 peptide expression was associated with greater PC risk in white men, and high serum SHLP2 concentration accurately ruled out PC in both racial groups, demonstrating potential as a novel diagnostic biomarker for PC [70]. However, no association was found between SHLP2 expression and PC clinical grade, therefore SHLP2 does not confer any prognostic value [70].

### 2.7. Other Types of Cancer

miPEP133, a pri-miORF-encoded peptide, was demonstrated by Kan and colleagues (2020) to be a favorable prognostic biomarker for nasopharyngeal carcinoma (NPC) [71]. Low miPEP133 mRNA levels were associated with advanced metastatic progression in NPC and significantly worse overall survival in NPC patients [71].

Sun and colleagues demonstrated that lncRNA ASH1L-AS1 encodes a peptide called APPLE, and both the lncRNA and peptide are significantly upregulated in preliminary acute myeloid

leukemia (AML) patients when compared to healthy controls or AML patients in complete remission [17]. AML patients with high ASH1L-AS1 expression are more likely to achieve 5-year leukemia-free survival than those with low expression [17]. Interestingly, ASH1L-AS1 expression can also be used to differentiate patients in complete remission status and those with preliminary diagnosis, further supporting the status of ASH1L-AS1 and APPLE as predictors of poor AML prognosis [17].

circMAPK1, a circORF-encoded peptide, exhibited downregulated expression in gastric cancer (GC) tissue compared to paired adjacent normal tissues [32]. High circMAPK1 expression is also linked with favorable clinicopathologic parameters (e.g., tumor size, lymphatic invasion, clinical stage) as well as longer overall survival [32].

MIAC, a lncORF-encoded peptide, was demonstrated to have higher expression in renal cell carcinoma (RCC) tissue compared with normal tissues in TCGA database screening as well as clinical samples [72]. RCC patients with high MIAC expression demonstrated significantly higher survival rates than those with low expression [72]. MIAC expression was also significantly higher in patients with early-stage renal cancer than those in advanced stages [72]. This result emphasizes that MIAC lncORF has potential clinical diagnostic and prognostic value in RCC.

To identify novel prognostic and diagnostic biomarkers for cancer, proteomic approaches can be applied. For example, in a recent study, global micro-/alt-protein quantitation in two human leukemia cell lines, K562 and MOLT4, identify 12 unannotated proteins that are differentially expressed in these cell lines[78]. Further quantification of ncORF protein in clinical samples will identify novel ncORF biomarkers for the diagnosis and prognosis of human cancers.

3. ncORFs as Regulators of Cancer Hallmarks

Disruption of regulatory circuits that govern cell homeostasis enables the transformation of normal cells into malignant cancers, driven by genetic alternations [79]. Over two decades ago, Hanahan and Weinberg proposed 6 essential hallmarks acquired by the cancer cell genotypes, which were further extended to 10 hallmarks including increased proliferation, reduced cell death, loss of tumor suppressor genes, increased angiogenesis, enabling replicative immortality, activating invasion and metastasis, immune escape, inflammation, genomic instability, and dysregulated metabolism[79,80]. Previous studies have shown that ncRNAs play important roles in regulation of various hallmarks of cancer[81]. In this review, we discuss the implication and molecular mechanism underlying the role of known microproteins in different cancer hallmarks (Figure 2).

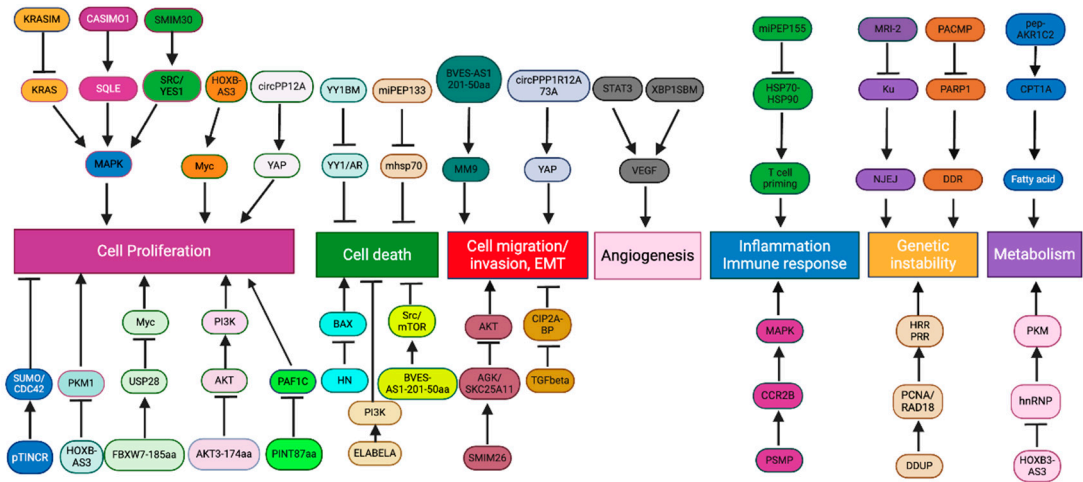


Figure 2. Regulation of cancer hallmarks by ncORF.

3.1. Cell Proliferation and Death

The development of cancer results in continual unregulated cell proliferation and inhibition of cell death [79]. Unlike normal cells, cancer cells can grow uncontrollably, invade normal tissues, and

spread throughout the body, leading to metastasis [79]. This loss of growth control is a net outcome of abnormal cell regulatory pathways [79].

### 3.1.1. Cell Proliferation

Normal cells are strictly regulated to inhibit uncontrolled proliferation [79]. However, cancer cells can grow uncontrollably by sustaining self-sufficient growth signals and evading growth suppressors [79]. Many proteins translated from ncORFs foster cell proliferation. For instance, a novel protein, Cancer-Associated Small Integral Membrane Open reading frame 1 (CASIMO1) is preferentially expressed in hormone receptor-positive breast tumors and regulates proliferation and cell cycle [82]. CASIMO1 interacts with squalene epoxidase (SQLE) to induce ERK phosphorylation [82]. Loss of SQLE mimics the effect of CASIMO1 deficiency on breast cancer cell proliferation, indicating the role of CASIMO1 in modulating the MAPK pathway through its interaction with SQLE (Figure 2)[82]. These results suggest the oncogenic role of CASIMO1 in breast cancer cells. Additionally, the proliferation of HCC cells is modulated by a lncRNA-encoded peptide, SMIM30, which functions as an adaptor protein for membrane anchoring of non-receptor tyrosine kinases, SRC/YES1 to promote downstream MAPK signaling activation [66]. Another example is HOXB-AS3, a lncRNA-encoded peptide that exhibits opposing effects in different cancer types. Overexpression of HOXB-AS3 protein enhances the proliferation and viability of oral squamous cell carcinoma cells through c-Myc mRNA stabilization [83]. Conversely, tumor growth of CRC is suppressed by HOXB-AS3 peptide, which blocks glucose metabolic reprogramming through inhibition of the splicing of PKM1 (Figure 2)[60]. A ncORF found in C20orf204 RNA encodes C20orf204-189AA, which enhances proliferation of HCC cells when overexpressed; the mechanistic details of this peptide in promoting HCC specific proliferation remain to be investigated [84]. The proliferative capability of cancer cells can also be promoted by other ncORF-encoded peptides, including ACLY-BP and circPPP1R12A-73aa [77,85]. Knockout of ACLY-BP significantly suppresses the proliferation of clear cell renal cell lines, suggesting its important role in maintaining the stability of ATP citrate lyase (ACLY), leading lipid metabolic shift to promote cell proliferation in clear cell RCC (ccRCC)(Figure 2)[85]. Increased proliferation ability was observed in colon cancer (CC) cells with circPPP1R12A-73aa overexpression, which subsequently activated YAP activity [77]. In addition, CRISPR knockout of 57 of the 553 ncORFs (10%) demonstrated growth-inhibitory effect[52].

Although many proteins translated from ncORFs often display oncogenic roles in cancers, some peptides can exhibit tumor suppressive roles. FBXW7-185aa, a novel protein encoded by the circular form of the FBXW7 gene, de-stabilizes cMyc by interacting with a de-ubiquitinating enzyme, USP28, to inhibit the proliferation and cell cycle acceleration in glioma cells (Figure 2)[65]. Two peptides encoded by circRNAs, AKT3-174aa and PINT87aa, have been discovered to suppress cell proliferation of GBM through two independent studies [64,86]. Specifically, AKT3-174aa negatively regulates RTK/PI3K pathway by blocking AKT phosphorylation at Thr308 site, and PINT87aa suppresses oncogenic transcriptional elongation through PAF1C [64,86]. The protein-coding capacity of *TINCR*, a ncRNA involved in epidermal cell differentiation, is recently identified [87]. Expression of TINCR protein significantly suppresses the tumor growth through interaction with SUMO/CDC42 complex in squamous cell carcinoma (Figure 2)[87]. KRASIM, another microprotein translated from lncRNA NCBP2-AS2, appears to reduce KRAS levels, suppressing ERK/MAPK signaling to inhibit the growth and proliferation of HCC cells (Figure 2)[88].

### 3.1.2. Cell Death

The apoptotic program can be initiated by various physiological signals, triggering a series of choreographed molecular steps involving in disruption of cellular membranes, cytoskeletal collapsing, chromatin condensation, and nuclear fragmentation [79]. The main apoptotic pathways are extrinsic and intrinsic as well as an additional T cell mediated perforin/granzyme pathway [89]. Although each pathway depends on specific activating signals, they converge on the same execution pathway triggered by caspase-3 activation [89]. Tumor cells can gain apoptotic resistance by expressing anti-apoptotic proteins (e.g., Bcl-2) or inhibiting proapoptotic proteins (e.g., Bax) [89].

Many peptides translated from ncORFs have been implicated in cell death regulation. For example, humanin (HN) inhibits apoptosis by preventing Bax translocation from cytosol to mitochondria (Figure 2)[90]. Inhibition of HN expression re-sensitizes the cells to Bax [90]. One of lncRNAs, BVES-AS1, encodes a short peptide, named BVES-AS1-201-50aa, which promotes cell viability by activating Src/mTOR signaling in CRC cells [91]. ELABELA is a 54aa peptide hormone that activates PI3K/AKT signaling via the apelin receptor to stimulate hESC growth and maintain self-renewal through inhibition of apoptotic pathways (Figure 2)[92]. Another lncRNA-encoded micropeptide regulating cell viability is YY1-blocking micropeptide (YY1BM) [93]. Under nutrient deprivation, YY1BM induces apoptosis by inhibiting the interaction of YY1 and androgen receptor in male esophageal squamous cell carcinoma (Figure 2)[93]. As previously described, pri-miRNAs, precursors of miRNAs, have been found to harbor ncORFs that encode functional peptides [28]. Microproteins encoded by these ncORFs are termed miRNA-encoded peptides (miPEPs). A novel miPEP, miPEP133, has demonstrated its tumor suppressive ability [71]. Overexpression of miPEP133 induces cell cycle arrest and apoptosis by blocking the interaction between mitochondrial heat shock protein 70 and its binding partner (Figure 2)[71].

### 3.2. Metastasis

#### 3.2.1. Cell Migration and Invasion

Metastases account for 90% of cancer-related mortality [94]. The ability of a tumor to colonize foreign environments enhances its survival, especially when the nutrients and space become limited at the primary site [79]. Cell migration and invasion are initial steps during metastatic dissemination involving the activation of cellular pathways that modulate the dynamics of cytoskeletons in tumor cells [95]. The epithelial-to-mesenchymal transition (EMT) is an important cellular program activated by cancer cells to acquire invasive phenotypes during metastatic event [96]. Loss of adherent junctions, acquisition of fibroblastic morphology, and expression of matrix-degrading proteases are common traits established during EMT [96]. A reversal of EMT traits occurs to recapitulate epithelial characteristics for colonization at a distant secondary place [96]. This process is often termed as mesenchymal to epithelial transition (MET) [96]. Regulation of F-actin dynamic is essential for cell motility, and loss of CASMO1 impacts actin assembly, dramatically reducing the migration ability of BC cells [82]. Moreover, two lncRNA-encoded peptides, SMIM30 and ZFAS1, independently promote cell migration and invasion of HCC cells [66,97], and their involvements of HCC metastatic events in vivo needs further exploration. BVES-AS1-201-50aa is one of many lncRNA-encoded proteins involved in tumor progression [91]. This 50-aa microprotein enhances migratory and invasive abilities of CRC cells in vitro [91]. Overexpression of BVES-AS1-201-50aa is thought to increase the production of metastasis-associated protein MM9 in CRC cells [91]. Through ribosome profiling and RNA-seq, a conserved and secreted peptide, Toddler also known as APELA, is found to be expressed by a noncoding RNA in zebrafish, modulating cell movements during gastrulation [98,99]. Later research reveals that APELA is upregulated in ovarian cancer, contributing to cell migration and cancer progression [99]. Other than lncRNA-encoded proteins, circRNA is another type of ncRNAs containing ncORFs. One small peptide, circPPP1R12A-73aa, is expressed by such ncORFs from circPPP1R12A and displays metastasis-promoting abilities in CC by activating YAP activity [77].

A lncRNA-encoded microprotein, SMIM26, behaves as a tumor suppressor by inhibiting EMT in ccRCC [100]. Overexpression of SMIM26 leads to reduced metastatic phenotypes, while silencing SMIM26 greatly enhances the migration and invasion of RCC cells [100]. Meng et al. (2023) reveal that SMIM26 interacts with AGK, a mitochondrial acylglycerol kinase, as well as glutathione transport regulator SKC25A11 to interfere with AKT signaling-mediated metastasis (Figure 2)[100]. The role of TGF- $\beta$  signal transduction is well recognized in cancer metastasis, and induced EMT phenotypes and cancer stemness can be achieved by TGF- $\beta$  treatment in breast cancer [18]. CIP2A-BP is a micropeptide encoded by lncRNA and inhibits metastatic capabilities when overexpressed in TNBC cells [18]. Expression of CIP2A-BP is usually reduced in TNBC by TGF- $\beta$  signaling to induce

tumor invasion and metastasis (Figure 2)[18]. Cell migration and invasion can also be inhibited by a previously discussed micropeptide, miPEP133, though its anti-metastatic role in vivo has not been fully illustrated [71].

### 3.2.2. Angiogenesis

Without blood supplies, nutrients and oxygen become depleted quickly, limiting the tumor from growing beyond a size of more than 1 mm in diameter and subsequently forcing it to enter a dormant state [101]. To sustain sufficient growth support, cancer cells promote vascularization by inducing angiogenesis, another hallmark for tumor development [101]. Normally, angiogenesis occurs during wound healing [101]. A fine-tuned balance of angiogenic activators and inhibitors controls the production of new blood vessels [101]. However, tumor cells can induce the angiogenic switch by tipping this balance to the pro-angiogenic state [101]. Compared to normal blood vessels, tumor vasculature is organized chaotically, resulting in abnormal blood flow and leaky vessels, which increase the likelihood of distant metastases [101]. Two lncRNA-encoded microproteins display opposing roles during angiogenic development in TNBC, including ASRPS and XBP1SBM [16,102]. ASRPS inhibits angiogenesis by regulating VEGF pathway through its interaction with STAT3, whereas XBP1SBM can upregulate VEGF expression level, promoting angiogenesis in TNBC [16,102].

### 3.3. Inflammation and Immune Responses

Tumor cells express specific antigens that are recognized by immune cells as non-self and targeted for destruction [103]. Antigen-presenting cells (APC) such as dendritic cells (DCs) recognize and display tumor-specific antigens, which elicit T-cell responses to destroy tumor cells [103]. However, tumor cells exploit various mechanisms to evade immune surveillance, such as downregulating antigens or inducing an immunosuppressive microenvironment [103]. A common strategy is to induce immune tolerance to cancer by expressing the ligands for the receptor of the Ig superfamily, PD-1 [103].

Advancements in proteomics technologies allow the identification of peptides or proteins derived from ncORFs. Several ncORF-encoded microproteins contribute to immune responses. For example, the identification of non-annotated ORFs led to the discovery of PC3-secreted microprotein (PSMP), a chemoattractant protein highly expressed in some prostate cancers. PSMP regulates inflammation during tumor development through binding to CCR2B receptor and activate downstream ERK/MAPK signaling [104]. Moreover, protein-coding ORFs within lncRNA Aw112010 express proteins essential for mucosal immunity [105]. Another micropeptide encoded by a non-protein coding transcript is miPEP155, which suppresses DC-driven auto inflammation and T cell priming by regulating antigen presentation through disrupting HSP70-HSP90 interaction [106]. Deep mining of non-canonical proteins also results in the discovery of tumor-specific peptides, highlighting their potential for establishing cancer immunotherapies [107,108].

### 3.4. DNA Damage Response and Genetic Instability

Genomic integrity relies on DNA monitoring and repair mechanisms to ensure a pristine state of the whole genome [79]. Genomic DNA is sensitive to exogenous and endogenous damage, resulting in genetic instability [68]. Diverse DNA damage response mechanisms have evolved to sustain the genome integrity, including canonical homologous recombination repair (HRR), non-homologous end-joining (NHEJ) pathways, and the post-replication repair (PRR) pathway [68]. Yet tumor cells accumulate substantial genomic alternations that allow them to acquire survival advantages during carcinogenesis [79]. Defects of DNA repair mechanisms explain the high mutational rate found in cancer cells [79]. Genomic instability enabled by malfunctions of DNA repair machinery then promotes clonal diversity and selection [79]. Known microproteins regulating genome stability include MRI-2, which interacts with DNA end-binding protein Ku in nucleus to stimulate NHEJ DNA repair pathway [109]. PACMP and DDUP are reported lncRNA-encoded microproteins that also trigger DNA damage responses through different mechanisms [68,110].

PACMP modulates DDR by preventing degradation of CTBP-interacting protein and enhancing PARP1-dependent poly(ADP-ribosyl)ation [110]. DNA damage induces upregulation of DDUP, which regulates PCNA ubiquitination and RAD18 retention at DNA damage sites, inducing HRR and PRR mechanisms [68]. The ubiquitin-like microprotein, TINCR, is reported to be a protective factor preventing UV exposure as accelerated skin lesions were observed in UV-treated mice with TINCR mutant [87].

### 3.5. Metabolism

Reprogramming of cellular metabolism is another core hallmark of cancer development, supporting the acquisition and maintenance of malignant phenotypes [111]. Anabolic growth in nutrient-depleted conditions is made possible by altering metabolic activity in cancer cells for survival [111]. The Warburg effect or aerobic glycolysis is a classic example of metabolic reprogramming found in cancer [111]. In the 1920s, Warburg discovered that cancer cells constitutively uptake glucose and convert it into lactate regardless of the presence of oxygen [111]. Elevated glycolytic intermediates are utilized as macromolecule precursors to fulfill the demands of proliferating cells [111]. Additionally, intermediates from the tricarboxylic acid cycle can be supplied into subsidiary pathways for macromolecule synthesis [111]. Thus, the ability of cancer cells to reprogram metabolic pathways supports tumorigenic proliferation and progression. The selective expression of glycolytic enzyme pyruvate kinase M (PKM) isoforms is regulated by splicing factors, such as hnRNP A1 [60]. Re-expression of PKM2 in human cancers is thought to promote aerobic glycolysis, enabling proliferative advantages [60]. A small peptide encoded by the lncRNA *HOXB-AS3* suppresses glucose reprogramming by antagonizing hnRNP A1-mediated PKM splicing in CRC (Figure 2) [60]. Lactate utilization is restricted by MP31, a microprotein encoded from 5' UTR of PTEN in GBM [112]. Loss of MP31 increased global lactate metabolism, enhancing tumorigenicity [112].

Abnormal lipid oxidation is prevalent during tumor development [113,114]. In addition to glucose, tumor cells adapt to nutrient-deprived microenvironment by increasing fatty acid metabolism [113]. A key enzyme of fatty acid oxidation (FAO), carnitine palmitoyl transferase 1A (CPT1A), is upregulated by an exosomal lncAKR1C2-encoded microprotein, pep-AKR1C2, inducing a metabolic switch towards FAO in gastric cancer [113,115]. Lipid metabolism and homeostasis can be influenced by microproteins, including CASIMO1 and ACLY-BP, through distinct mechanisms, enhancing proliferation in BC and ccRCC, respectively [82,85].

## 4. ncORF in Cancer Therapy

Given the fact that ncORFs or microproteins play important roles in various human diseases including cancer, their roles as therapeutic targets have been proposed or demonstrated recently [116]. Targeted therapy is a high precision treatment designed to attack certain molecules that play oncogenic roles in tumors [55]. Its goal is to minimize non-specific cytotoxicity while destroying the growth and survival of cancer cells [55]. Molecular targeted therapeutic agents act on growth factors, cell surface antigens, and signaling pathways that initiate and maintain cancer hallmarks to block cancer progression [117]. Many anti-cancer therapies target different anti-apoptotic components, such as Bcl-2 family proteins, to induce apoptosis in tumor cells [117]. Molecular targeted therapies have great advantages in treating different types of cancers, such as breast, gastric, colon, and lung cancers [117]. Compared to conventional chemotherapy, inhibitors of EGFR-tyrosine kinase improve clinical outcomes in NSCLC patients with EGFR mutations [117]. Despite the observed advantages, limitations exist and restrict the efficiency of targeted therapy in cancer patients [117]. Since targeted treatment requires specific biomarkers to be effective, patients who do not express such target will respond poorly [117]. Several treatment options have been proposed for cancer [115]. Cancer cells are enriched with ncORF MHC-I associated peptides (MAPs), suggesting their potential as a source of tumor antigens for immunotherapeutic applications. Additionally, pharmacological targeting of cancer-specific ncRNAs to silence pro-oncogenic miniprotein translation could lead to downstream inhibition of previously undruggable targets. Miniproteins can also provide valuable insights for drug design by identifying protein-protein

interaction hotspots or inhibitor binding sites. Exploring miniprotein conjugation to other drugs may offer potential applications in molecular glues or heterobifunctional compounds. Miniproteins are also promising candidates as non-antibody binding scaffolds. Furthermore, the differential expression of ncORFs and miniproteins in cancer enhances their potential as biomarkers, particularly in cases involving differentially expressed secreted miniproteins.

Drug resistance occurs when cancer cells tolerate a treatment, greatly limiting the efficiency of anti-tumor therapies in many patients [117,118]. Genomic instability is one factor contributing to drug resistance, which can be intrinsic or acquired [118]. Innate resistance exists in patients even before administering anti-cancer drugs due to pre-existing genetic perturbations, tumor heterogeneity, or activation of defense mechanisms [118]. For example, resistance to cisplatin treatment was found in gastric cancer patients with HER2 overexpression, which upregulated the expression of EMT-related transcriptional factor Snail [119]. Moreover, studies show that Snail and Slug not only mediate EMT but also promote stem-like characteristics, conferring resistance to radiotherapy and chemotherapy [119,120]. Alternatively, relapse to treatment can occur when new mutations are acquired in tumor cells, resulting in a gradual decrease in therapy efficacy [118]. For instance, a point mutation from threonine-to-isoleucine in BCR-ABL kinase domain can cause resistance or relapse of tyrosine kinase inhibitor treatment in patients with chronic myelogenous leukemia [121]. Several microproteins have been shown to affect drug efficiency in cancers. Radiation sensitivity was induced by overexpressing AKT3-174aa in GBM cells, while loss of AKT3-174aa expression regained resistance to radiation [86]. Chemoresistance can be acquired by evading DNA damage-induced genetic instability [118]. DDUP is a lncRNA-encoded microprotein and functions as a regulator of DDR, promoting resistance to DNA-damaging chemotherapies [68]. Inhibition of DDUP by ATR inhibitor Berzosertib re-sensitized ovarian cancer cells to cisplatin treatment, indicating its vital role in affecting drug response [68]. The tumor-suppressive role of a microprotein, N1DARP, was recognized in pancreatic cancer since it could suppress Notch1 signaling pathways, which contribute to chemoresistance [122].

## 5. Conclusion and Future Perspective

The recognition of ncORFs has attracted increasing attention and uncovers new aspects of regulating biological processes. The protein-coding capacity of human genome is greatly underestimated and remains to be fully explored. Improved approaches prompt the identification of many ncORF-encoded peptides involving in cancer pathogenesis. Even though increasing number of proteins expressed from ncORFs are unveiled, only a small set of these peptides are functionally characterized. Regulatory functions of many microproteins translated from ncRNAs are reported to influence tumor progression. Lots of ncRNAs-derived microproteins can act as tumor suppressors or inducers. Thus, understanding the role of these ncRNA-encoded proteins offers critical prospects for cancer therapies.

In this review, we describe different types of ncRNAs reported to have coding potential, mention present techniques used for capturing ncORFs, and summarize functional proteins encoded by ncORFs implicated in cancer research. The discovery of ncORF-encoded microproteins expands our understanding of tumor development as many of them display prognostic and/or carcinogenic roles in human cancers. Despite more and more studies have focused on ncORFs, only a small portion of ncRNA-encoded microproteins has been identified and validated, and their mechanisms of action in cancer remain mostly unclear. Therefore, identifying and characterizing noncanonical proteins may lead to the establishment of novel biomarkers and drug targets for cancer diagnosis and therapeutics.

**Author Contributions:** Conceptualization, A.G. and X.Y.; writing—original draft preparation, A.G. and C.C.; writing—review and editing, A.G. and X.Y.; visualization, A.G. and C.C.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the Canadian Institute of Health Research (CIHR), grant numbers 186142 and 148629.

**Institutional Review Board Statement:** Not applicable

**Informed Consent Statement:** Not applicable

**Data Availability Statement:** Not applicable

**Acknowledgments:** Not applicable

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Liu, Y., S. Zeng, and M. Wu, *Novel insights into noncanonical open reading frames in cancer*. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 2022. **1877**.
2. Consortium, I.H.G.S., *Finishing the euchromatic sequence of the human genome*. Nature, 2004. **431**: p. 931-945.
3. Wright, B.W., et al., *The dark proteome: translation from noncanonical open reading frames*. Trends Cell Biol, 2022. **32**: p. 243-258.
4. Hofman, D.A., et al., *Translation of non-canonical open reading frames as a cancer cell survival mechanism in childhood medulloblastoma*. Mol Cell., 2024. **84**: p. 261-276.
5. Ferreira, J.P., et al., *Engineering ribosomal leaky scanning and upstream open reading frames for precise control of protein translation*. Bioengineered. , 2014. **5**: p. 186-192.
6. Samandi, S., et al., *Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins*. Elife 2017. **6**: p. e27860.
7. Jayaram, D.R., et al., *Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of PKCs*. Proc Natl Acad Sci U S A., 2021. **118**: p. e2018899118.
8. Wu, Q., et al., *Translation of small downstream ORFs enhances translation of canonical main open reading frames*. EMBO J., 2020. **39**: p. e104763.
9. Brunet, M.A., et al., *The FUS gene is dual-coding with both proteins contributing to FUS-mediated toxicity*. EMBO Rep. , 2021. **22**: p. e50640.
10. Valdivia-Francia, F. and A. Sandoel, *No country for old methods: New tools for studying microproteins*. iScience, 2024. **27**: p. 108972.
11. Hann, S.R., et al., *A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas*. Cell, 1988. **52**: p. 185-195.
12. Mattick, J.S., et al., *Long non-coding RNAs: definitions, functions, challenges and recommendations*. Nat Rev Mol Cell Biol., 2023. **24**: p. 430-447.
13. Statello, L., et al., *Gene regulation by long non-coding RNAs and its biological functions*. Nat Rev Mol Cell Biol., 2021. **22**: p. 96-118.
14. Patraquim, P., et al., *Translation and natural selection of micropeptides from long non-canonical RNAs*. Nat Commun, 2022. **13**: p. 6515.
15. Ge, Q., et al., *Micropeptide ASAP encoded by LINC00467 promotes colorectal cancer progression by directly modulating ATP synthase activity*. J Clin Invest. , 2021. **131**: p. e152911.
16. Wang, Y., et al., *LncRNA-encoded polypeptide ASRPS inhibits triple-negative breast cancer angiogenesis*. J Exp Med., 2020. **217**: p. jem.20190950.
17. Sun, L., et al., *The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation*. Mol Cell. , 2021. **81**: p. 4493-4508.
18. Guo, B., et al., *Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression*. EMBO J. , 2020. **39**: p. e102190.
19. Ho, L., et al., *ELABELA Is an Endogenous Growth Factor that Sustains hESC Self-Renewal via the PI3K/AKT Pathway*. Cell Stem Cell. , 2015. **17**: p. 435-447.
20. Li, M., et al., *Micropeptide MIAC Inhibits HNSCC Progression by Interacting with Aquaporin 2*. J. Am. Chem. Soc., 2020. **142**: p. 6708-6716.
21. Lanz, R.B., et al., *A Steroid Receptor Coactivator, SRA, Functions as an RNA and Is Present in an SRC-1 Complex*. Cell, 1999. **97**: p. 17-27.
22. Yan, Y., et al., *The steroid receptor RNA activator protein (SRAP) controls cancer cell migration/motility*. FEBS Lett., 2015. **589**: p. 4010-4018.
23. Cheetham, S.W., G.J. Faulkner, and M.E. Dinger, *Overcoming challenges and dogmas to understand the functions of pseudogenes*. Nat Rev Genet., 2020. **21**: p. 191-201.
24. Ji, Z., et al., *Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins*. eLife, 2015. **4**: p. e08890.

25. Ransohoff, J.D., Y. Wei, and P.A. Khavari, *The functions and unique features of long intergenic non-coding RNA*. Nat Rev Mol Cell Biol., 2018. **19**: p. 143-157.
26. Apcher, S., et al., *Translation of pre-spliced RNAs in the nuclear compartment generates peptides for the MHC class I pathway*. Proc Natl Acad Sci U S A. , 2013. **110**: p. 17951-17956.
27. Ratti, M., et al., *MicroRNAs (miRNAs) and Long Non-Coding RNAs (lncRNAs) as New Tools for Cancer Therapy: First Steps from Bench to Bedside*. Target Oncol., 2020. **15**: p. 261-278.
28. Motameny, S., et al., *Next Generation Sequencing of miRNAs - Strategies, Resources and Methods*. Genes (Basel). , 2010. **1**: p. 70-84.
29. Chen, R., et al., *Engineering circular RNA for enhanced protein production*. Nat Biotechnol., 2023. **41**: p. 262-272.
30. Misir, S., N. Wu, and B.B. Yang, *Specific expression and functions of circular RNAs*. Cell Death Differ., 2022. **29**: p. 481-491.
31. Yang, Y., et al., *Extensive translation of circular RNAs driven by N(6)-methyladenosine*. Cell Res. , 2017. **27**: p. 626-641.
32. Jiang, T., et al., *A novel protein encoded by circMAPK1 inhibits progression of gastric cancer by suppressing activation of MAPK signaling*. Mol Cancer. , 2021. **20**: p. 66.
33. Merry, T.L., et al., *Mitochondrial-derived peptides in energy metabolism*. Am J Physiol Endocrinol Metab., 2020. **319**: p. E659-E666.
34. Yang, H., et al., *Widespread stable noncanonical peptides identified by integrated analyses of ribosome profiling and ORF features*. Nat Commun, 2024. **15**: p. 1932.
35. Martinez, T.F., et al., *Accurate annotation of human protein-coding small open reading frames*. Nat Chem Biol., 2020. **16**: p. 458-468.
36. Leong, A.Z.-X., et al., *Short open reading frames (sORFs) and microproteins: an update on their identification and validation measures*. J Biomed Sci., 2022. **29**: p. 19.
37. Hanada, K., et al., *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**: p. 632-640.
38. Brito-Estrada, O., K.R. Hassel, and C.A. Makarewich, *An Integrated Approach for Microprotein Identification and Sequence Analysis*. J Vis Exp., 2022.
39. Zhu, M. and M. Gribskov, *MiPepid: MicroPeptide identification tool using machine learning*. BMC Bioinformatics, 2019. **20**: p. 559.
40. Skarszewski, A., et al., *uPEPPERoni: An online tool for upstream open reading frame location and analysis of transcript conservation*. BMC Bioinformatics, 2014. **15**: p. 36.
41. Lin, M.F., I. Jungreis, and M. Kellis, *PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions*. Bioinformatics, 2011. **27**: p. i275-282.
42. Ingolia, N.T., et al., *Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling*. Science, 2009. **324**: p. 218-223.
43. Giacomini, G., et al., *A Neural Network Approach for the Analysis of Reproducible Ribo-Seq Profile*. Algorithms, 2022. **15**: p. 274.
44. Kearse, M.G. and J.E. Wilusz, *Non-AUG translation: a new start for protein synthesis in eukaryotes*. Genes Dev., 2017. **31**: p. 1717-1731.
45. Slavoff, S.A., et al., *Peptidomic discovery of short open reading frame-encoded peptides in human cells*. Nat Chem Biol., 2013. **9**: p. 59-64.
46. Guttman, M., et al., *Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins*. Cell, 2013. **154**: p. 240-251.
47. Brar, G.A. and J.S. Weissman, *Ribosome profiling reveals the what, when, where and how of protein synthesis*. Nat Rev Mol Cell Biol., 2015. **16**: p. 651-664.
48. Aspden, J.L., et al., *Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq*. Elife, 2014. **3**: p. e03528.
49. Mohsen, J.J., A.A. Martel, and S.A. Slavoff, *Microproteins-Discovery, structure, and function*. Proteomics, 2023. **23**(23-24): p. e2100211.
50. López, E., et al., *Functional phosphoproteomic mass spectrometry-based approaches*. Clin Transl Med., 2012. **1**: p. 20.
51. Chen, J., et al., *Pervasive functional translation of noncanonical human open reading frames*. Science, 2020. **367**: p. 1140-1146.

52. Prensner, J.R., et al., *Noncanonical open reading frames encode functional proteins essential for cancer cell survival*. Nat Biotechnol, 2021. **39**(6): p. 697-704.
53. Schlesinger, D. and S.J. Elsässer, *Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins*. FEBS J., 2022. **289**: p. 53-74.
54. Hofman, D.A., et al., *Translation of non-canonical open reading frames as a cancer cell survival mechanism in childhood medulloblastoma*. Mol Cell, 2024. **84**(2): p. 261-276.e18.
55. Zhu, S., et al., *Peptides/Proteins Encoded by Non-coding RNA: A Novel Resource Bank for Drug Targets and Biomarkers*. Front Pharmacol., 2018. **9**: p. 1295.
56. Erady, C., et al., *Pan-cancer analysis of transcripts encoding novel open-reading frames (nORFs) and their potential biological functions*. NPJ Genom Med., 2021. **6**: p. 4.
57. Zhou, H., et al., *Micropeptides: potential treatment strategies for cancer*. Cancer Cell International. Cancer Cell Int., 2024. **24**: p. 134.
58. Setrerrahmane, S., et al., *Cancer-related micropeptides encoded by ncRNAs: Promising drug targets and prognostic biomarkers*. Cancer Lett., 2022. **547**: p. 215723.
59. Carlomagno, N., et al., *Diagnostic, Predictive, Prognostic, and Therapeutic Molecular Biomarkers in Third Millennium: A Breakthrough in Gastric Cancer*. Biomed Res Int., 2017. **2017**: p. 7869802.
60. Huang, J.-Z., et al., *A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth*. Mol Cell., 2017. **68**: p. 171-184.
61. Meng, N., et al., *Small Protein Hidden in lncRNA LOC90024 Promotes "Cancerous" RNA Splicing and Tumorigenesis*. Adv Sci (Weinh). 2020. **7**: p. 1903233.
62. Zhu, S., et al., *An oncopeptide regulates m6A recognition by the m6A reader IGF2BP1 and tumorigenesis*. Nat Commun, 2020. **11**: p. 1685.
63. Zhang, M., et al., *A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis*. Oncogene, 2018. **37**: p. 1805-1814.
64. Zhang, M., et al., *A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma*. Nat Commun, 2018. **9**: p. 4475.
65. Yang, Y., et al., *Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis*. J Natl Cancer Inst., 2018. **110**: p. 304-315.
66. Pang, Y., et al., *Peptide SMIM30 promotes HCC development by inducing SRC/YES1 membrane anchoring and MAPK pathway activation*. J Hepatol., 2020. **73**: p. 1155-1169.
67. Xiao, M.-H., et al., *Downregulation of a mitochondrial micropeptide, MPM, promotes hepatoma metastasis by enhancing mitochondrial complex I activity*. Mol Ther., 2022. **30**: p. 714-725.
68. Yu, R., et al., *lncRNA CTBP1-DT-encoded microprotein DDUP sustains DNA damage response signalling to trigger dual DNA repair mechanisms*. Nucleic Acids Research, 2022. **50**: p. 8060-8079.
69. Ren, L., et al., *The DDUP protein encoded by the DNA damage-induced CTBP1-DT lncRNA confers cisplatin resistance in ovarian cancer*. Cell Death Dis., 2023. **14**: p. 568.
70. Xiao, J., et al., *Low circulating levels of the mitochondrial-peptide hormone SHLP2: novel biomarker for prostate cancer risk*. Oncotarget., 2017. **8**: p. 94900-94909.
71. Kang, M., et al., *Identification of miPEP133 as a novel tumor-suppressor microprotein encoded by miR-34a pri-miRNA*. Mol Cancer., 2020. **19**: p. 143.
72. Li, M., et al., *Micropeptide MIAC inhibits the tumor progression by interacting with AQP2 and inhibiting EREG/EGFR signaling in renal cell carcinoma*. Mol Cancer., 2022. **21**: p. 181.
73. Xu, S., et al., *lncRNA HOXB-AS3 promotes growth, invasion and migration of epithelial ovarian cancer by altering glycolysis*. Life Sci., 2021. **264**: p. 118636.
74. Jiang, W., et al., *lncRNA HOXB-AS3 exacerbates proliferation, migration, and invasion of lung cancer via activating the PI3K-AKT pathway*. J Cell Physiol., 2020. **235**: p. 7194-7203.
75. Papaioannou, D., et al., *The long non-coding RNA HOXB-AS3 regulates ribosomal RNA transcription in NPM1-mutated acute myeloid leukemia*. Nat Commun, 2019. **10**: p. 5351.
76. Xing, Y., et al., *Long non-coding RNA (lncRNA) HOXB-AS3 promotes cell proliferation and inhibits apoptosis by regulating ADAM9 expression through targeting miR-498-5p in endometrial carcinoma*. J Int Med Res., 2021. **49**: p. 03000605211013548.
77. Zheng, X., 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**: p. 47.

78. Cao, X., et al., *Comparative Proteomic Profiling of Unannotated Microproteins and Alternative Proteins in Human Cell Lines*. J Proteome Res, 2020. **19**(8): p. 3418-3426.
79. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. Cell, 2000. **100**: p. 57-70.
80. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. Cell, 2011. **144**(5): p. 646-74.
81. Merino-Valverde, I., E. Greco, and M. Abad, *The microproteome of cancer: From invisibility to relevance*. Exp Cell Res, 2020. **392**(1): p. 111997.
82. Polycarpou-Schwarz, M., et al., *The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation*. Oncogene, 2018. **37**: p. 4750-4768.
83. Leng, F., et al., *A micro-peptide encoded by HOXB-AS3 promotes the proliferation and viability of oral squamous cell carcinoma cell lines by directly binding with IGF2BP2 to stabilize c-Myc*. Oncol Lett. , 2021. **22**: p. 697.
84. Lara, S.B.D., et al., *C20orf204, a hepatocellular carcinoma-specific protein interacts with nucleolin and promotes cell proliferation*. Oncogenesis, 2021. **10**: p. 31.
85. Zhang, S., et al., *LncRNA-Encoded Micropeptide ACLY-BP Drives Lipid Deposition and Cell Proliferation in Clear Cell Renal Cell Carcinoma via Maintenance of ACLY Acetylation*. Mol Cancer Res., 2023. **21**: p. 1064-1078.
86. Xia, X., 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**: p. 131.
87. Morgado-Palacin, L., et al., *The TINCR ubiquitin-like microprotein is a tumor suppressor in squamous cell carcinoma*. Nat Commun, 2023. **14**: p. 1328.
88. Xu, W., et al., *Ribosome profiling analysis identified a KRAS-interacting microprotein that represses oncogenic signaling in hepatocellular carcinoma cells*. Sci China Life Sci., 2020. **63**: p. 529-542.
89. Vitale, I., et al., *Apoptotic cell death in disease - current understanding of the NCCD 2023*. Cell Death Differ., 2023. **30**: p. 1097-1154.
90. Guo, B., et al., *Humanin peptide suppresses apoptosis by interfering with Bax activation*. Nature 2003. **423**: p. 456-461.
91. Zhang, W., et al., *Peptide encoded by lncRNA BVES-AS1 promotes cell viability, migration, and invasion in colorectal cancer cells via the SRC/mTOR signaling pathway*. PLoS One, 2023. **18**: p. e0287133.
92. Ho, L., et al., *ELABELA Is an Endogenous Growth Factor that Sustains hESC Self-Renewal via the PI3K/AKT Pathway*. Cell Stem Cell, 2015. **17**(4): p. 435-47.
93. Wu, S., et al., *A Novel Micropeptide Encoded by Y-Linked LINC00278 Links Cigarette Smoking and AR Signaling in Male Esophageal Squamous Cell Carcinoma*. Cancer Res., 2020. **80**: p. 2790-2803.
94. Fares, J., et al., *Molecular principles of metastasis: a hallmark of cancer revisited*. Sig Transduct Target Ther. , 2020. **5**: p. 28.
95. Friedl, P. and S. Alexander, *Cancer invasion and the microenvironment: plasticity and reciprocity*. Cell, 2011. **147**: p. 992-1009.
96. Seyfried, T.N. and L.C. Huysentruyt, *On the origin of cancer metastasis*. Crit Rev Oncog., 2013. **18**: p. 43-73.
97. Guo, Z.-W., et al., *Translated Long Non-Coding Ribonucleic Acid ZFAS1 Promotes Cancer Cell Migration by Elevating Reactive Oxygen Species Production in Hepatocellular Carcinoma*. Front Genet., 2019. **10**: p. 1111.
98. Pauli, A., et al., *Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors*. Science, 2014. **343**: p. 1248636.
99. Yi, Y., et al., *APELA promotes tumor growth and cell migration in ovarian cancer in a p53-dependent manner*. Gynecol Oncol., 2017. **147**: p. 663-671.
100. Meng, K., et al., *LINC00493-encoded microprotein SMIM26 exerts anti-metastatic activity in renal cell carcinoma*. EMBO Rep., 2023. **24**: p. e56282.
101. Geiger, T.R. and D.S. Peeper, *Metastasis mechanisms*. Biochim Biophys Acta., 2009. **1796**: p. 293-308.
102. Wu, S., et al., *A micropeptide XBP1SBM encoded by lncRNA promotes angiogenesis and metastasis of TNBC via XBP1s pathway*. Oncogene, 2022. **41**: p. 2163-2172.
103. Adam, J.K., B. Odhav, and K.D. Bhoola, *Immune responses in cancer*. Pharmacol Ther., 2003. **99**: p. 113-132.
104. Pei, X., et al., *PC3-Secreted Microprotein Is a Novel Chemoattractant Protein and Functions as a High-Affinity Ligand for CC Chemokine Receptor 2*. J Immunol., 2014. **192**: p. 1878-1886.
105. Jackson, R., et al., *The translation of non-canonical open reading frames controls mucosal immunity*. Nature 2018. **564**: p. 434-438.
106. Niu, L., et al., *A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation*. Sci Adv., 2020. **6**: p. eaaz2059.

107. Chong, C., et al., *Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes*. Nat Commun, 2020. **11**: p. 1293.
108. Irajizad, E., et al., *A Comprehensive Search of Non-Canonical Proteins in Non-Small Cell Lung Cancer and Their Impact on the Immune Response*. Int. J. Mol. Sci., 2022. **23**: p. 8933.
109. Slavoff, S., et al., *A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining*. J Biol Chem., 2014. **289**: p. 10950-10957.
110. Zhang, C., et al., *Micropeptide PACMP inhibition elicits synthetic lethal effects by decreasing CtIP and poly(ADP-ribosylation)*. Mol Cell., 2022. **82**: p. 1297-1312.
111. Deberardinis, R. and N. Chandel, *Fundamentals of cancer metabolism*. Sci Adv., 2016. **2**: p. e1600200.
112. Huang, N., et al., *An Upstream Open Reading Frame in Phosphatase and Tensin Homolog Encodes a Circuit Breaker of Lactate Metabolism*. Cell Metab., 2021. **33**: p. 128-144.
113. Zhu, K.-G., et al., *The microprotein encoded by exosomal lncAKR1C2 promotes gastric cancer lymph node metastasis by regulating fatty acid metabolism*. Cell Death Dis., 2023. **14**: p. 708.
114. Carracedo, A., L. Cantley, and P. Pandolfi, *Cancer metabolism: fatty acid oxidation in the limelight*. Nat Rev Cancer., 2013. **13**: p. 227-232.
115. Posner, Z., I. Yannuzzi, and J.R. Prensner, *Shining a light on the dark proteome: Non-canonical open reading frames and their encoded miniproteins as a new frontier in cancer biology*. Protein Sci, 2023. **32**(8): p. e4708.
116. Hassel, K.R., O. Brito-Estrada, and C.A. Makarewich, *Microproteins: Overlooked regulators of physiology and disease*. iScience, 2023. **26**(6): p. 106781.
117. Lee, Y.T., Y.J. Tan, and C.E. Oon, *Molecular targeted therapy: treating cancer with specificity*. Eur. J. Pharmacol., 2018. **834**: p. 188-196.
118. Wang, X., H. Zhang, and X. Chen, *Drug resistance and combating drug resistance in cancer*. Cancer Drug Resist., 2019. **2**: p. 141-160.
119. Huang, D., et al., *Cisplatin resistance in gastric cancer cells is associated with HER2 upregulation-induced epithelial-mesenchymal transition*. Sci. Rep., 2016. **6**: p. 20502.
120. Kurrey, N.K., et al., *Snail and Slug mediate radioresistance and chemoresistance by antagonizing p53-mediated apoptosis and acquiring a stem-like phenotype in ovarian cancer cells*. Stem Cells., 2009. **27**: p. 2059-2068.
121. Quintás-Cardama, A., H.M. Kantarjian, and J.E. Cortes, *Mechanisms of primary and secondary resistance to imatinib in chronic myeloid leukemia*. Cancer Control., 2009. **16**: p. 122-131.
122. Zhai, S., et al., *A microprotein N1DARF encoded by LINC00261 promotes Notch1 intracellular domain (N1ICD) degradation via disrupting USP10-N1ICD interaction to inhibit chemoresistance in Notch1-hyperactivated pancreatic cancer*. Cell Discov., 2023. **9**: p. 95.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.