

非编码 RNA 来源的小肽：“微不足道”却“功能强大”*

陈晓彤, 赵文龙, 孙林玉, 王文涛, 陈月琴

中山大学生命科学学院, 广东 广州 510275

摘要: 非编码 RNA(ncRNA, non-coding RNA)长久以来被认为不具有编码能力。近年来随着研究技术和生物信息学工具的迅速发展, 研究发现在基因组的非编码区域上存在大量小开放阅读框(sORFs, small/short open reading frames), 其翻译产物被称作小 ORF 编码肽(SEPs, sORF encoded peptides)或小肽(micropeptides)。部分小肽被证实存在于细胞内稳定存在并独立于其来源 RNA 发挥重要作用。本文系统总结了非编码 RNA 来源小肽的鉴定方法、可编码小肽的 RNA 类型以及其研究困难和瓶颈, 并重点回顾了疾病和植物中发现的功能小肽, 以期对小肽的筛选鉴定提供思考, 对小肽作为药物研发或者农作物增产的关键靶点提供新的思路 and 方向。

关键词: 非编码 RNA; 小肽; 非经典翻译; 鉴定方法; 调控机制

中图分类号: Q71 **文献标志码:** A **文章编号:** 2097-0137(2023)03-0001-13

Micropeptides derived from non-coding RNAs: Tiny but powerful

CHEN Xiaotong, ZHAO Wenlong, SUN Linyu, WANG Wentao, CHEN Yueqin

School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

Abstract: It was long presumed that non-coding RNAs (ncRNAs) are lacking in protein-coding potential. However, recent advances in technology and tools have led to an important finding that a number of small open reading frames (sORFs) were found in different kind of ncRNAs, and their translated products have been termed sORF encoded peptides (SEPs) or micropeptides. Some micropeptides have been confirmed to exist stably in cells and play important roles independently of their source RNA. In this review, we summarize the identification methods of micropeptides derived from ncRNAs, the types of RNA that can encode micropeptides, and focus on the functional micropeptides found in diseases and plants. The purpose of the review is to provide a thought on the screening and identification of micropeptides, and provide new ideas for micropeptides as potentials for drug development or crop yield improvement.

Key words: non-coding RNA; micropeptide; non-canonical translation; identification methods; regulation mechanism

随着人类基因组计划的完成以及 ENCODE 计划的开展, 科学家发现, 约 75% 的基因组可以产生转录本 (Derrien et al., 2012; Djebali et al., 2012)。早期对于蛋白质编码基因的鉴定和注释主要集中在以 ATG 起始的能够编码至少 100 个氨基酸

的多肽的 ORF, 因此过去普遍认为, 能够形成遗传信息最终产物的蛋白质编码基因仅占整个基因组的 2%, 剩余的转录本构成了非编码 RNA 转录组 (Derrien et al., 2012; Djebali et al., 2012; Guttman et al., 2013)。在非编码 RNA 转录组中, 长链非编

* 收稿日期: 2022-10-30

录用日期: 2023-02-03

网络首发日期: 2023-03-12

基金项目: 国家重点研发计划项目(2022YFA1300020); 广东省自然科学基金杰出青年基金(2021B1515020002)

作者简介: 陈晓彤(1997年生), 女; 研究方向: 非编码 RNA 生物学; E-mail: chenxt223@mail2.sysu.edu.cn

通信作者: 陈月琴(1964年生), 女; 研究方向: 非编码 RNA 生物学; E-mail: lsscyq@mail.sysu.edu.cn

码 RNA 是数量最多的成员,在各种生物学过程中扮演了重要的角色,包括维持细胞干性、调控生长发育以及参与肿瘤发生发展等重大生命活动(Esteller, 2011; Hangauer et al., 2013; Wang et al., 2013a; Iyer et al., 2015; Brazão et al., 2016; Delás et al., 2017)。

随着转录组学和蛋白质组学等生物信息学技术发展,越来越多的研究表明,一些传统上被认为不编码蛋白的 RNA 区域(包括长链非编码 RNA 以及信使 RNA(mRNA, messenger RNA)的 5' 和 3' 非翻译区(UTR, untranslated region)等)实际上也存在开放阅读框(ORF, open reading frame)(Orr et al., 2020)。这些开放阅读框的长度通常小于 300 nt (nucleotide), 被称为小开放阅读框(Orr et al., 2020)。由核糖体印记技术(ribosome footprint profiling)发展而来的翻译组测序(Ribo-Seq, ribosome profiling sequencing), 为发现新的翻译模式提供了技术手段(Ingolia et al., 2012)。Ribo-Seq 结合质谱技术(MS, mass spectrometry), 许多 sORF 被发现可以翻译出长度通常小于 100 个氨基酸的小肽(Slavoff et al., 2013)。这些长度短于 100 个氨基酸的小肽在生物体内发挥重要功能, 包括参与个体发育、肌肉收缩和 DNA 修复等(Lu et al., 2004; Ma et al., 2014; Ma et al., 2016; Olexiouk et al., 2018)。

由于这些非编码 RNA 来源的小肽长期以来被认为并不存在, 所以被称为“幽灵蛋白质组”(ghost proteome)或“隐藏的蛋白质组”(hidden proteome)(Yang et al., 2019; Cardon et al., 2021)。当前, 功能性 sORF 的鉴定已成为基因组注释中的主要挑战, 本文将从非编码 RNA 来源小肽的鉴定方法、可编码小肽的非编码 RNA 种类、小肽的功能等几个方面来系统综述这类新发现的小肽, 重点论述这类小肽的功能机制, 以期为非编码 RNA 来源小肽的系统鉴定、编码规律和功能研究提供新视角。

1 非编码 RNA 来源小肽的鉴定方法

小肽由于其肽段长度较短, 容易降解等特点, 难以被传统翻译分析方法所捕捉。因此, 为了探寻当前未注释和未充分研究的“隐藏蛋白质组”, 需要更精准的鉴定方法。

翻译起始的标准扫描模型认为, 核糖体与 RNA 5' 帽结构结合形成起始前复合体, 该复合体在 mRNA 上扫描, 当发现以起始密码子 AUG 为中

心的 Kozak 序列后开始进行翻译且多肽延伸, 直至遇到终止密码子结束翻译, 核糖体从转录本上脱落(Moteki et al., 2002; Sonenberg et al., 2009)。由于翻译的场所在细胞质, 而核内的 ncRNA 无法接触翻译机器, 因此对于具有潜在翻译能力的 ncRNA 的初步筛选是寻找具有 m7G 结构且定位于胞质的 ncRNA(A)。(1)分析 ncRNA 序列是否具有 7-甲基鸟苷(m7G)帽子结构以及 Kozak 序列;(2)利用核质分离以及荧光原位杂交(FISH, fluorescence in situ hybridization)实验, 检测 ncRNA 是否稳定存在于细胞质中。

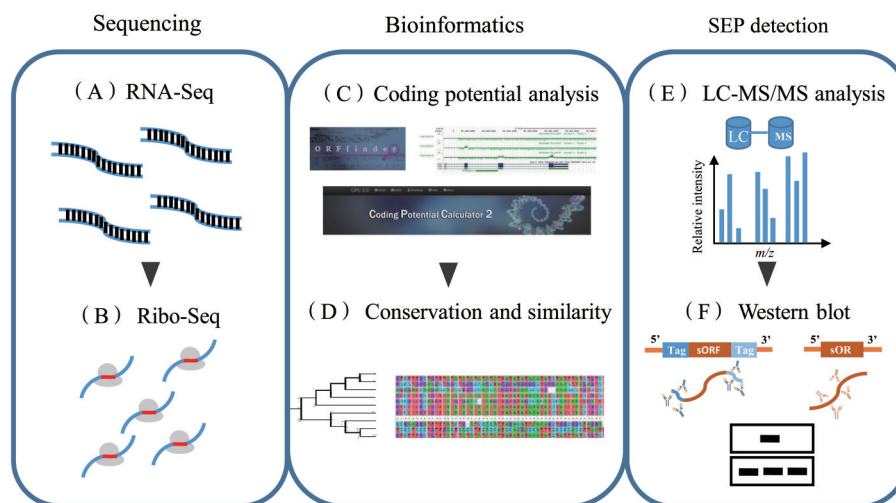
除了标准扫描模型外, 核糖体还存在泄露扫描、分流、通读和内部核糖体进入位点(IRES, internal ribosome entry site)等替代翻译模式, 研究发现多数真核生物的蛋白质编码基因缺乏最佳的 Kozak 区间序列, 且翻译可以在近源起始密码子(CUG/GUG/ACG 等)处起始(Ivanov et al., 2011; Kearsse et al., 2017; Yang et al., 2019; Cao et al., 2020)。Ribo-Seq 从多核糖体复合物中提取完整的 RNA 转录物, 并通过核酸酶处理降解未受核糖体结合保护的 RNA 片段, 因此可预测潜在的 sORF(Olexiouk et al., 2018)。因此, 通过结合 Ribo-Seq 技术可以全局鉴定具有潜在翻译能力的 ncRNA。翻译起始测序(TI-Seq, translation initiation sequencing)通过 lactimidomycin(LTM)或 harringtonine(HARR)等翻译抑制剂诱导核糖体在起始密码子处阻滞, 可预测起始密码子, 特别是非经典起始密码子(Ingolia et al., 2012)(B)。

一些研究显示 ncRNA 本身可以通过结合核糖体发挥翻译调控功能(Carrieri et al., 2012; Yoon et al., 2012; Tran et al., 2016)。因此, 为了排除核糖体仅停留在 ncRNA 上不进行翻译的情况, 需要进一步确认 ncRNA 上的开放阅读框是否能够进行有效翻译。通常可采用以下方法:(1)生物信息学预测。通过 RNA 编码预测工具, 例如 sORFfinder(Hanada et al., 2010), PhyloCSF(Lin et al., 2011), CPAT(Wang et al., 2013b), CPC/CPC2(Kong et al., 2007; Kang et al., 2017), IRESite(Mokrejs et al., 2006), M6AMRFS(Qiang et al., 2018)等, 寻找 ncRNA 上潜在的开放阅读框, 并通过查找序列上下区间中的翻译调节元件(如 IRES、m6A 修饰等)进行序列的编码能力预测(图 1C);(2)保守性分析, 功能编码序列通常表现出较高的跨物种密码子保守性, 虽然 lncRNA 保守性比 mRNA 低, 但通

常具有功能活性的蛋白编码 ORF 序列保守性比较高, 因此可以通过分析序列片段的保守性来寻找有效 ORF (图 1D)。(3)转录组学分析。通过 Ribo-Seq 测序技术可以检测核糖体结合的 ncRNA 转录本, 进一步的生物信息学分析算法可以对潜在的活跃翻译 ORF 实现密码子分辨率的预测。

sORF 编码的小肽通常比蛋白质更易降解, ncRNA 如果翻译了却没有产生稳定表达的蛋白, 那么核糖体通读产生的小肽可能只是一个副产物。因此对蛋白是否真实表达并且稳定存在的检测非常重要, 可通过以下 5 种方式筛选。(1)质谱鉴定。质谱技术提供小肽存在的直接证据, 可以鉴定小肽在生理条件下的氨基酸组成以及丰度 (图 1E)。

(2)标签肽验证。在 ORF 序列 N 或 C 末端进行标签标记, 通过蛋白质印迹法 (western blot) 验证标签肽是否翻译。(3)内源抗体检测。生产靶向小肽的特异氨基酸序列抗体, 通过 western blot 确认小肽的大小是否与预期相符, 在内源细胞系样品中验证小肽在体内条件下是否真实存在 (图 1F)。(4)体外翻译。将 ncRNA 全序列经过原核/真核体外转录以及翻译系统, 结合 ^{35}S 同位素放射性自显影技术, 检测 ncRNA 是否可以翻译小肽以及可以翻译几条小肽。(5)开放阅读框突变。使用起始密码子突变或者移码突变, 可以确认开放阅读框及其产生小肽的准确性。



结合测序技术、生物信息学分析以及蛋白质鉴定方法对非编码RNA来源小肽进行筛选鉴定。

测序技术(A) RNA-Seq与(B) Ribo-Seq用于大规模筛选具有翻译潜能的候选RNA;生物信息学平台对候选RNA进行(C)编码能力预测以及(D)序列保守性分析;(E)质谱技术和(F)蛋白质印迹法验证小肽的真实翻译。

图1 非编码RNA来源小肽的鉴定方法

Fig. 1 Methods for identification of small peptides derived from non-coding RNA

2 可编码小肽的非编码RNA种类

非编码RNA在转录组中占据极大的比例, 形成了高度复杂的家族, 包括长非编码RNA (lncRNA, long non-coding RNA)、微小RNA (miRNA, microRNA)、环状RNA (circRNA, circular RNA)、核糖体RNA (rRNA, ribosome RNA)、小干扰RNA (siRNA, small interfering RNA)等 (Kapranov et al., 2007)。目前对于非编码翻译的研究主要集中在包括mRNA的5'和3' UTR以及lncRNA、pri-miRNA和circRNA等非编码RNA区域中 (Orr et al., 2020)。

2.1 mRNA非翻译区(UTR)

UTR位于成熟mRNA两端, 分为5' UTR和3' UTR。过去认为UTR通常不编码蛋白质, 主要行使对蛋白质编码区(CDS, coding sequence)的翻译调控功能 (图2A)。随着越来越多的sORF在UTR区域被发现, 人们将位于mRNA的5' UTR上的ORF命名为上游ORF (upORF, upstream ORF), 位于3' UTR的ORF称为下游ORF (dORF, downstream ORF) (Calvo et al., 2009; Couso et al., 2017; Wu et al., 2020a)。

sORF在5' UTR中大量存在, 根据终止密码子的位置, upORF可分为完全上游ORF (cuORF,

completely upstream ORF) 和上游重叠 ORF (uoORF, upstream overlapping ORF) (Calvo et al., 2009; Ye et al., 2015)。目前研究发现, 许多具有较长 5' UTR 的转录本的蛋白质翻译水平会显著下降, upORF 的存在可能会翻译产生小肽, 进而导致核糖体在 mRNA 上的提前脱离, 通过阻止部分或全部核糖体对 CDS 区的扫描来抑制 mRNA 的常规蛋白翻译 (Calvo et al., 2009; Ye et al., 2015)。当然 upORF 不一定总会影响 CDS 的翻译, 这取决于 upORF 的终止密码子与 CDS 区的距离, 以及 upORF 前后序列的起始能力强弱 (Wagner et al., 2020)。Rodriguez et al. (2019) 发现 31% 的神经母细胞瘤转录本中, 一共鉴定了 4 954 个可翻译的 upORFs, 且主要是通过非经典起始密码子起始翻译, 所得小肽可以作为 CDS 翻译的顺式调节因子。例如, 从 GADD34 的 5' UTR 中的 upORF 翻译而来的 SEP 通过 C 端保守的 3' 氨基酸序列介导核糖体释放从而抑制主 CDS 的翻译 (Young et al., 2015)。相比 upORF, 位于 3' UTR 上的 dORF 的功能较少被报道, 其中一些 dORF 的翻译也被证实参与主 CDS 的翻译调控中 (Couso et al., 2017; Wu et al., 2020b)。例如 Wu et al. (2020b) 发现 dORF 可作为 CDS 的翻译增强子, 当对 dORF 的起始密码子进行突变以抑制其翻译时, CDS 的表达水平随之降低, 暗示了 dORF 编码的小肽与 CDS 翻译之间存在的联系。

2.2 长链非编码 RNA (lncRNA)

相对于其他非编码 RNA, lncRNA 具有数量众多并且功能多样等特点, 在可产生小肽的非编码 RNA 中, lncRNA 是最被关注的一类 ncRNA (图 2B)。与其他非编码 RNA 不同的是, 大部分 lncRNA 与 mRNA 有着相似的特点。例如, 它们均由 RNA 聚合酶 II 转录出来, 转录完成后均有 m⁷G 加帽以及 Poly A 加尾的过程, 经过可变剪切的加工, 并且它们都容易在细胞质聚集, 这些特点使得 lncRNA 具备翻译小肽的潜能 (Ruiz-Orera et al., 2014)。目前发现一些 lncRNA 编码功能小肽, 如 lncRNA ASHIL-AS1 编码位于内质网的小肽 APPLE, 通过调控翻译维持癌细胞高速合成, 促进 AML 发展 (Sun et al., 2021); lncRNA HOXB-AS3 编码保守的 53 aa (amino acids) 小肽, 通过调节肿瘤能量代谢来抑制结直肠癌 (CRC, colorectal carcinoma) 细胞的生长、集落形成、迁移、侵袭和肿瘤发生 (Huang et al., 2017); LINC00691 编码的小肽 SPAR

可以调控肌肉细胞的再生能力等 (Matsumoto et al., 2017)。但是也有研究提出质疑, 认为仅仅是 RNA 结构以及核糖体的结合并不能作为一个转录本的绝对可翻译条件, 实验污染会导致核糖体结合的 RNA 中不仅有 lncRNA, 还包括了一些核内的非编码 RNA。这些核内非编码 RNA 由于空间阻断缺乏翻译机会; 而采用核糖体释放速率算法可以将核糖体结合 lncRNA 与翻译的 mRNA 显著区分, 这暗示了大量与核糖体结合 lncRNA 可能未被真实翻译 (Guttman et al., 2013)。的确, 也有研究表明 lncRNA 可通过结合核糖体参与 mRNA 翻译调控中, 如 lincRNA-p21 (Yoon et al., 2012)、AS-RBM15 (Tran et al., 2016) 以及 antisense Uchl1 (Carrieri et al., 2012) 等均结合在核糖体上, 但是不翻译成小肽, 而是以 RNA 的形式发挥翻译调控的功能。这些研究表明, 功能小肽的准确鉴定仍然具有很大挑战。

2.3 环状 RNA (circRNA)

环状 RNA 是一类共价闭合单链 RNA, 由前体转录本 (pre-RNA) 的可变剪接而来, 包括内含子来源的环状 RNA 和外显子来源的环状 RNA, 后者占总环状 RNA 比重最大 (Li et al., 2018)。与经典 RNA 顺式剪接不同, 外显子来源环状 RNA 由 RNA 的反式剪接生成。在成环过程中, 外显子序列中上游外显子 3' 剪切位点同下游外显子 5' 剪切位点剪切连接成环状结构, 进而形成成熟的环状转录本 (Ashwal-Fluss et al., 2014; Zhang et al., 2014; Ivanov et al., 2015; Starke et al., 2015; Wang et al., 2015)。由此可见, 环状 RNA 具有环化以及部分基因的外显子序列, 因此, 具有潜在 ORF 的特点, 暗示了环状 RNA 具有翻译的潜能 (图 2C)。此外, 大部分环状 RNA 主要存在于细胞质中, 在空间上为环状 RNA 的翻译提供了可行性 (Sinha et al., 2022)。

与 mRNA 不同, 环状 RNA 的闭合结构使其缺少 5' 端 m⁷G 帽子, 导致翻译起始复合物 eIF4F 无法结合并介导核糖体进入环状 RNA 从而启动翻译过程 (Jeck et al., 2013; Guo et al., 2014; Schneider et al., 2016)。近年来陆续有文章报道环状 RNA 上存在多个开放阅读框, 并且可以通过帽子非依赖途径实现核糖体进入, 这些途径包括: (1) 通过 IRES 介导的翻译起始。研究表明, 许多环状 RNA 上存在 IRES 序列, 这些 IRES 序列可以通过自身结构或者结合帽子非依赖翻译起始蛋白来招募核糖体起

始翻译(Yang et al., 2019)。(2)通过m⁶A甲基化修饰介导的翻译起始(Yang et al., 2017a)。研究表明mRNA 5'UTR上存在的m⁶A修饰可以介导帽子非依赖的翻译起始过程,由于m⁶A修饰在环状RNA上分布广泛,环状RNA可以通过m⁶A阅读蛋白YTHDF3招募翻译起始复合物与核糖体从而起始翻译(Legnini et al., 2017; Pamudurti et al., 2017; Yang et al., 2017a)。

尽管大部分环状RNA自身序列并不是很长,但是由于其共价闭合环状的特点,开放阅读框可以跨过成环位点直到出现终止密码子(Liang et al., 2019; Zhang et al., 2021)。因此,一部分环状RNA编码的“小肽”会超过100个氨基酸,包括SHPRH-146aa(Begum et al., 2018)、AKT3-174aa(Xia et al., 2022)、FBXW7-185aa(Yang et al., 2018)、circDIDO1-529aa(Zhang et al., 2021)和 β -catenin-370aa(Liang et al., 2019)等。然而,很多具有明显IRES元件和开放阅读框的环状RNA无法翻译出蛋白,提示明确环状RNA翻译规律仍是一个亟须解决的科学问题。最近研究发现环状RNA的翻译亦或受到压力影响,例如细胞饥饿可以增强circMbl翻译效果(Pamudurti et al., 2017);热刺激可以促进包含m⁶A的环状RNA表达报告基因如绿色荧光蛋白GFP(Yang et al., 2017a)。这些研究进展表明了部分环状RNA翻译具有一定的时空特异性。

2.4 pri-miRNA

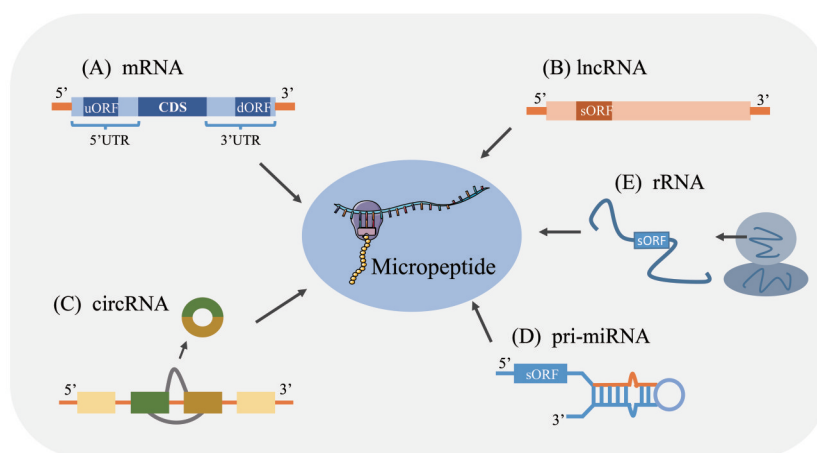
miRNA的生物发生包括两个步骤:(1)通过RNA聚合酶II转录产生初级miRNA(pri-miRNA)转录物;(2)由Dicer like 1(DCL1)蛋白将pri-

miRNA加工成前体miRNA(pre-miRNA)和最终成熟miRNA(Carthew et al., 2009)。最近研究表明,原始病毒可能含有编码调节性小肽(miPEPs)的短开放阅读框(图2D)。这些短肽影响相关miRNA的积累(Fang et al., 2017; Laressergues et al., 2015)。虽然,这种调节的分子机制仍未被破译,但是目前已经证明miPEPs、miPEP171d、miPEP172c、miPEP858a和miPEP165a可以调节植物的生长和发育(Yadav et al., 2021)。

关于pri-miRNAs的肽编码潜力的第一个证据是:Laressergues et al.(2015)发现,蒺藜状苜蓿的pri-miRNA171b和唐松草的pri-miR165a含有编码调节肽的sORF。这些短肽特异性地正向调节其相应成熟miRNA的积累。MiPEP171b过表达导致侧根密度降低,与miPEP171b过表达植株的表型一样。类似地,miPEP165a过表达导致根伸长。此外,为了验证miPEP过表达和相应miRNA积累之间的正相关性,同时分析了其他5种主要miRNAs的开放阅读框。这些miPEP(miPEP160b、miPEP164a、miPEP319、miPEP169d和miPEP171e)的过度表达或外部应用导致其相应miRNAs的更高积累(Laressergues et al., 2015)。

2.5 核糖体RNA(rRNA)

核糖体RNA编码小肽的报道较少(图2E)。目前发现线粒体16S rRNA编码的小肽Humanin参与细胞凋亡等多种生命活动的调控(Lee et al., 2013)、线粒体12S rRNA编码的小肽MOTS-c可以通过调控胰岛素敏感性来抑制饮食诱导的肥胖(Lee et al., 2015)。



小肽可从(A) mRNA的5'和3'非翻译区(UTR)、(B) lncRNA、(C) pri-miRNA和(D) circRNA等非编码RNA区域中翻译而来。

图2 可编码小肽的非编码RNA种类

Fig. 2 Classification of non-coding RNA encoding micropeptides

3 小肽的功能

虽然近年来在真核基因转录本中挖掘出大量的 sORF, 但是目前已被鉴定的具有体内翻译能力和生物活性的小肽数量仍非常少(Vitorino et al., 2021)。越来越多的研究证实, 小肽在哺乳动物和植物中均有存在, 且在多种生物进程中发挥着重要功能, 如 RNA 去帽过程、DNA 修复、压力通路、凋亡、肌肉形成、代谢稳态、钙离子稳态等(Yeasmin et al., 2018)。转录组测序结果显示, 编码小肽的非编码 RNA 如 lncRNA 往往在癌症中异常表达, 并可以编码与患者总生存期和治疗反应相关的小肽。这些研究结果提示非编码 RNA 来源的小肽可能参与癌症发生发展, 具有潜在的临床应用价值(Hanahan et al., 2011; Merino-Valverde et al., 2020; Wu et al., 2020b; Ye et al., 2020; Prensner et al., 2021)。研究显示许多小肽在癌症中失调表达, 通过调节增殖、凋亡、代谢和细胞炎症反应影响肿瘤进程(Yang et al., 2017b; Zhang et al., 2018)。在植物中, 也发现可以通过外源合成 pri-miRNA 来源的小肽来改善植物的农艺性状(Yadav et al., 2021)。

3.1 小肽参与调控细胞增殖

在生理条件下, 细胞通过严格控制有丝分裂信号来维持正常的结构和稳态。相比之下, 癌细胞最突出的特点就是过度的有丝分裂导致的无限增殖以及永生化(图 3A)。最近研究表明, 一些小肽参与调节了癌细胞的有丝分裂信号传导(Polycarpou-Schwarz et al., 2018; Pang et al., 2020; Xu et al., 2020; Zheng et al., 2021)。在癌细胞中显著高表达的小肽通常通过激活信号通路来促进癌细胞增殖。例如, 癌症相关小整体膜开放阅读框 1(CASIMO1, cancer-associated small integral membrane open reading frame 1)来源于非编码 RNA NR_029453 上的一个 sORF, 在乳腺癌中显著高表达(Polycarpou-Schwarz et al., 2018)。研究发现 CASIMO1 编码的小肽 SMIM22 与角鲨烯环氧化酶(SQLE, squalene epoxidase)相互作用, 导致 SQLE 蛋白积累和脂滴聚集增加, SQLE 蛋白可促进 ERK 磷酸化和 MAPK 通路激活, 进而导致 G0/G1 细胞周期停滞, 促进癌细胞增殖。同样, circPPP1R12A 编码的功能小肽 circPPP1R12A-73aa(PPP1R12A-C), 在结直肠癌组织中显著增加, 通过激活 Hippo-YAP 信号通路促进 CRC 生长和转移(Zheng et al., 2021)。LINC00998 编码的小肽

SMIM30 在肝癌患者中高度表达, SIMM30 与非受体酪氨酸激酶 SRC/YES1 结合, 驱动其膜锚定和磷酸化, 激活下游 MAPK 信号通路, 促进肝癌细胞在体外和体内的增殖和迁移(Pang et al., 2020)。相反, 在癌细胞中显著低表达的小肽则发挥了抑制癌细胞生长增殖的作用。例如, lncRNA NCBP2-AS2 编码的小肽 KRASIM 与 KRAS 蛋白及其蛋白水平相互作用, 抑制 ERK 信号活性, 从而抑制肝细胞癌(HCC, hepatocellular carcinoma)的细胞生长和增殖(Xu et al., 2020)。

sORFs 除了通过调控信号通路来影响有丝分裂之外, 还可以通过其他方式调节癌细胞的增殖。例如, lncRNA LINC00467 编码的 ATP 合酶相关肽(ATP synthase-associated peptide, ASAP), 通过与 ATP5A 和 ATP5C 相互作用提高 ATP 合酶活性和线粒体耗氧率, 从而促进体外和体内 CRC 细胞增殖(Ge et al., 2021)。LINC-PINT 编码的小肽 PINT87aa 与 forkhead box M1(FOXM1)的 DNA 结合域结合, 减少其靶基因抑制素 2(prohibitin 2, PHB2)的转录, 从而发挥抗增殖和抗衰老作用(Xiang et al., 2021)。circPINTexon2 编码的 PINT87aa 直接与聚合酶相关因子复合物(polymerase associated factor complex, PAF1c)相互作用, 参与 PAF1/RNA II 聚合酶复合物调控, 使 RNA 聚合酶 II 在特定癌基因启动子(如细胞周期蛋白 D1、CPEB1、c-MYC 和 SOX2)处暂停, 从而在体外和体内抑制胶质母细胞瘤细胞增殖(Zhang et al., 2018)。在结肠癌中低表达的 HOXB-AS3 peptide 可以通过调控 mRNA 的可变剪切来影响癌细胞的生长、转移和侵袭(Huang et al., 2017)。

3.2 小肽参与调控细胞凋亡

调节细胞死亡和存活之间的平衡对于维持组织稳态至关重要。细胞凋亡作为一种由基因编程的自主细胞死亡形式, 是正常发育和组织稳态所必需的(Delbridge et al., 2012; Childs et al., 2014; Aubrey et al., 2018)。研究表明, 一些非编码 RNA 来源的小肽参与细胞凋亡功能调控。lncRNA LINC00278 编码小肽 YY1BM(Yin Yang 1-binding micropeptide), 可在营养剥夺下通过雄激素受体途径诱导细胞凋亡, 在食管鳞状细胞癌(ESCC, esophageal squamous cell carcinoma)中具有肿瘤抑制活性(Wu et al., 2020c)。在肿瘤内注射纯化的小肽 YY1BM 在异种移植模型中显示出治疗效果, 表明其具有作为肿瘤抑制剂的潜力。同样, 胃肠道

特异性 lncRNA LINC00675 被发现可编码小肽 FORCP (FOXA1-regulated conserved small protein), 在 CRC 中显著低表达。FORCP 定位于内质网 (ER, endoplasmic reticulum), 通过与 BRI3BP 相互作用从而抑制 CRC 细胞增殖, 并在 ER 应激下诱导细胞凋亡 (Li et al., 2020)。

线粒体是细胞凋亡调控中心, 多数凋亡刺激因子通过线粒体激活细胞凋亡途径。Humanin 是在线粒体基因组中鉴定出的第一个 sORF, 由线粒体 16S rRNA 编码, 研究发现 Humanin 可通过介导 BCL-2 蛋白质家族的细胞内定位来调节内源性或线粒体凋亡途径 (Guo et al., 2003; Luciano et al., 2005; Lee et al., 2013)。Humanin 可上调 BCL-2 的表达并抑制 BAX 表达以抑制细胞凋亡 (Guo et al., 2003)。此外, Humanin 还可以与其他 BCL-2 蛋白家族成员 (如 BID 和 BAX) 相互作用, 并调节它们的易位以抑制凋亡体的产生 (Luciano et al., 2005)。PIGBOS 是一种由 PIGB 基因的反义链 RNA 编码的新型小肽, 定位于线粒体外膜, 通过 CLCC1 蛋白与 ER 相互作用, PIGBOS 的下调通过增加对化学诱导的 UPR 的敏感性来诱导细胞凋亡 (Chu et al., 2019)。虽然 PIGBOS 功能的分子机制尚未被描述, 但有理由认为, 其可能通过使细胞对未折叠蛋白质反应 (UPR, unfolded protein response) 敏感并迫使它们进入细胞凋亡, 其在癌细胞中的抑制可能具有治疗潜力。

3.3 小肽参与调控细胞物质代谢与能量代谢

细胞代谢是指细胞通过对葡萄糖、氨基酸、脂肪酸等燃料的分解, 产生用于细胞活动的 ATP (Judge et al., 2020)。这些过程主要发生在人体的能量工厂——线粒体中, 并需要多种酶和辅助因子的参与 (Fennie et al., 2004)。实验生物学和生物信息学分析均表明, 小肽在线粒体内膜 (IMM, inner mitochondrial membrane) 上大量富集, 参与细胞的代谢活动 (Kim et al., 2017) (图 3C)。小肽 BRAWNIN 由 C12orf73 基因编码, 在心脏和骨骼肌细胞中高度表达并定位于 IMM 中 (Zhang et al., 2020)。小肽 BRAWNIN 通过与 UQCRC1 的相互作用来促进呼吸链复合物 III (CIII, respiratory chain complex III) 组装, 通过反馈回路调节细胞能量状态。敲低 BRAWNIN 可导致细胞 ATP 合成减少, 从而促进 AMPK 活化。小肽 MtlN 与 BRAWNIN 类似, 在心脏和骨骼肌中高表达, 与线粒体三功能蛋白 (MTP, mitochondrial trifunctional

protein) 的亚基 HADHB 互作, 因此, MtlN 被认为参与脂肪酸氧化的调节 (Makarewich et al., 2018)。除 MTP 外, MtlN 还与心磷脂相互作用, 心磷脂对于维持膜完整性和组装呼吸用多蛋白复合物至关重要 (Stein et al., 2018)。MtlN 的过表达增加了基础和最大呼吸速率, 减少了糖酵解 (Chugunova et al., 2019)。

LncRNA HOXB-AS3 被证明可以翻译产生参与 RNA 剪接的 53AA 小肽。Huang 及其合作者发现, HOXB-AS3 在结直肠癌细胞中下调, 改变了丙酮酸激酶 M (PKM, pyruvate kinase M) 前 mRNA 的剪接形式, 重新表达胚胎同种型 PKM2, 有利于糖酵解活性; 与此同时, HOXB-AS3 来源小肽的表达有利于促进氧化磷酸化的成体亚型 (PKM1) 的表达。总的来说, HOXB-AS3 来源小肽在结直肠癌细胞系中的过表达可以通过改变其对葡萄糖代谢的使用而减弱其致癌能力 (Huang et al., 2017)。线粒体基因组 12s rRNA 上的开放阅读框编码的小肽 MOTS-c 被发现在体外可增加葡萄糖摄取、糖酵解活性和 AMPK 活化, 同时降低耗氧率, 改善了与体内肥胖相关的代谢参数 (Lee et al., 2015)。

3.4 小肽调控炎症免疫通路

炎症是由先天免疫系统发起的一系列细胞反应, 通过消除病原体以及恢复细胞和机体稳态来维持宿主健康 (El-Kenawi et al., 2017; Evavold et al., 2019; Medzhitov, 2008)。已有研究表明, 线粒体为免疫细胞识别外来抗原以及合成促炎细胞因子和趋化因子提供能量, 在调节炎症反应方面起关键作用 (Subramanian et al., 2013; Jo et al., 2016; West et al., 2017; Forrester et al., 2018; Neagu et al., 2019; Andrieux et al., 2021)。小肽在线粒体中的富集也暗示其对细胞炎症免疫反应的调控作用 (Kim et al., 2017) (图 3D)。

炎症小体是一种大分子蛋白复合物, 其中 IL-1 β 前体通过半胱天冬酶 1 (caspase 1) 转化为生物活性 IL-1 β , 引发 IL-1 介导的炎症 (Weber et al., 2010)。线粒体小肽 -47 (Mm47, mitochondrial micropeptide-47) 被发现在活化的巨噬细胞中差异表达, 其对免疫刺激的下调和线粒体定位使其在 Nlrp3 介导的炎性小体反应具有调节作用 (Bhatta et al., 2020)。然而, Mm47 在调节 Nlrp3 寡聚化程度和 OMM 招募方面的分子机制仍然未知。此外, 线粒体小肽 MOTS-c 和 Humanin 也被报道能够通过调节衰老细胞的线粒体活性来加剧衰老相关分泌表型 (SASP,

senescence-associated secretory phenotype)的促炎作用(Kim et al., 2018; Mendelsohn et al., 2018)。

3.5 小肽调控植物生长发育

在植物中,近年研究发现 pri-miRNA 编码的小肽(miPEPs)发挥了重要的作用,特别是在农艺性状方面。miR172c 是大豆结瘤的正向调节因子(Wang et al., 2019),其过度表达导致结节数量的增加,抑制则具有相反的效果。Couzigou et al. (2016)表明,用合成的大豆 miR172c 前体编码的小肽 miPEP172c 处理大豆可显著增加根瘤数,而不影响其他根系发育特征,与在大豆中过量表达 miR172c 的结果一致。进一步研究还发现,用 miPEP172c 处理大豆还可以增加了 miR172c 的积累,说明小肽 miPEP 可以上调 miRNAs 表达(图 3E)。并且,与对照植株相比,经 miPEP172c 处理的植株中 ENOD40-1、NIN、NSP1 和 Hb2 表达也显著提高,表明 miPEP172c 在植物发育过程以及植物-微生物相互作用中都发挥重要作用。这些进展揭示了调节性 miPEPs 在改善作物农艺性状中的重要性。

类似地,发现由 pri-miR858a 编码的小肽 miPEP858a 调节拟南芥类黄酮的生物合成和发育(Sharma et al., 2020)。外源应用合成的 miPEP858a 可恢复由 miPEP858a 缺失导致的拟南芥植物发育缺陷。此外,启动子报告基因分析表明,miPEP858a 具有调控自身启动子活性和 GUS 基因转录的作用,暗示 miPEPs 具有特定的生物学功能(图 3F)。

不定根的形成是葡萄无性繁殖的主要障碍。最近的一项研究发现了外施 miR-171d 编码的小肽 miPEP171d1 在不定根形成中的重要作用(Chen et al., 2020)。将外源 miPEP171d1 导入欧洲葡萄组培苗中,可促进 miR171d 的积累和不定根的发育(图 3G)。这些结果显示了一种新的 miRNA 调节机制,该机制与其相关的 miPEP 有关。在拟南芥中,通过对 pri-miRNA 来源小肽 miPEP164a、miPEP165a 和 miPEP319a 等的研究(Yadav et al., 2021),发现这些小肽对植株生长都具有明显的促进作用。与对照植株相比,喷洒、浇水、堆肥、添加肥料等外施 miPEPs 的植株在茎高、早花、花数增加和花柄高度方面均表现出明显优势。外源施用 miPEPs 时,植株内的 miPEP 总量增加,因此总体上调节了相关 miRNAs 的积累。有趣的是,miPEPs 可以在外源施用下发挥作用,这一发现将在农业上有

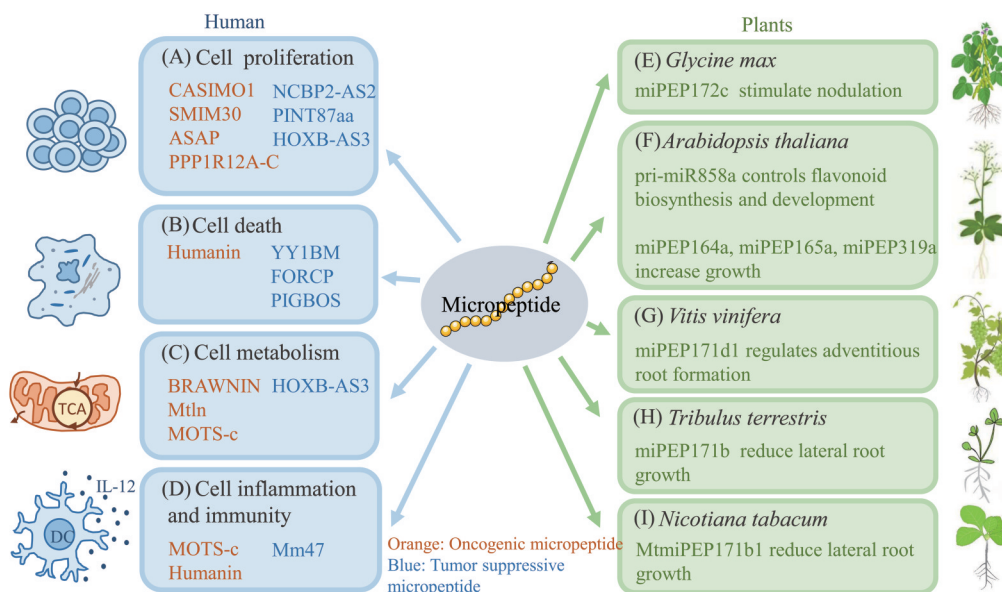
广泛的用途。

在蒺藜苜蓿中外施和过表达 miPEP171b 均能使植株内的 miR171b 表达升高并抑制其靶基因(Lauressergues et al., 2015)(图 3H)。miPEP171b1 过表达导致 Mtmir171b 累积,从而负调控靶基因 HAM 和 HAM2,使植株侧根数量减少,表明其在根发育中的重要性。Mtmir171b 和 MtmirPEP171b1 在烟草中的过表达也具有相似的结果(Lauressergues et al., 2015)。通过对 MtmirORF171b 和 MtmirPEP171b1 以及突变 ORF MtmirORF171b 的相对表达量发现, MtmirORF171b 对 MtmirPEP171b1 的翻译作用增加了 Mtmir171b 的积累(Lauressergues et al., 2015)。外施和过表达 MtmirPEP171b1 可增加各自 pri-miRNAs、pre-miRNAs 和成熟 miRNAs 的积累(图 3I)。而用 ATT 替代 pri-miR171b ORF1 中的 ATG 起始密码子后,仍有少量的 miR171b 产生,这表明 miPEP171b 充当了调节介质并使相关的 miRNA 积累增加(Lauressergues et al., 2015)。

外源 miPEPs 如何进入植物细胞内是限制其在植物中发挥功能的重要问题。一般认为,内吞作用和被动扩散是 miPEPs 内化的主要途径(Ormancey et al., 2020)。在用荧光染料标记 MiPEP165a 对拟南芥进行外源处理后,观察到 miPEP165a 能快速进入根冠和分生组织区,而根部其余部分的渗透延迟(Ormancey et al., 2020)。通过研究进入植株体内的网格蛋白介导途径和膜微结构域途径发现;miPEP165a 在根冠、分生组织和根的分化区是被动进入,而在根细胞成熟区的进入受到影响。其中,rem1-2 在 miPEP 吸收中受到强烈影响。进一步通过使用化学抑制剂 TyrA23 和 M β CD 分别抑制网格蛋白介导和膜微域途径。结果表明,这些分子阻断了施用 miPEP165a 对根长的正向表型效应。这说明 miPEPs 可能具有局部效应,而不是系统效应。

4 小结与展望

虽然许多具有 sORF 的 RNA 曾经被归类为 ncRNA,因而在以往的研究中忽视了对这些具有潜在翻译能力的 sORF 鉴定,但是越来越多研究结果表明具有生物活性的非编码 RNA 来源小肽真实存在并且在生物体内发挥不可替代的功能,使我们重新认识这些非编码 RNA 的功能形式(Lu et al., 2004; Ma et al., 2014; Ma et al., 2016; Olexiouk et al., 2018)。本文总结了识别、筛选和鉴定 sORF 的



小肽在人类与植物中均有存在,且在多种生物进程中发挥着重要的作用。

在人类中,小肽通过参与调控(A) 细胞增殖、(B) 细胞死亡、(C) 细胞物质与能量代谢以及(D) 炎症反应等影响疾病的进程;在植物中,来源于 pri-miRNA 的小肽在(E) 大豆 *Glycine max*、(F) 拟南芥 *Arabidopsis thaliana*、(G) 葡萄 *Vitis vinifera*、(H) 蒺藜苜蓿 *Tribulus terrestris* 和(I) 烟草 *Nicotiana tabacum* 等多个物种中参与调控植物的生长发育,可用于改良植物的农艺性状。

图3 小肽的功能

Fig. 3 The function of micropeptides

方法,以及编码小肽的非编码RNA种类,揭示了非编码RNA来源小肽在生物体内的多重作用机制与生物功能。这些进展揭示了非编码RNA来源的小肽具有成为抗肿瘤药物、治疗靶点以及肿瘤生物标志物的巨大潜力,例如作为新型抗肿瘤药物,小肽相比其来源RNA具有更高的活性,更低的免疫原性和更低的细胞毒性,因此可以合成肿瘤抑制微肽并直接递送到癌症组织(Zhu et al., 2018; Bakhti et al., 2022)。由于小肽易于代谢降解的特点,也使在植物中应用外源合成小肽相比使用转基因手段和化学农药具有更高的安全性,将成为一种新的改善植物农艺性状的手段(Yadav et al., 2021)。这些新发现使研究者开始重新审视过去的非编码RNA研究,是否由于技术限制了对其编码能力的预测和验证?其编码小肽是否可能发挥与母体RNA相似的功能,甚至其母体RNA“表现”出的功能是否实际上由其编码小肽执行?分子研究手段的不断发展让小肽的功能与RNA自身活性得以区分,也强调了在未来的非编码RNA研究中分离RNA功能和小肽功能的必要性。此外,一些研究发现mRNA也具有与其编码蛋白无关的非编码功能,例如一种编码早期减数分裂诱导蛋白1(Emi1)的mRNA转录本,可从其3'UTR加工成熟

一种功能性长非编码RNA——端粒酶RNA(TER)(Logeswaran et al., 2022)。目前已有定义统一将既能以RNA形式发挥功能,又能以蛋白形式发挥功能的RNA称为cncRNA(coding and non-coding RNAs),也叫双功能RNA,包括了具有编码功能的ncRNA和具有非编码功能的mRNA(Kumari et al., 2015)。未来的小肽研究可能会将更多的非编码RNA列入双功能RNA的队伍中。

同时,尽管一些非编码RNA来源的小肽的重要功能已经被揭示,但是还有许多问题尚未解决。例如,在小肽的筛选和翻译验证方面,如何进一步优化现有检测技术减少假阳性,找到更多可翻译的小肽?决定sORF成功翻译小肽的关键因素是什么?是否有特定的上下文序列决定sORF能否进行翻译?无义介导的mRNA衰变(NMD, nonsense-mediated mRNA decay)作为监视机制可降解错误的mRNA,并对虚假翻译做出反应(Lykke-Andersen et al., 2015; Supek et al., 2021)。含有sORF的lncRNA很可能成为NMD的目标,那么NMD靶向lncRNA的程度如何?同时,在对小肽的功能研究方面,如何系统性大量验证小肽的不同生理功能?如何克服小蛋白的研究障碍(如蛋白丰度低、不稳定、缺乏特异性抗体等)?如何改善和优化小肽的

结合亲和力和长效呈递能力来实现小肽在疾病中和植物中的实际应用? 因此, 未来我们不仅仅需要努力扩大和完善小肽检测技术, 还需要继续发展各种各样的实验手段来验证和研究小肽的时空

表达模式及生物学功能, 丰富和完善小肽在生物体中发挥功能的各种分子机制, 为利用小肽作为药物研发或者农作物增产的关键靶点和实际应用提供新的思路 and 方向。

参考文献:

- ANDRIEUX P, CHEVILLARD C, CUNHA-NETO E, et al, 2021. Mitochondria as a cellular hub in infection and inflammation[J]. *Int J Mol Sci*, 22(21): 11338.
- ASHWAL-FLUSS R, MEYER M, PAMUDURTI N, et al, 2014. circRNA biogenesis competes with pre-mRNA splicing[J]. *Mol Cell*, 56(1): 55-66.
- AUBREY B J, KELLY G L, JANIC A, et al, 2018. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression?[J]. *Cell Death Differ*, 25(1): 104-113.
- BAKHTI S Z, LATIFI-NAVID S, 2022. Non-coding RNA-encoded peptides/proteins in human cancer: The future for cancer therapy [J]. *Curr Med Chem*, 29(22): 3819-3835.
- BEGUM S, YIU A, STEBBING J, et al, 2018. Novel tumour suppressive protein encoded by circular RNA, circ-SHPRH, in glioblastomas[J]. *Oncogene*, 37(30): 4055-4057.
- BHATTA A, ATIANAND M, JIANG Z, et al, 2020. A mitochondrial micropeptide is required for activation of the Nlrp3 inflammasome [J]. *J Immunol*, 204(2): 428-437.
- BRAZÃO T F, JOHNSON J S, MÜLLER J, et al, 2016. Long noncoding RNAs in B-cell development and activation [J]. *Blood*, 128(7): e10-e19.
- CALVO S E, PAGLIARINI D J, MOOTHA V K, 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans [J]. *Proc Natl Acad Sci USA*, 106(18): 7507-7512.
- CAO X, SLAVOFF S A, 2020. Non-AUG start codons: Expanding and regulating the small and alternative ORFeome [J]. *Exp Cell Res*, 391(1): 111973.
- CARDON T, FOURNIER I, SALZET M, 2021. Shedding light on the ghost proteome[J]. *Trends Biochem Sci*, 46(3): 239-250.
- CARRIERI C, CIMATTI L, BIAGIOLI M, et al, 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat[J]. *Nature*, 491(7424): 454-457.
- CARTHEW R W, SONTHEIMER E J, 2009. Origins and mechanisms of miRNAs and siRNAs[J]. *Cell*, 136(4): 642-655.
- CHEN Q J, DENG B H, GAO J, et al, 2020. A miRNA-encoded small peptide, vvi-miPEP171d1, regulates adventitious root formation[J]. *Plant Physiol*, 183(2): 656-670.
- CHILDS B G, BAKER D J, KIRKLAND J L, et al, 2014. Senescence and apoptosis: Dueling or complementary cell fates?[J]. *EMBO Rep*, 15(11): 1139-1153.
- CHU Q, MARTINEZ T F, NOVAK S W, et al, 2019. Regulation of the ER stress response by a mitochondrial microprotein [J]. *Nat Commun*, 10(1): 4883.
- CHUGUNOVA A, LOSEVA E, MAZIN P, et al, 2019. *LINC00116* codes for a mitochondrial peptide linking respiration and lipid metabolism[J]. *Proc Natl Acad Sci USA*, 116(11): 4940-4945.
- COUSO J P, PATRAQUIM P, 2017. Classification and function of small open reading frames [J]. *Nat Rev Mol Cell Biol*, 18(9): 575-589.
- COUZIGOU J M, ANDRÉ O, GUILLOTIN B, et al, 2016. Use of microRNA-encoded peptide miPEP172c to stimulate nodulation in soybean[J]. *New Phytol*, 211(2): 379-381.
- DELÁS M J, SABIN L R, DOLZHENKO E, et al, 2017. lncRNA requirements for mouse acute myeloid leukemia and normal differentiation[J]. *eLife*, 6: e25607.
- DELBRIDGE A R, VALENTE L J, STRASSER A, 2012. The role of the apoptotic machinery in tumor suppression [J]. *Cold Spring Harb Perspect Biol*, 4(11): a008789.
- DERRIEN T, JOHNSON R, BUSSOTTI G, et al, 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression [J]. *Genome Res*, 22(9): 1775-1789.
- DJEBALI S, DAVIS C A, MERKEL A, et al, 2012. Landscape of transcription in human cells [J]. *Nature*, 489(7414): 101-108.
- EL-KENAWI A, RUFFELL B, 2017. Inflammation, ROS, and mutagenesis [J]. *Cancer Cell*, 32(6): 727-729.
- ESTELLER M, 2011. Non-coding RNAs in human disease [J]. *Nat Rev Genet*, 12(12): 861-874.
- EVAVOLD C L, KAGAN J C, 2019. Inflammasomes: Threat-assessment organelles of the innate immune system [J]. *Immunity*, 51(4): 609-624.
- FANG J, MORSALIN S, RAO V, et al, 2017. Decoding of non-coding DNA and non-coding RNA: Pri-micro RNA-encoded novel peptides regulate migration of cancer cells [J]. *J Pharmaceut Sci Pharmacol*, 3(1): 23-27.
- FERNIE A R, CARRARI F, SWEETLOVE L J, 2004. Respiratory metabolism: Glycolysis, the TCA cycle and mitochondrial electron transport [J]. *Curr Opin Plant Biol*, 7(3): 254-261.
- FORRESTER S J, KIKUCHI D S, HERNANDES M S, et al, 2018. Reactive oxygen species in metabolic and inflammatory signaling [J]. *Circ Res*, 122(6): 877-902.
- GE Q, JIA D, CEN D, et al, 2021. Micropeptide ASAP encoded by *LINC00467* promotes colorectal cancer progression by directly modulating ATP synthase activity [J]. *J Clin Invest*, 131(22): e152911.

- GUO B, ZHAI D, CABEZAS E, et al, 2003. Humanin peptide suppresses apoptosis by interfering with Bax activation [J]. *Nature*, 423(6938): 456–461.
- GUO J U, AGARWAL V, GUO H, et al, 2014. Expanded identification and characterization of mammalian circular RNAs [J]. *Genome Biol*, 15(7): 409.
- GUTTMAN M, RUSSELL P, INGOLIA N, et al, 2013. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins[J]. *Cell*, 154(1): 240–251.
- HANADA K, AKIYAMA K, SAKURAI T, et al, 2010. sORF finder: A program package to identify small open reading frames with high coding potential[J]. *Bioinformatics*, 26(3): 399–400.
- HANAHAHAN D, WEINBERG R, 2011. Hallmarks of cancer: The next generation[J]. *Cell*, 144(5): 646–674.
- HANGAUER M J, VAUGHN I W, MCMANUS M T, 2013. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs [J]. *PLoS Genet*, 9(6): e1003569.
- HUANG J Z, CHEN M, CHEN D, et al, 2017. A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth [J]. *Mol Cell*, 68(1): 171–184.
- INGOLIA N T, BRAR G A, ROUSKIN S, et al, 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments [J]. *Nat Protoc*, 7(8): 1534–1550.
- IVANOV A, MEMCZAK S, WYLER E, et al, 2015. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals[J]. *Cell Rep*, 10(2): 170–177.
- IVANOV I P, FIRTH A E, MICHEL A M, et al, 2011. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences[J]. *Nucleic Acids Res*, 39(10): 4220–4234.
- IYER M K, NIKNAFS Y S, MALIK R, et al, 2015. The landscape of long noncoding RNAs in the human transcriptome [J]. *Nat Genet*, 47(3): 199–208.
- JECK W R, SORRENTINO J A, WANG K, et al, 2013. Circular RNAs are abundant, conserved, and associated with ALU repeats [J]. *RNA*, 19(2): 141–157.
- JO E K, KIM J K, SHIN D M, et al, 2016. Molecular mechanisms regulating NLRP3 inflammasome activation[J]. *Cell Mol Immunol*, 13(2): 148–159.
- JUDGE A, DODD M S, 2020. Metabolism[J]. *Essays Biochem*, 64(4): 607–647.
- KANG Y J, YANG D C, KONG L, et al, 2017. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features[J]. *Nucleic Acids Res*, 45(W1): W12–W16.
- KAPRANOV P, CHENG J, DIKE S, et al, 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription[J]. *Science*, 316(5830): 1484–1488.
- KEARSE M G, WILUSZ J E, 2017. Non-AUG translation: A new start for protein synthesis in eukaryotes[J]. *Genes Dev*, 31(17): 1717–1731.
- KIM S J, XIAO J, WAN J, et al, 2017. Mitochondrially derived peptides as novel regulators of metabolism[J]. *J Physiol*, 595(21): 6613–6621.
- KIM S J, MEHTA H H, WAN J, et al, 2018. Mitochondrial peptides modulate mitochondrial function during cellular senescence [J]. *Aging*, 10(6): 1239–1256.
- KONG L, ZHANG Y, YE Z Q, et al, 2007. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine [J]. *Nucleic Acids Res*, 35 (Web Server Issue): W345–W349.
- KUMARI P, SAMPATH K, 2015. cncRNAs: Bi-functional RNAs with protein coding and non-coding functions [J]. *Semin Cell Dev Biol*, 47–48: 40–51.
- LAURESSERGUES D, COUZIGOU J M, CLEMENTE H S, et al, 2015. Primary transcripts of microRNAs encode regulatory peptides[J]. *Nature*, 520(7545): 90–93.
- LEE C, YEN K, COHEN P, 2013. Humanin: A harbinger of mitochondrial-derived peptides? [J]. *Trends Endocrinol Metab*, 24(5): 222–228.
- LEE C, ZENG J, DREW B, et al, 2015. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance[J]. *Cell Metab*, 21(3): 443–454.
- LEGNINI I, Di TIMOTEO G, ROSSI F, et al, 2017. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis[J]. *Mol Cell*, 66(1): 22–37.
- LI X, YANG L, CHEN L L, 2018. The biogenesis, functions, and challenges of circular RNAs[J]. *Mol Cell*, 71(3): 428–442.
- LI X L, PONGOR L, TANG W, et al, 2020. A small protein encoded by a putative lncRNA regulates apoptosis and tumorigenicity in human colorectal cancer cells[J]. *eLife*, 9: e53734.
- LIANG W C, WONG C W, LIANG P P, et al, 2019. Translation of the circular RNA circ β -catenin promotes liver cancer cell growth through activation of the Wnt pathway [J]. *Genome Biol*, 20(1): 84.
- LIN M F, JUNGREIS I, KELLIS M, 2011. PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions[J]. *Bioinformatics*, 27(13): i275–i282.
- LOGESWARAN D, LI Y, AKHTER K, et al, 2022. Biogenesis of telomerase RNA from a protein-coding mRNA precursor [J]. *Proc Natl Acad Sci USA*, 119(41): e2204636119.
- LU P D, HARDING H P, RON D, 2004. Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response[J]. *J Cell Biol*, 167(1): 27–33.
- LUCIANO F, ZHAI D, ZHU X, et al, 2005. Cytoprotective peptide humanin binds and inhibits proapoptotic Bcl-2/Bax family protein BimEL[J]. *J Biol Chem*, 280(16): 15825–15835.
- LYKKE-ANDERSEN S, JENSEN T H, 2015. Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes [J]. *Nat Rev Mol Cell Biol*, 16(11): 665–677.
- MA J, WARD C C, JUNGREIS I, et al, 2014. Discovery of human

- sORF-encoded polypeptides (SEPs) in cell lines and tissue[J]. *J Proteome Res*, 13(3): 1757–1765.
- MA J, DIEDRICH J K, JUNGREIS I, et al, 2016. Improved identification and analysis of small open reading frame encoded polypeptides[J]. *Anal Chem*, 88(7): 3967–3975.
- MAKAREWICH C A, BASKIN K K, MUNIR A Z, et al, 2018. MOXI is a mitochondrial micropeptide that enhances fatty acid β -oxidation[J]. *Cell Rep*, 23(13): 3701–3709.
- MATSUMOTO A, PASUT A, MATSUMOTO M, et al, 2017. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide [J]. *Nature*, 541(7636): 228–232.
- MEDZHITOV R, 2008. Origin and physiological roles of inflammation[J]. *Nature*, 454(7203): 428–435.
- MENDELSON A R, LARRICK J W, 2018. Mitochondrial-derived peptides exacerbate senescence [J]. *Rejuvenation Res*, 21(4): 369–373.
- MERINO-VALVERDE I, GRECO E, ABAD M, 2020. The microproteome of cancer: From invisibility to relevance[J]. *Exp Cell Res*, 392(1): 111997.
- MOKREJS M, VOPÁLENSKÝ V, KOLENATY O, et al, 2006. IRESite: The database of experimentally verified IRES structures (www.iresite.org) [J]. *Nucleic Acids Res*, 34(Database issue): D125–D130.
- MOTEKI S, PRICE D, 2002. Functional coupling of capping and transcription of mRNA[J]. *Mol Cell*, 10(3): 599–609.
- NEAGU M, CONSTANTIN C, POPESCU I D, et al, 2019. Inflammation and metabolism in cancer cell—Mitochondria key player [J]. *Front Oncol*, 9: 348.
- OLEXIOUK V, van CRIEKINGE W, MENSCHAERT G, 2018. An update on sORFs.org: a repository of small ORFs identified by ribosome profiling[J]. *Nucleic Acids Res*, 46(D1): D497–D502.
- ORMANCEY M, RU A L, DUBOÉ C, et al, 2020. Internalization of miPEP165a into *Arabidopsis* roots depends on both passive diffusion and endocytosis-associated processes[J]. *Int J Mol Sci*, 21(7): 2266.
- ORR M W, MAO Y, STORZ G, et al, 2020. Alternative ORFs and small ORFs: Shedding light on the dark proteome [J]. *Nucleic Acids Res*, 48(3): 1029–1042.
- PAMUDURTI N R, BARTOK O, JENS M, et al, 2017. Translation of CircRNAs[J]. *Mol Cell*, 66(1): 9–21.
- PANG Y, LIU Z, HAN H, et al, 2020. Peptide SMIM30 promotes HCC development by inducing SRC/YES1 membrane anchoring and MAPK pathway activation [J]. *J Hepatol*, 73(5): 1155–1169.
- POLYCARPOU-SCHWARZ M, GROSS M, MESTDAGH P, et al, 2018. The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation [J]. *Oncogene*, 37(34): 4750–4768.
- PRENSNER J R, ENACHE O M, LURIA V, et al, 2021. Noncanonical open reading frames encode functional proteins essential for cancer cell survival[J]. *Nat Biotechnol*, 39(6): 697–704.
- QIANG X, CHEN H, YE X, et al, 2018. M6AMRFS: Robust prediction of N6-methyladenosine sites with sequence-based features in multiple species[J]. *Front Genet*, 9: 495.
- RODRIGUEZ C M, CHUN S Y, MILLS R E, et al, 2019. Translation of upstream open reading frames in a model of neuronal differentiation[J]. *BMC Genomics*, 20(1): 391.
- RUIZ-ORERA J, MESSEGUER X, SUBIRANA J A, et al, 2014. Long non-coding RNAs as a source of new peptides [J]. *eLife*, 3: e03523.
- SCHNEIDER T, HUNG L H, SCHREINER S, et al, 2016. CircRNA-protein complexes: IMP3 protein component defines subfamily of circRNPs[J]. *Sci Rep*, 6: 31313.
- SHARMA A, BADOLA P K, BHATIA C, et al, 2020. Primary transcript of miR858 encodes regulatory peptide and controls flavonoid biosynthesis and development in *Arabidopsis* [J]. *Nat Plants*, 6(10): 1262–1274.
- SINHA T, PANIGRAHI C, DAS D, et al, 2022. Circular RNA translation, a path to hidden proteome [J]. *Wiley Interdiscip Rev RNA*, 13(1): e1685.
- SLAVOFF S A, MITCHELL A J, SCHWAID A G, et al, 2013. Peptidomic discovery of short open reading frame-encoded peptides in human cells[J]. *Nat Chem Biol*, 9(1): 59–64.
- SONENBERG N, HINNEBUSCH A G, 2009. Regulation of translation initiation in eukaryotes: Mechanisms and biological targets [J]. *Cell*, 136(4): 731–745.
- STARKE S, JOST I, ROSSBACH O, et al, 2015. Exon circularization requires canonical splice signals [J]. *Cell Rep*, 10(1): 103–111.
- STEIN C S, JADIYA P, ZHANG X, et al, 2018. Mitoregulin: A lncRNA-encoded microprotein that supports mitochondrial super complexes and respiratory efficiency [J]. *Cell Rep*, 23(13): 3710–3720.
- SUBRAMANIAN N, NATARAJAN K, CLATWORTHY M, et al, 2013. The adaptor MAVS promotes NLRP3 mitochondrial localization and inflammasome activation [J]. *Cell*, 153(2): 348–361.
- SUN L, WANG W, HAN C, et al, 2021. The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation[J]. *Mol Cell*, 81(21): 4493–4508.
- SUPEK F, LEHNER B, LINDEBOOM R G H, 2021. To NMD or not to NMD: Nonsense-mediated mRNA decay in cancer and other genetic diseases[J]. *Trends Genet*, 37(7): 657–668.
- TRAN N T, SU H, KHODADADI-JAMAYRAN A, et al, 2016. The AS-RBM15 lncRNA enhances RBM15 protein translation during megakaryocyte differentiation[J]. *EMBO Rep*, 17(6): 887–900.
- VITORINO R, GUEDES S, AMADO F, et al, 2021. The role of micropeptides in biology[J]. *Cell Mol Life Sci*, 78(7): 3285–3298.
- WAGNER S, HERRMANNOVÁ A, HRONOVÁ V, et al, 2020.

- Selective translation complex profiling reveals staged initiation and co-translational assembly of initiation factor complexes [J]. *Mol Cell*, 79(4): 546–560.
- WANG L, PARK H J, DASARI S, et al, 2013a. CPAT: Coding–Potential Assessment Tool using an alignment–free logistic regression model [J]. *Nucleic Acids Res*, 41(6): e74.
- WANG L, SUN Z, SU C, et al, 2019. A GmNINa–miR172c–NNC₁ regulatory network coordinates the nodulation and autoregulation of Nodulation pathways in soybean [J]. *Mol Plant*, 12(9): 1211–1226.
- WANG Y, CHEN L, CHEN B, et al, 2013b. Mammalian ncRNA–disease repository: A global view of ncRNA–mediated disease network [J]. *Cell Death Dis*, 4(8): e765.
- WANG Y, WANG Z, 2015. Efficient backsplicing produces translatable circular mRNAs [J]. *RNA*, 21(2): 172–179.
- WEBER A, WASILIEW P, KRACHT M, 2010. Interleukin–1beta (IL–1beta) processing pathway [J]. *Sci Signal*, 3(105): cm2.
- WEST A P, SHADEL G S, 2017. Mitochondrial DNA in innate immune responses and inflammatory pathology [J]. *Nat Rev Immunol*, 17(6): 363–375.
- WU P, MO Y, PENG M, et al, 2020a. Emerging role of tumor–related functional peptides encoded by lncRNA and circRNA [J]. *Mol Cancer*, 19(1): 22.
- WU Q, WRIGHT M, GOGOL M M, et al, 2020b. Translation of small downstream ORFs enhances translation of canonical main open reading frames [J]. *EMBO J*, 39(17): e104763.
- WU S, ZHANG L, DENG J, et al, 2020c. A novel micropeptide encoded by Y–linked LINC00278 links cigarette smoking and AR signaling in male esophageal squamous cell carcinoma [J]. *Cancer Res*, 80(13): 2790–2803.
- XIA X, LI X, LI F, et al, 2022. Correction: A novel tumor suppressor protein encoded by circular AKT3 RNA inhibits glioblastoma tumorigenicity by competing with active phosphoinositide–dependent Kinase–1 [J]. *Mol Cancer*, 21(1): 124.
- XIANG X, FU Y, ZHAO K, et al, 2021. Cellular senescence in hepatocellular carcinoma induced by a long non–coding RNA–encoded peptide PINT87aa by blocking FOXM1–mediated PHB2 [J]. *Theranostics*, 11(10): 4929–4944.
- XU W, DENG B, LIN P, et al, 2020. Ribosome profiling analysis identified a KRAS–interacting microprotein that represses oncogenic signaling in hepatocellular carcinoma cells [J]. *Sci China Life Sci*, 63(4): 529–542.
- YADAVA, SANYALI, RAIS P, et al, 2021. An overview on miRNA–encoded peptides in plant biology research [J]. *Genomics*, 113(4): 2385–2391.
- YANG L, TANG Y, HE Y, et al, 2017a. High Expression of LINC01420 indicates an unfavorable prognosis and modulates cell migration and invasion in nasopharyngeal carcinoma [J]. *J Cancer*, 8(1): 97–103.
- YANG Y, FAN X, MAO M, et al, 2017b. Extensive translation of circular RNAs driven by N⁶–methyladenosine [J]. *Cell Res*, 27(5): 626–641.
- YANG Y, GAO X, ZHANG M, et al, 2018. Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis [J]. *J Natl Cancer Inst*, 110(3): 304–315.
- YANG Y, WANG Z, 2019. IRES–mediated cap–independent translation, a path leading to hidden proteome [J]. *J Mol Cell Biol*, 11(10): 911–919.
- YE M, ZHANG J, WEI M, et al, 2020. Emerging role of long noncoding RNA–encoded micropeptides in cancer [J]. *Cancer Cell Int*, 20: 506.
- YE Y, LIANG Y, YU Q, et al, 2015. Analysis of human upstream open reading frames and impact on gene expression [J]. *Hum Genet*, 134(6): 605–612.
- YEASMIN F, YADA T, AKIMITSU N, 2018. Micropeptides encoded in transcripts previously identified as long noncoding RNAs: A new chapter in transcriptomics and proteomics [J]. *Front Genet*, 9: 144.
- YOON J H, ABDELMOHSEN K, SRIKANTAN S, et al, 2012. LincRNA–p21 suppresses target mRNA translation [J]. *Mol Cell*, 47(4): 648–655.
- YOUNG S K, WILLY J A, WU C, et al, 2015. Ribosome reinitiation directs gene–specific translation and regulates the integrated stress response [J]. *J Biol Chem*, 290(47): 28257–28271.
- ZHANG M, ZHAO K, XU X, et al, 2018. A peptide encoded by circular form of LINC–PINT suppresses oncogenic transcriptional elongation in glioblastoma [J]. *Nat Commun*, 9(1): 4475.
- ZHANG S, RELJIĆ B, LIANG C, et al, 2020. Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly [J]. *Nat Commun*, 11(1): 1312.
- ZHANG X O, WANG H B, ZHANG Y, et al, 2014. Complementary sequence–mediated exon circularization [J]. *Cell*, 159(1): 134–147.
- ZHANG Y, JIANG J, ZHANG J, et al, 2021. CircDIDO1 inhibits gastric cancer progression by encoding a novel DIDO1–529aa protein and regulating PRDX2 protein stability [J]. *Mol Cancer*, 20(1): 101.
- ZHENG X, CHEN L, ZHOU Y, et al, 2021. Correction to: A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo–YAP signaling [J]. *Mol Cancer*, 20(1): 42.
- ZHU S, WANG J, HE Y, et al, 2018. Peptides/proteins encoded by non–coding RNA: A novel resource bank for drug targets and biomarkers [J]. *Front Pharmacol*, 9: 1295.