

· 生命科学与新药探索 ·

长链非编码 RNA 编码的微肽研究进展

张雪, 李臣诚, 王颖*

(中国药科大学生命科学与技术学院, 江苏 南京 211198)

[摘要] 长链非编码 RNA (lncRNA) 是由超过 200 个核苷酸构成的 RNA 分子, 展现出相对较低的序列保守性, 长期以来在生物学领域内被视为“转录噪声”, 即无功能性的 RNA 分子。然而, 近年来, 随着研究的不断深入, 科学家们揭示出在 lncRNA 中潜藏着许多小开放阅读框, 其中部分能够编码微肽。这些微肽已被证实可参与多种细胞进程及基因表达调控网络, 并扮演着至关重要的角色。这一发现为生命活动的进一步探索以及疾病的临床诊断与治疗开辟了新的研究方向。对 lncRNA 编码的微肽在病理与生理过程中的作用、微肽的亚细胞定位与功能机制, 以及微肽研究方法的进展进行综述, 旨在为新型微肽诊疗一体化药物的开发提供思路与参考。

[关键词] 长链非编码 RNA; 小开放阅读框; 微肽; 肿瘤

[中图分类号] O629.72

[文献标志码] A

[文章编号] 1001-5094 (2024) 11-0823-15

DOI: 10.20053/j.issn1001-5094.2024.11.003

Advances in Research on Micropeptides Encoded by Long Non-coding RNAs

ZHANG Xue, LI Chencheng, WANG Ying

(School of Life Science and Technology, China Pharmaceutical University, Nanjing 211198, China)

[Abstract] Long non-coding RNAs (lncRNAs) are RNA molecules constituted by more than 200 nucleotides, exhibiting relatively low sequence conservation. For a long time, they had been regarded as “transcriptional noise”, that is, non-functional RNA molecules, in the biological field. However, in recent years, as research has advanced, scientists have unveiled numerous small open reading frames (sORFs) within lncRNAs, some of which can encode micropeptides. These micropeptides have been confirmed to participate in various cellular processes and gene expression regulatory networks, playing crucial roles. This discovery has opened new research directions for further exploration of life activities and clinical diagnosis and treatments of diseases. This review summarizes the roles of lncRNA-encoded micropeptides in pathological and physiological processes, the subcellular localization and functional mechanisms of micropeptides, and the progress in micropeptide research methodologies, aiming to provide insights and references for the development of novel integrated micropeptide-based diagnostics and therapeutics.

[Key words] long non-coding RNA; small open reading frame; micropeptide; tumor

1 引言

在生物学领域, 长链非编码 RNA (long non-coding RNA, lncRNA) 的研究一直是近年来的热点之一。作为基因组中占比高达 98% 的非蛋白质编码序列的重要组成部分, lncRNA 曾一度被视为“转录噪声”, 未受到足够重视^[1-2]。然而, 随着 DNA 元件百科全书 (ENCODE) 计划的深入研究和二代测序技术的发展, 人们发现 lncRNA 可通过多种机制参与各种生命活动, 调节关键的生物学过程, 如介

导转录激活、调控异染色质形成、结合小分子 RNA (miRNA) 等^[3]。还有一些研究报道, lncRNA 可通过反式或顺式作用机制对附近基因进行调控, 并且发现了广泛而复杂的 RNA-RNA 和蛋白质-RNA 相互作用网络^[4-6]。一些 lncRNA 可以产生内源性小干扰 RNA (siRNA) 和 miRNA 前体、改变蛋白质定位、调节蛋白质活性^[7-8]。它们不仅能够转录及转录后水平上调节蛋白编码基因的表达, 还广泛参与包括细胞分化、个体发育在内的多种重要生命过程^[9]。更重要的是, lncRNA 的异常表达与多种人类重大疾病的发生密切相关, 这使得它们成为疾病诊疗中的重要潜在靶点。

lncRNA 的研究并未止步于此。近年来, 越来越多的研究表明, lncRNA 中含有大量小于 300 个

接受日期: 2024-03-01

项目资助: 国家自然科学基金 (No. 82400545)

* 通信作者: 王颖, 副研究员;

研究方向: 长链非编码 RNA 及其产物微肽的作用机制研究;

E-mail: wy_cpu@126.com

碱基的小开放阅读框 (sORF), 部分 sORF 可以翻译为小于 100 个氨基酸的肽, 其被称为微肽^[10-12]。微肽在钙离子稳态、胚胎发育、线粒体代谢以及恶性肿瘤的发生发展中发挥着重要调节作用^[13]。这一发现不仅补充了经典的“中心法则”, 更提示我们在以往研究中可能忽视了 lncRNA 编码的微肽在生命调控过程的重要功能。

鉴于 lncRNA 编码的微肽的潜在生物活性及其与疾病的相关性, 对 lncRNA 编码的微肽的研究进展进行综述显得尤为重要。本文将从 lncRNA 编码的微肽在病理及生理过程中的作用、定位与分子机制以及鉴定方法等方面对其研究进展进行综述。同时, 本文也将探讨目前研究中面临的挑战和未来的研究方向, 以期对相关领域的研究者提供借鉴和参考, 并为 lncRNA 编码的微肽作为新型药物靶点和疾病生物标志物的开发提供思路。

2 lncRNA 的分类与其编码的微肽

分子生物学中心法则的基本原理是通过基因组的自我复制、转录、翻译, 最终使遗传信息流向蛋白质。蛋白质作为生命构成的基石, 在各类生命活动中发挥着不可或缺的作用。然而, 人类基因组中只有 2% 的基因可以编码蛋白质, 其余大部分则转录为非编码 RNA, 仅有少数得以表征^[1-2]。

lncRNA 作为非编码 RNA 的一个重要分支, 近年来随着研究的深入, 已被明确定义为由 RNA 聚合酶 II 催化转录、长度超越 200 个核苷酸且具有多聚腺苷酸化修饰的 RNA 分子^[14-16]。依据 lncRNA 与蛋白质编码基因在基因组上的相对位置, 可将其分为 4 种主要类型。1) 反义 lncRNA (antisense lncRNA): 转录方向与蛋白质编码基因方向相反, 并与 mRNA 外显子存在部分或完全重叠; 2) 基因间 lncRNA (intergenic lncRNA, LincRNA): 定位于蛋白质编码基因间隔区域; 3) 正义重叠 lncRNA (sense overlapping lncRNA): 转录方向与蛋白质编码基因方向相同, 并与 mRNA 一个或多个外显子重叠; 4) 内含子 lncRNA (intronic lncRNA): 基因位点在蛋白质编码基因的内含子内, 但不与同一链上的任何外显子重叠^[17] (见图 1)。

最近的研究表明, 生物体的发育过程不仅受到蛋白质的调控, lncRNA 在其中同样扮演着至关重要的角色^[17]。这些 lncRNA 通过多种方式调节生物的生理功能, 例如: 通过影响组蛋白修饰酶调节表观遗传^[18-19]; 通过竞争性结合 miRNA, 充当 RNA 海绵^[20]; 调控转录和剪接, 影响蛋白质翻译过程^[21]。此外, lncRNA 中含有 sORF, 可以翻译成为微肽, 这些微肽影响着生物体内多种不同的生理与病理过程, 因此关注微肽的功能尤为重要。

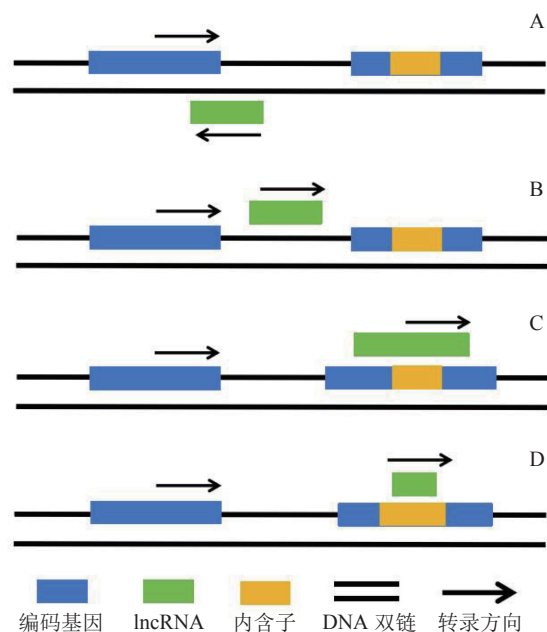


图 1 4 类长链非编码 RNA 示意图
A: 反义 lncRNA; B: 基因间 lncRNA; C: 正义重叠 lncRNA; D: 内含子 lncRNA

图 1 4 类长链非编码 RNA 示意图
Figure 1 Schematic diagram of four types of long non-coding RNAs

在生物信息学领域, 精确鉴别 sORF 是一项挑战。大多数算法将 100 个氨基酸作为最小检测单元, 这种过滤策略显著降低了假阳性率, 但也可能导致部分包含 sORF 的基因被错误注释为非编码 RNA^[3]。直到 1996 年, 研究者首次在大豆基因 *ENOD40* 中鉴定出由 sORF 编码的 2 个短肽 (分别为 12 和 24 个氨基酸), 并于 2002 年通过对 *ENOD40* 的进一步研究正式确认了这类短肽的功能^[14,22-24]。随着转录组学和蛋白质组学技术的飞速发展, 微肽的识别与鉴定取得了显著进展, 多个微肽数据库已经上线。

目前, 研究人员在真菌、动物和植物等多种生物类群中都发现了微肽, 并且这些微肽在生命过程中发挥着重要的作用, 例如影响代谢调控、维持 DNA 稳态、参与应激响应, 以及癌症和心血管疾病的发生发展^[13]。由此可见, 微肽为疾病机制、诊断靶标、治疗分子等方面的研究提供了新的理论基础, 深入挖掘微肽的功能机制将成为未来重要的研究方向。

3 lncRNA 编码的微肽在病理生理过程中的作用

lncRNA 编码的微肽展现出了多样的生物学功能。近年来, 微肽在生理与病理过程中的功能机制逐渐被阐明^[14,25], 极大地推动了该研究领域的深入发展。以下将介绍 lncRNA 编码的微肽如何通过与特定蛋白质、RNA 分子或细胞膜受体相互作用来调节细胞内的信号传导通路、基因表达模式以及细胞代谢过程, 从而在肿瘤、心血管疾病等多种病理过程及正常生理过程中发挥调节作用。

3.1 在癌症中的作用

3.1.1 头颈部鳞状细胞癌 头颈部鳞状细胞癌 (HNSCC) 是最常见的癌症之一, 研究人员在 lncRNA *RP11-469H8.6* 转录本中发现了一个包含 51 个密码子的新 sORF, 并将其翻译的微肽命名为抑制肌动蛋白细胞骨架的微肽 (MIAC)^[12]。MIAC 可与水通道蛋白 2 (AQP2) 相互作用并下调整合素 Beta4 (ITGB4) 和胞裂蛋白 2 (SEPT2), 从而影响纤维状肌动蛋白 (F-actin) 的细胞骨架形成。在人舌鳞癌细胞 (CAL27) 中稳定过表达 MIAC 后, 细胞增殖迁移能力降低。同时, MIAC 过表达可使 HNSCC 模型小鼠的肿瘤生长受到显著抑制。这些信息进一步证明了 MIAC 在肌动蛋白细胞骨架形成和肿瘤转移调控中的重要作用, 也揭示了 AQP2、SEPT2 和 ITGB4 的表达在 HNSCC 细胞增殖和迁移中的关键作用。

3.1.2 结直肠癌 6-甲基腺苷 (m^6A) 是真核 RNA 中最普遍的修饰物, RNA m^6A 修饰在转录调控中的作用离不开 m^6A 结合蛋白 (也称 m^6A 阅读器), 这类蛋白直接介导了多种生物学功能的实现。lncRNA *LINC00266-1* 可以编码含有 71 个氨基酸的微肽, 后者可与胰岛素样生长因子 2 mRNA 结合蛋白 1

(IGF2BP1, 一种 m^6A 阅读器) 等 RNA 结合蛋白相互作用, 因此被命名为“RNA 结合调节肽”(RBRP)^[8]。在原发性结直肠癌 (CRC) 中, RBRP 表达水平升高并结合 IGF2BP1, 从而增强 IGF2BP1 对靶 RNA (如致癌基因 *c-Myc* 的 mRNA) 上 m^6A 修饰结构的识别, 然后招募 RNA 稳定剂到 m^6A 修饰的靶 RNA 上, 最终使靶 RNA 的稳定性和表达量增加。总的来说, 由 *LINC00266-1* 编码的 RBRP 与 m^6A 阅读器相互作用, 可调节 m^6A 阅读器对靶 RNA 的 m^6A 修饰结构的识别, 从而发挥致癌作用。

3.1.3 食管癌 食管癌是消化道肿瘤中较为常见的类型, 也是全球第六大致死性癌症, 每年约有 30 万人死于食管癌。食管癌有食管鳞状细胞癌 (ESCC) 和腺癌 (EAC) 2 种类型, ESCC 是亚洲食管癌的主要亚型, 虽然多模式治疗改善了食管癌的疗效和预后, 但总体上 5 年生存率仍然较低。有研究发现了一个 Y 连锁的 lncRNA *LINC00278*^[26], 其在男性 ESCC 中下调。*LINC00278* 可以编码含有 21 个氨基酸的微肽, 该微肽可以和调控蛋白 YY1 结合, 因而被命名为 YY1BM。在机制上, YY1BM 可以抑制 YY1 与雄激素受体 (AR) 的相互作用, 进而使真核生长因子 2 激酶 (eEF2K) 的表达降低, 并在营养剥夺 (ND) 下诱导 ESCC 细胞凋亡, 可作为 ESCC 的潜在抗癌微肽。

3.1.4 乳腺癌 乳腺癌 (BC) 是最常见的恶性肿瘤之一, 在女性中发病率很高, 已严重威胁到女性的健康乃至生命。乳腺癌发病率通常在女性绝经前后较高, 主要集中在乳房腺上皮组织, 有一定的遗传性。因此, 阐明乳腺癌相关机制对于其临床治疗具有重要意义。最近, 许多研究揭示了 lncRNA 编码的微肽与乳腺癌的密切关联^[27-28]。

三阴性乳腺癌 (TNBC) 是乳腺癌的一种亚型, 具有侵袭性, 且预后性较差。研究发现, 谷氨酰胺 (Gln) 代谢在 TNBC 和非三阴性乳腺癌 (non-TNBC) 中有着关键作用。lncRNA *MLLT4-AS1* 可以编码含有 21 个氨基酸的微肽, 其可与剪接型 X-box 结合蛋白 1 (X-box binding protein 1 spliced, XBP1s) 相互作用, 因而被命名为 XBP1s 结合多肽 (XBP1SBM)^[29], 该微肽在 TNBC 组织和 Gln 缺失的 TNBC 细胞

系中上调。Gln 缺失可诱导 XBP1s 转录从而促进 XBP1SBM 表达上调, 同时 XBP1SBM 阻断了非剪接型的 X-box 蛋白 1 (X-box binding protein 1 unspliced, XBP1u) 和 XBP1s 之间的相互作用, 并将 XBP1s 保留在细胞核中以增强血管内皮生长因子 (VEGF) 的表达。在人内皮细胞、小鼠异种移植模型和小鼠自发性 BC 模型中进行的研究表明, XBP1SBM 可以改善 TNBC 中的 Gln 水平, 以促进血管生成和转移, 这种机制为 TNBC 提供了一种新的预后生物标志物和治疗靶点选择。

TNBC 组织中的 lncRNA *LINC00908*^[30] 可编码含有 60 个氨基酸的微肽 ASRPS (转录因子蛋白质 STAT3 的一种小调节肽)。与非 TNBC 组织相比, ASRPS 在 TNBC 组织中表达较低, 同时 ASRPS 低表达与 TNBC 患者生存率低有关。在机制上, ASRPS 通过卷曲螺旋结构域 (CCD) 与 STAT3 相互作用, 使得 STAT3 磷酸化减少, 降低了 VEGF 的表达。在人内皮细胞、小鼠异种移植 BC 模型和自发性 BC 模型中, ASRPS 过表达减少了血管生成。在小鼠异种移植乳腺癌模型中, ASRPS 的下调促进了肿瘤生长。各项研究证据表明, 微肽 ASRPS 具有治疗 TNBC 的潜力。

转化生长因子 (TGF)- β 信号通路在乳腺癌转移中起关键作用。有学者鉴定了一种由 lncRNA *LINC00665* 编码的含有 52 个氨基酸的微肽, 并命名为 CIP2A-BP, 其在乳腺癌细胞系中被 TGF- β 下调^[31]。具体而言, TGF- β 通过激活 Smad 信号通路, 诱导翻译抑制蛋白 4E-BP 的表达, 从而抑制真核翻译起始因子 (eIF4E), 导致 CIP2A-BP 的翻译减少。CIP2A-BP 可以与肿瘤致癌基因 *CIP2A* 直接作用, 增加了蛋白磷酸酶 2A (PP2A) 的活性, 从而抑制磷脂酰肌醇 3-激酶 (PI3K)/磷酸激酶 B (AKT)/核因子- κ B (NF- κ B) 通路及下游基因的表达。研究发现, 在乳腺癌自发模型小鼠中, CIP2A-BP 可以抑制肿瘤肺转移, 提高小鼠总生存率, 可作为乳腺癌细胞迁移和侵袭的抑制剂发挥作用, 也是 TNBC 的预后标志物。

3.1.5 胶质母细胞瘤 有研究发现 lncRNA *LINC-PINT*^[32] 是一种肿瘤抑制性基因间 lncRNA, 参与构成多梳蛋

白抑制复合物 2 (PRC2), 环状结构的 *LINC-PINT* 可编码含有 87 个氨基酸的微肽 PINT87aa。该微肽可与聚合酶相关因子复合物 (PAF1c) 结合, 提示其在转录延伸过程中具有抑制作用。同时, 与正常组织相比, 胶质母细胞瘤 (GBM) 中 PINT87aa 表达水平下调, 说明该微肽可能在 GBM 中发挥重要作用。随后的研究发现, 过表达 PINT87aa 的细胞表现出 G₁ 期停滞和细胞增殖减少, 而 PINT87aa 敲除细胞显示出细胞周期和细胞增殖速率增加, 表明该微肽可能抑制 GBM 细胞增殖。在异种移植肿瘤动物中进行的研究显示, PINT87aa 敲除后肿瘤体积显著增加, 进一步支持了 PINT87aa 在体内的抗癌作用。

3.1.6 黑色素瘤 皮肤黑色素瘤 (SKCM) 是皮肤瘤中恶性程度最高的肿瘤类型, 主要特征为色素性皮肤病的明显变化。其死亡率较高, 且易在早期发生转移, 因此早期诊断与治疗很重要。lncRNA *MELOE* 位于基因 *HDAC4* 的反义方向的内含子中, 包含多个 sORF^[33]。研究表明, *MELOE* 可以编码 3 种微肽, 即 MELOE-1、MELOE-2 和 MELOE-3, 长度分别为 39、44 和 54 个氨基酸。其中, MELOE-1 和 2 的翻译是通过内部核糖体结合位点 (IRES) 依赖性机制实现的, 且在黑色素瘤细胞中特异性高表达, 与黑色素瘤的 T 细胞免疫识别有关^[34-36], 可能有助于癌细胞的肿瘤发生及转移。然而, MELOE-3 通过经典的帽依赖性翻译, 在黑色素瘤和黑色素细胞中均高表达, 免疫原性较差。总的来说, 在 *MELOE* 抗原家族中, IRES 依赖性抗原是免疫原性最强的, 也是黑色素瘤免疫治疗的最佳靶点^[33]。

3.2 在其他病理生理过程中的作用

3.2.1 心血管疾病 内皮细胞 (EC) 是一种异质性细胞系, 可调节脉管系统的稳态。在血管生成过程中, EC 被促血管生成因子激活后开始发芽、增殖、迁移, 然后形成新的毛细血管并恢复到静止状态。研究发现, lncRNA *LINC00961* 编码的含有 75 个氨基酸的微肽 SPAAR (氨基酸反应的小调节多肽) 可以调节内皮细胞功能, 并且 *LINC00961* 和 SPAAR 均反向调节血管生成^[37]。此外, SPAAR 的缺失会影响小鼠的心脏发育、心肌动力学和心血管功能, 这表明 SPAAR 有助于成年期的生长发育以及基础心血管功

能, 从而降低心肌梗死的风险^[38]。上述研究结果或可为心血管疾病的临床治疗提供新的科学依据和策略。

3.2.2 肺动脉高压 肺动脉高压 (PH) 是一种肺血管疾病, 其特征是肺血管阻力逐渐增加, 导致劳力性呼吸困难症状逐渐恶化, 直至右心室衰竭, PH 中的一个重要病理过程与缺氧引起的肺动脉平滑肌细胞 (PASMC) 的增殖有关^[39]。lncRNA *Rps4l* 被发现于缺氧诱导的 PH 中下调, 其编码的微肽 RPS4XL 可以抑制 PASMC 增殖并减少 PASMC 诱导的 PH 死亡^[40-41]。此外, 微肽 RPS4XL 与核糖体蛋白 S6 (RPS6) 相互作用, 调节 RPS6 的磷酸化从而参与 PH 相关因子的通路, 最终影响缺氧诱导的 PH 病理过程。RPS4XL 与 RPS6 相互作用的发现为缺氧诱导的 PH 的治疗和早期诊断提供了新的思路^[39]。

3.2.3 mRNA 降解 lncRNA *LINC01420* 基因位于 X 染色体上, 含有 3 个外显子, 由 2 个内含子隔开。mRNA 降解相关的 RNA 蛋白质颗粒 (P-body) 是一种细胞质结构, 主要包含未翻译的 mRNA、RNA 降解相关的蛋白质以及其他与 mRNA 的代谢和调控有关的因子。研究人员发现 *LINC01420* 中的一个 ORF 可编码含有 68 个氨基酸的微肽, 并将其命名为 NoBody^[42]。NoBody 和 P-body 数量之间存在负相关关系, 并且和 mRNA 脱帽蛋白存在相互作用关系, 从而参与 mRNA 稳态调节和降解过程。同时, 研究表明 NoBody 的缺失导致无义介导的 mRNA 降解 (NMD) 底物的稳态水平降低, 这与 NoBody 对细胞中 mRNA 降解的负调控结果一致, 但其具体的调控机制还需要进一步的探究。总的来说, NoBody 是 mRNA 脱帽复合物的一种新成分, 其在 mRNA 稳态和降解过程中具有潜在的调控作用。

3.2.4 肌肉发育 最近有研究表明^[43], lncRNA 作为肌肉生成和骨骼肌再生的调节因子具有重要作用。有研究鉴定了一种可在骨骼肌中特异性富集的和肌生成相关的 lncRNA *lnc-mg*。对原代骨骼肌细胞的分析表明, *lnc-mg* 在肌源性分化过程中逐渐增加, 其过表达改善了细胞分化^[43]。研究发现 lncRNA *MyolncR4* 在肌肉分化过程中被上调, *MyolncR4* 可编码含有 56 个氨基酸的微肽^[44], 其被命名为 LEMP, 并在脊椎动物物种中高度保守。LEMP 可促

进小鼠肌肉形成和再生, 并与多种线粒体蛋白相关, 或可通过调节线粒体功能来调节肌肉发育和再生。

在分析骨骼肌特异性 lncRNA 时, Anderson 等^[45]发现了 *LINC00948* 中以前未被识别的 ORF, 其编码一个保守的含有 46 个氨基酸的微肽, 并将该微肽命名为肌调节蛋白 (MLN)。MLN 形成单个跨膜 α 螺旋, 与肌浆网 (SR) 膜中的心肌肌浆网 Ca^{2+} -ATP 酶 (SERCA) 相互作用并调节 Ca^{2+} , 是骨骼肌中主要的 SERCA 抑制性微肽。SERCA 是横纹肌性能的关键调节剂, 它作为主要的 Ca^{2+} ATP 酶, 负责将细胞溶质 Ca^{2+} 再摄取到 SR 中。MLN 与 SR 膜中的 SERCA 形成稳定的复合物, 并且 MLN 直接影响 SR Ca^{2+} 水平和机体最大运动表现。

3.2.5 胰腺 β 细胞内稳态调节 有研究鉴定了一种在 β 细胞和神经细胞中富集的 lncRNA *TUNAR* 所编码的微肽^[3], 该微肽高度保守且含有 48 个氨基酸, 被命名为 β 细胞和神经细胞调节蛋白 (BNLN)。研究发现, 过表达的 BNLN 降低了胰腺 β 细胞中内质网钙水平, 维持了内质网稳态, 并提高了葡萄糖刺激的胰岛素分泌。除胰腺外, *TUNAR* 在大脑、垂体、睾丸、输卵管和子宫中丰度高, 但还需要进一步的研究来发现该 lncRNA 编码的微肽在其中发挥的功能。

3.3 小结

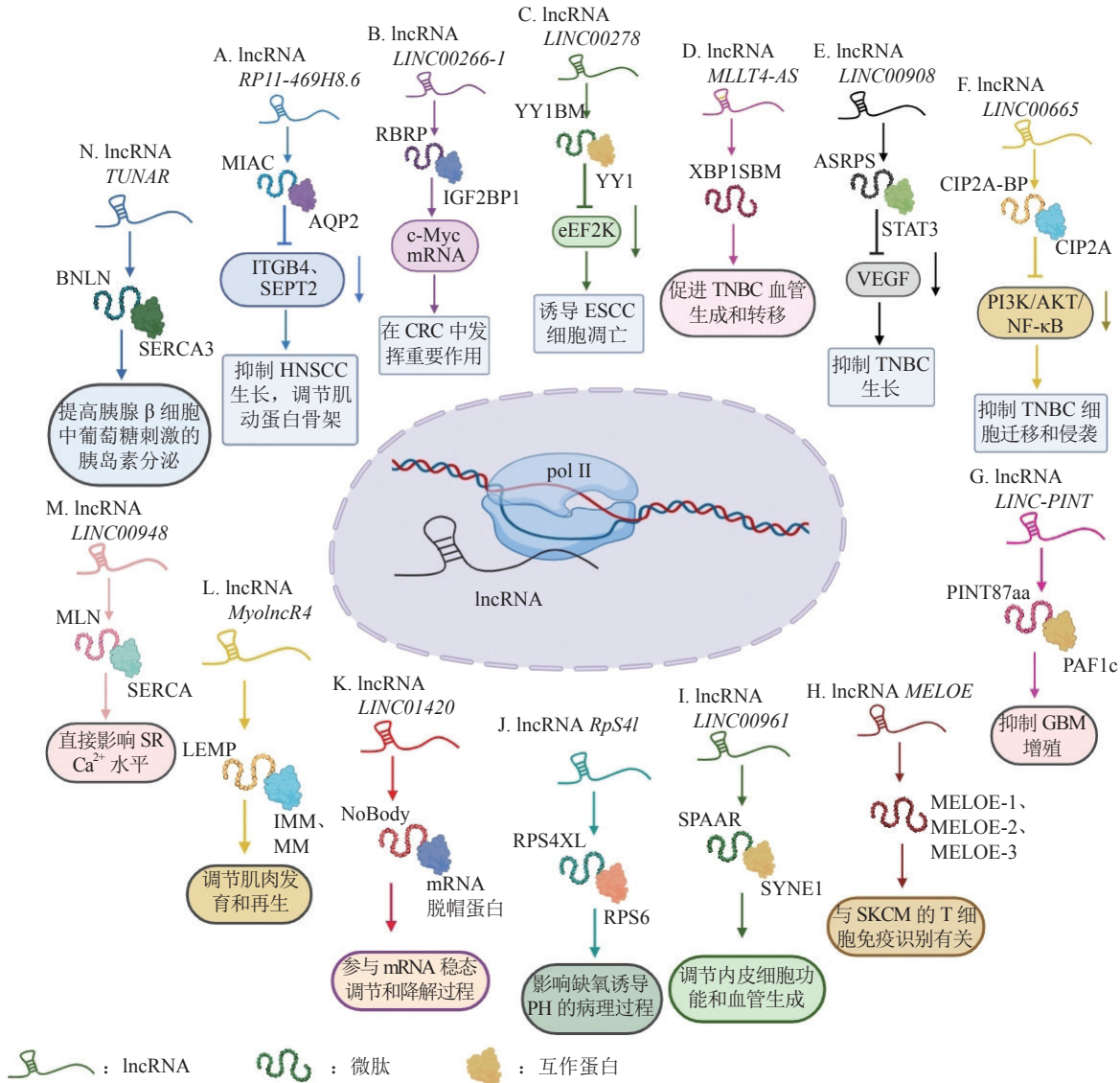
如前所述, lncRNA 编码的微肽在病理生理过程中发挥着广泛而重要的作用。未来, 随着研究的深入, 我们有望发现更多 lncRNA 编码的微肽并揭示其功能, 为疾病的诊断与治疗提供更加精准和有效的策略。图 2 概括了 lncRNA 编码的微肽通过与蛋白质相互作用调控疾病进程及细胞内多种生命活动的复杂网络, 展示了这一研究领域的重要性和发展前景。

4 lncRNA 编码的微肽的定位

随着人们对 lncRNA 编码的微肽研究的持续深化, 大量研究揭示了微肽在多种细胞与组织中的独特分布特征及其功能多样性。Shi 等^[46]通过对 HEK293 细胞的转录组测序 (RNA sequencing, RNA-seq) 联合核糖体印迹测序 (ribosome sequencing, Riboseq), 鉴定出 77 种已知的微肽在膜上高度富集。他们发现, 大约 30% 的微肽具有生物膜富集的趋势,

且与内质网应激反应、肌肉发育和抗病毒免疫反应有关。Aspden 等^[47]利用多聚核糖体测序技术、肽标记转染及免疫印迹鉴定了一系列具有不同亚细胞定位的微肽, 并阐释了各自的功能差异。这些研究

强调了微肽亚细胞定位可能与其特异性功能有重要的联系, 为微肽研究领域开辟了新的探索方向。以下将详细介绍分布在细胞膜、细胞质与细胞器, 以及细胞核等区域的微肽所发挥的具体生物学功能。



A: 微肽 MIAC 抑制 HNSCC 肿瘤生长; B: 微肽 RBRP 在 CRC 中发挥抑癌作用; C: 微肽 YY1BM 诱导 ESCC 细胞凋亡; D: 微肽 XBP1SBM 促进 TNBC 中血管生成; E: 微肽 ASRPS 抑制 TNBC 生长; F: 微肽 CIP2A-BP 抑制 TNBC 细胞迁移和侵袭; G: 微肽 PINT87aa 抑制 GBM 细胞增殖; H: 微肽 MELOE-1、MELOE-2 和 MELOE-3 在皮肤黑色素瘤中的相关作用; I: 微肽 SPAAR 在内皮细胞血管生成中的作用; J: 微肽 RPS41XL 影响缺氧诱导的 PH; K: 微肽 NoBody 参与 mRNA 稳态调节和降解; L: 微肽 LEMP 调节肌肉发育和再生; M: 微肽 MLN 影响肌浆中钙离子水平; N: 微肽 BNLN 调节胰腺 β 细胞的胰岛素分泌

图 2 lncRNA 编码的微肽在病理与生理过程中的调节作用示意图

Figure 2 Schematic diagram of the regulatory roles of lncRNA-encoded micropeptides in pathological and physiological processes

4.1 定位于细胞膜

有研究发现 lncRNA *Gm11549* 可编码一种含有

61 个氨基酸的微肽, 将其命名为 NEMEP^[48]。蛋白质拓扑预测分析表明, NEMEP 是一种单程跨膜蛋白,

研究人员预测发现其在第 7~29 位氨基酸间具有 α -螺旋结构域。同时, 免疫荧光及免疫印迹实验显示在小鼠胚胎干细胞 (mESC) 中 NEMEP 位于细胞膜上。在随后的研究中发现, NEMEP 可以和细胞膜上的葡萄糖转运蛋白 (GLUT) 1 和 3 相互作用并在中胚层分化过程中增加葡萄糖摄取, 维持细胞内的能量代谢。总的来说, 该微肽在细胞膜上的位置和特异性表达使其在调节葡萄糖转运和代谢方面发挥关键作用。

4.2 定位于细胞质与细胞器

4.2.1 细胞质基质 据报道, 人类 lncRNA *MIR155HG* 在炎症树突状细胞 (DC) 中高表达, 且可以编码一种含有 17 个氨基酸的微肽 miPEP155 (P155)^[49]。对该微肽的定位研究表明, 其主要分布于细胞质中, 并与细胞质中热休克蛋白 70 (HSC70) 选择性结合, 从而调节炎症 DC 中主要组织相容性复合体 (MHC) II 类分子呈递, 显著改善小鼠炎症 DC 驱动的自身炎症。此外, P155 能够调节抗原呈递, 可作为治疗自身免疫性疾病的良好候选肽类药物。

4.2.2 内质网 lncRNA *ASHIL-AS1* 可编码一种微肽 APPLE^[50], 对该微肽进行亚细胞分离, 结果显示其在细胞质中的分布比较突出。值得注意的是, 该微肽主要与内质网标志物共定位, 而与其他细胞器标志物无关。研究人员在随后的基因本体 (GO) 分析中发现, APPLE 的相互作用组在“翻译”相关类别中最为丰富, 这与 APPLE 在内质网中富集这一结果共同说明了该微肽可能在翻译过程中起作用。最后, 实验研究证明 APPLE 可以和多聚腺苷酸结合蛋白 (PABPC) 1 相互作用, 促进 mRNA 环化和翻译启动。

Yang 等^[51]发现, lncRNA *LINC00998* 上的 sORF 编码了一种含有 59 个氨基酸和 2 个跨膜螺旋的微肽, 并将其命名为 SMIM30。该微肽是一种新型的跨膜肽。通过膜质分离实验和激光共聚焦实验对 SMIM30 进一步研究发现, SMIM30 与内质网膜和线粒体膜标记物表现出相同的定位模式, 表明该微肽主要位于内质网膜和线粒体膜中。众所周知, 内质网 (ER) 和细胞质之间的钙通量主要通过 ER 膜上的跨膜受体和钙离子泵调节。其中, 肌醇-24, 1, 4-三磷酸受体 (InsP5R) 负责将钙从 ER 释放到细胞质, 而 SERCA 钙泵负责将钙从细胞质输送到内质

网。研究发现, 微肽 SMIM30 可通过 SERCA 起作用, 降低胞质钙水平, 这对于维持细胞内钙稳态至关重要。随后, Shi 等^[46]发现 SMIM30 通过跨膜域 (TM) 靶向线粒体抗病毒信号蛋白 (MAVS), 以干扰其与 RNA 传感器蛋白 RIG-I 的相互作用, 并将 SMIM30 重新命名为抗病毒免疫中的微蛋白 1 (MAVI1)。总的来说, MAVI1 的内质网膜和线粒体膜定位对于其在钙离子稳态调节和抗病毒先天免疫应答中的作用至关重要。

4.2.3 线粒体 *LINC01013* 是一种基因间 lncRNA, 其分布于人脐带血管内皮 (HUVEC) 细胞中, 与关键纤维化基因 *CCN2* 协调表达, 提示其在纤维化中的潜在作用^[52]。研究人员通过 RNA-seq 技术分析发现, 该 lncRNA 具有一定的编码潜力。进一步研究 *LINC01013* ORF 编码微肽的亚细胞定位, 结果表明, 该微肽与线粒体基质标志蛋白明显共定位, 表明 *LINC01013* ORF 编码微肽定位于线粒体基质中。研究数据显示, *LINC01013* 直接与 TGF- β 1 介导的成纤维细胞活化有关, 并且这种效应至少部分是由于 sORF 编码的微肽而介导的。线粒体介导的代谢驱动因素广泛激活成纤维细胞, 通过活性氧激活 p38 丝裂原活化蛋白激酶 (p38MAPK) 和细胞外信号调节激酶 1/2 (ERK1/2) 途径而起作用, 说明 *LINC01013* ORF 编码微肽的促纤维化作用可能是通过整体代谢应激引起的。

研究人员发现^[53], lncRNA *LINC00493* 可被 PABPC4 识别并转移到核糖体中, 从而翻译出含有 95 个氨基酸的微肽, 并将该微肽命名为 SMIM26。SMIM26 在透明细胞肾细胞癌 (ccRCC) 中下调, 与低总生存率相关。蛋白质编码序列分析显示, SMIM26 的 N 端 (第 1~35 位氨基酸) 和 C 端 (第 52~95 位氨基酸) 具有进化保守性。利用 TMHMM 软件进行预测, 可见 SMIM26 的跨膜螺旋片段具有 2 个 α 螺旋 (第 1~12 位氨基酸和第 36~95 位氨基酸) 和 1 个跨膜结构域 (第 13~35 位氨基酸)。随后的激光共聚焦结果显示, SMIM26 与线粒体共定位, 主要位于线粒体外膜。SMIM26 通过其 N 末端与酰基甘油激酶 (AGK) 和谷胱甘肽转运调节因子 (SLC25A11) 相互作用, 来抑制 ccRCC 生长和转

移性肺定植。这种相互作用增加了 AGK 的线粒体定位, 随后抑制 AGK 介导的 AKT 磷酸化。上述研究表明, 线粒体 SMIM26 在线粒体蛋白代谢机制中起着重要作用, 提示 SMIM26 可能成为治疗 ccRCC 的一种潜在选择。

lncRNA *1810058I24Rik* 被 Zheng 等^[54] 证明可编码线粒体定位的微肽 Stmp1。为了确定 Stmp1 的亚细胞器位置, 该课题组从 NIH3T3 细胞中提取线粒体蛋白, 蛋白质印迹显示 Stmp1 与线粒体内膜或线粒体外膜中的蛋白质共定位。在小鼠体内, 微肽 Stmp1 的缺乏可以抑制小胶质细胞和炎症小体 Nlrp2 的活性, 从而保护视网膜神经节细胞免受视网膜缺血/再灌注损伤 (IR) 损伤。此外, 在小鼠中敲除或在原发性小胶质细胞中敲低 Stmp1 可以促进线粒体融合, 导致线粒体膜电位受损, 活性氧 (ROS) 产生减少, 从而使有氧糖酵解减少, 炎症减轻。同时, Stmp1 通过抑制代谢传感器解偶联蛋白 2 (Ucp2) 的表达和激活 Ca^{2+} 的表达来控制线粒体功能, 从而触发炎症小体 Nlrp2 激活, 为视网膜 IR 的分子机制提供了新的线索, 并可能为 IR 相关的神经退行性疾病提供新的治疗思路。

4.2.4 外泌体 Jiang 等^[55] 发现 lncRNA *LINC02381* 在 GBM 中高表达, 预测该 lncRNA 可编码相应的微肽, 通过 smORFunction (一种用于预测 sORF 和微肽功能的工具) 和 iLoc-lncRNA (一种用于预测 lncRNA 的亚细胞位置的工具) 对 *LINC02381* 编码微肽的位置信息和序列信息进行分析, 结果表明 *LINC02381* 位于细胞质中, 其编码的微肽位于外泌体中。进一步研究发现, *LINC02381* 编码的微肽与溶质载体家族 2 成员 10 (SLC2A10) 相关, 有可能调控铁死亡。*LINC02381* 编码的微肽在外泌体中表达丰富且稳定, 可以调节 GBM 的细胞增殖、侵袭、血管生成、免疫逃逸和耐药性, 有望成为 GBM 新的生物标志物。

4.3 定位于细胞核

Tormini 等^[56] 发现 *Linc-mipep* 和 *Linc-wrb* 分别编码大小为 87 和 93 个氨基酸的微肽, 并命名为 hmg1a 和 hmg1b。该课题组进一步发现, 这 2 种微肽富集在细胞核中, 提示其可能与细胞核的一些功能有关。随后的研究表明, 这 2 种微肽的缺失对

染色质可及性有影响, 特别是突变体中的染色质可及性在谷氨酸电离受体 *N*-甲基-*D*-天冬氨酸 (NMDA) 受体亚基 1 等区域发生改变, 从而改变了多个转录因子 (TF) 结合位点的可及性, 这进一步说明该微肽的核定位对其相关功能的实现具有重要作用, 为后续机制探究提供了方向。

Mitoregulin (MOXI) 是一种高度保守的微肽, 由 56 个氨基酸组成, 由人 lncRNA *LINC00116* 和小鼠 lncRNA *1500011K16Rik* 编码, 最初被证实存在于小鼠心脏和骨骼肌组织的线粒体内膜中, 与线粒体三功能蛋白结合, 可增强脂肪酸 β 氧化和呼吸效率。然而, Li 等^[57] 还发现 MOXI 对肾纤维化具有调节作用, 揭示了 MOXI 的全新功能, 即易位到细胞核中调控基因转录。进一步研究表明, MOXI 的 T49 位点磷酸化对其核积累至关重要, 可促进 MOXI 与 *N*-乙酰转移酶 14 (Nat14) 和原癌基因蛋白 (c-Jun) 形成复合物, 从而促进 α -1 型胶原蛋白基因 (*Colla1*, 一种纤维化相关基因) 转录。同时, 在成纤维细胞中进行的研究还发现, 用 TGF- β 1 处理细胞后的 12 h 内, MOXI 在线粒体中显著增加, 但在 24 h 后, 其核积累变得明显, 线粒体中 MOXI 轻微下降。MOXI 在细胞核和线粒体之间的双向通信, 为肾纤维化的研究提供了重要思路。

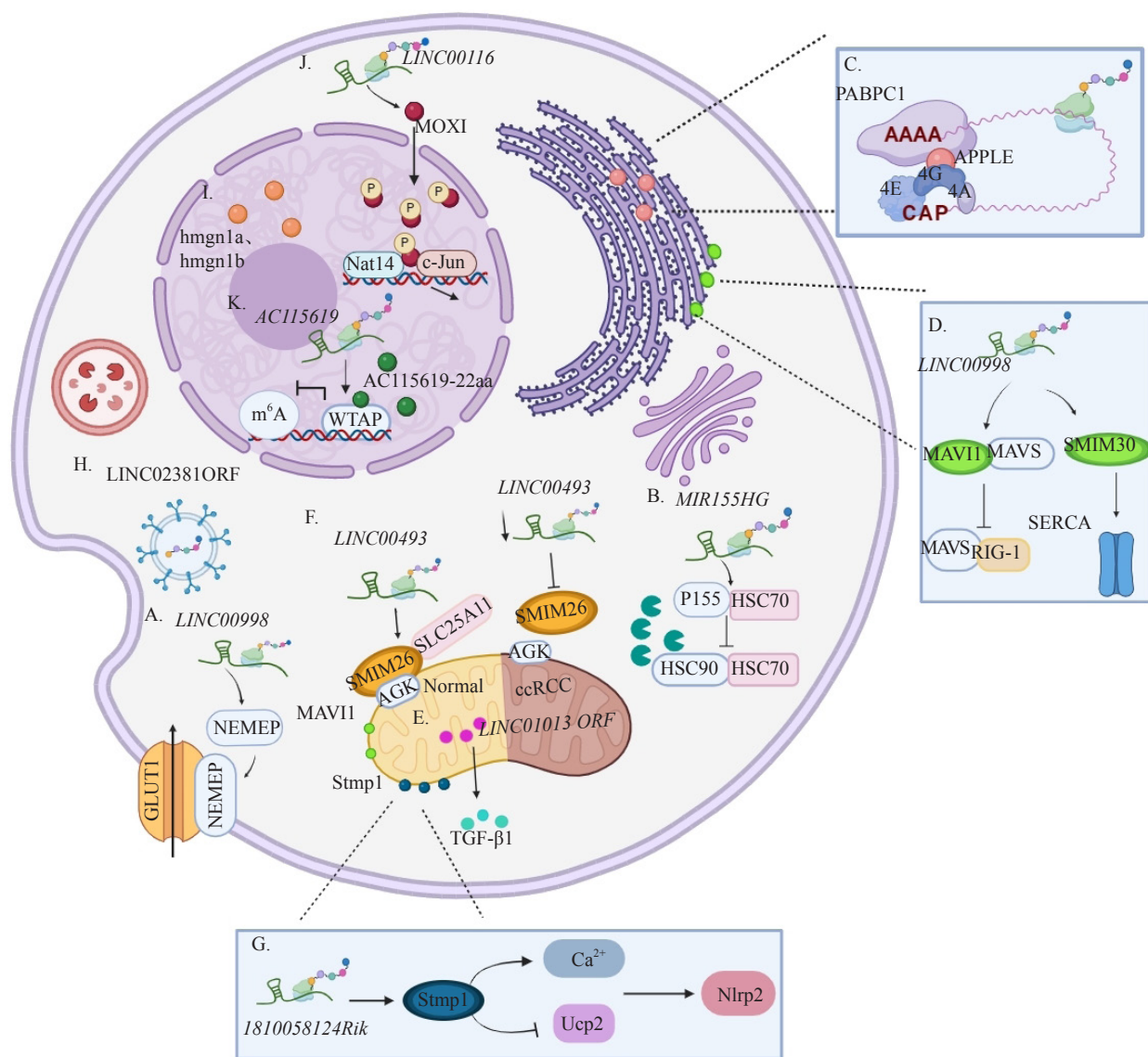
lncRNA *AC115619* 在肝细胞癌 (HCC) 中低表达, 并编码一种微肽 AC115619-22aa^[58]。免疫荧光染色显示, AC115619-22aa 与 m⁶A RNA 修饰相关蛋白 WTAP 在细胞核内共定位, 提示该微肽可能在细胞核中发挥相关功能。m⁶A RNA 修饰是一种常见的 RNA 修饰形式, 其在细胞核中发生后, 被修饰的 RNA 分子可通过核孔复合物被转运到细胞质中, 进一步影响 RNA 的稳定性、转录后调控及翻译过程等。细胞核作为转录和 RNA 修饰的主要地点, 对 m⁶A 修饰的调控起着重要作用, 而蛋白 WTAP 是 m⁶A 甲基转移酶复合物的核心成员之一, 可以调节这一修饰反应的催化活性和特异性。对 AC115619-22aa 的研究发现, 其通过与 WTAP 结合来阻碍 m⁶A 甲基转移酶复合物的组装, 从而抑制 HCC 的进展。

4.4 小结

如前所述, lncRNA 编码的微肽在细胞内的亚

细胞定位模式多样, 且可通过复杂精细的机制调控细胞活动 (见图 3)。未来可对微肽的亚细胞定位

及生物学功能进一步研究, 进而为疾病的精准治疗提供新思路。



A: 微肽 NEMEP 位于质膜上, 可以和细胞膜上的 GLUT 相互作用; B: 微肽 P155 定位于细胞质中, 可改善小鼠 DC 驱动的自身炎症; C: 微肽 APPLE 定位于内质网中并参与 mRNA 翻译过程; D: 微肽 SMIM30 定位于内质网及线粒体膜上, 在抗病毒免疫应答中发挥作用; E: *LINC01013* 编码的微肽位于线粒体基质中, 与 TGF- β 1 介导的成纤维细胞活化有关; F: 微肽 SMIM26 定位于线粒体外膜, 可以抑制 ccRCC 生长和转移性肺定植; G: 微肽 Stmp1 位于线粒体外膜, 可以改善 Nlrp2 相关炎症; H: 由 lncRNA *LINC02381* 编码的微肽位于外泌体中, 可调节 GBM 的细胞增殖、侵袭、血管生成等; I: 微肽 hmgn1a 和 hmgn1b 位于细胞核, 对染色质可及性有一定影响; J: 微肽 MOXI 在细胞核中促进基因转录; K: 微肽 AC115619-22aa 在细胞核中参与 RNA 的修饰

图 3 lncRNA 编码的微肽在亚细胞结构中的不同定位及调节作用

Figure 3 Different localization and regulatory roles of lncRNA-encoded micropeptides in subcellular structures

5 lncRNA 编码的微肽的研究方法

过去, 鉴定 lncRNA 编码的微肽及其内在特性的技术有限, 使得 lncRNA 编码的微肽的发现进程

显著受阻。然而, 近年来, 生物技术与工具的日新月异已将 lncRNA 编码的微肽的研究推向了新的热潮, 为发掘新型可编码微肽提供了强有力的支撑与

手段, 使得 lncRNA 编码的微肽逐渐获得了学界的广泛接纳与认可^[3,11]。以下将介绍当前微肽研究中常用的关键技术与方法。

5.1 生物信息学

生物信息学的核心研究领域涵盖基因组学 (Genomics) 和蛋白质组学 (Proteomics), 专注于解析核酸与蛋白质序列中蕴含的结构与功能信息, 已成为生物学研究中不可或缺的手段与工具^[59]。通过生物信息学的方法可以将目标元素与随机出现的噪声区分开来, 其中, 序列保守性和序列相似性是常用的参数^[60-61]。序列相似性搜索通过比对待预测微肽与已知蛋白质的结构域, 揭示其编码潜力和潜在功能。常用的预测软件有 sORFfinder^[62]、HAltORF^[63]、PhyloCSF^[64] 等。编码-非编码识别工具 (CNIT)^[65] 仅凭借对固有序列组成的分析进行分类, 无需依赖注释信息, 展现出跨物种应用的潜力。新开发的 MicroPeptide (MiPepid)^[66] 工具基于微生物蛋白作为训练集的逻辑回归模型来预测 sORF。支持向量机 (SVM) 分类器与编码潜力预测器 (CCPred) 通过整合蛋白质折叠预测因子等附加特征, 进一步增强了预测能力^[61]。尽管生物信息学技术的持续进步极大地推动了相关领域的探索, 但在准确预测和注释 ssORF 等方面遇到的难题, 仍凸显了这些信息的复杂性与挑战^[66]。

5.2 核糖体印迹测序分析

Ribo-seq 是一种新兴的实验技术, 它结合了高通量测序技术, 可以在全基因组水平上监测蛋白质的翻译水平, 为探索细胞内蛋白质翻译调控机制开辟了新途径。Ribo-seq 可以提供具有单核苷酸分辨率的全基因组主动翻译快照^[61], 显著推动了 sORF 的检测流程, 并在该领域展现出至关重要的作用。目前 Ribo-seq 建库测序的原理是对细胞使用翻译抑制剂, 使活跃翻译的核糖体锚定于 mRNA 序列或起始密码子上^[66-67]。随后, 将细胞裂解并加入 RNA 酶, 消化掉未受核糖体保护的 mRNA 区域, 分离单个核糖体并纯化其上保留的、未被消化的 mRNA 短片段, 进行建库测序和相应的数据分析。尽管这些工具用于 sORF 检测的效能仍有待研究, 但深度学习算法的应用为基于 Ribo-seq 数据的 sORF 检测策略的优化提供了可能。

5.3 转录组测序技术

RNA-seq 作为一种高通量测序技术, 可对 mRNA、miRNA 和 lncRNA 等进行深度测序分析, 从而反映出它们的表达水平, 为翻译过程提供证据^[61]。这项技术在转录本结构解析、非编码区域功能探索、基因表达水平量化和全新转录本发掘等多个研究领域均展示出较大的优势。RNA-seq 能够对每个转录本片段序列进行直接测定, 即使是只有拷贝数极低的稀有转录本也能够检测到。此外, RNA-seq 技术适用于任何物种的全基因组分析, 并能检测未知基因, 从而揭示新的转录本, 因此其在 lncRNA 和微肽的研究中, 也是一种十分重要的工具。

5.4 液相色谱-串联质谱

通过液相色谱-串联质谱 (LC-MS/MS) 分析技术, 我们可得到目标样品的质谱图, 将其与已知蛋白数据库中的理论质谱图进行比对, 即可鉴定出由 sORF 编码但尚未被注释的微肽。该方法可以直接检测到基因编码的产物, 因此一直被认为是微肽研究领域的“黄金标准”^[68]。尽管 Ribo-seq 技术为 sORF 研究领域带来了革命性的变化, 但其所提供的翻译证据并不一定意味着功能性微肽的产生。因此, 基于 LC-MS/MS 技术对翻译产物进行鉴定, 从而获得一组高置信度的 sORF 仍然至关重要^[61]。然而, 微肽的长度较短^[69], 表达水平较低, 在 MS 样品制备过程中这些小的翻译产物可能面临降解风险, 为微肽的质谱鉴定增加了难度, 亟需开发更为灵敏的检测方法^[70]。

5.5 基于成簇的规律间隔的短回文重复序列及其相关蛋白 9 的基因编辑技术

成簇的规律间隔的短回文重复序列 (CRISPR) 筛选技术是一种大规模的基因功能丧失实验方法, 有助于揭示特定细胞类型中特定功能或表型的关键基因或基因序列。Cas9 是 CRISPR 相关核酸酶, CRISPR/Cas9 是最新出现的一种由 RNA 指导的, 利用 Cas9 核酸酶对靶向基因进行编辑的技术。该项技术通过在 DNA 中的特定目标位点进行双链切割, 诱导细胞利用其 DNA 修复机制进行修复, 是一种有针对性且实用的方法, 可以确定 sORF 的编码潜力。尽管该技术无法直接预测 sORF, 并且需要事

先对其核苷酸序列进行预测分析, 但它是验证蛋白质表达、亚细胞定位和功能的独特工具^[71]。此外, CRISPR-Cas9 使功能丧失方法成为可能, 可通过分析缺乏目标基因后系统的表型变化, 定义该基因的生物学功能。然而, 一些微生物蛋白并不总能产生可检测的表型, 这在一定程度上削弱了该方法的优势^[61]。

综上所述, 在 lncRNA 编码的微肽的研究中有多种方法可提供有效证据, 通常使用的描绘潜在 sORF 的高通量方法可归类为生物信息学分析、测序策略和质谱 (MS) 分析。生物信息学方法使用系统

发育保守特征作为功能预测的首要证据; 而通过核糖体分析 (Ribo-seq) 或 RNA-seq 进行测序, 提供了翻译证据; LC-MS/MS 可提供翻译产物的直接证据。

6 结语与展望

在过去数十年间, 众多研究不断揭示出 lncRNA 在转录组中蕴藏的可翻译的 sORF。这些 sORF 可以编码小于 100 个氨基酸的微肽, 并在多种生物进程中发挥着不可或缺的作用。本文回顾了当前与病理功能和生理作用紧密相关的 lncRNA 编码微肽的研究 (见表 1)。

表 1 lncRNA 编码的微肽及其功能

Table 1 The lncRNA-encoded micropeptides and their functions

lncRNA 名称	lncRNA 编码微肽名称	氨基酸数量	微肽功能	参考文献
<i>RP11-469H8.6</i>	MIAC	51	抑制肿瘤细胞运动、增殖、迁移和肿瘤生长	[12]
<i>LINC00266-1</i>	RBRP	71	调节 m ⁶ A 阅读器对靶 RNA 和肿瘤发生的识别	[8]
<i>LINC00278</i>	YY1BM	21	在营养剥夺 (ND) 下诱导 ESCC 细胞凋亡	[26]
<i>MLLT4-AS1</i>	XBP1SBM	21	改善 TNBC 中的 Gln 水平, 以促进 TNBC 血管生成和转移	[29]
<i>LINC00908</i>	ASRPS	60	降低了 TNBC 中 VEGF 的表达	[30]
<i>LINC00665</i>	CIP2A-BP	52	抑制 BC 细胞的迁移和侵袭	[31]
<i>LINC-PINT</i>	PINT87aa	87	抑制 GBM 细胞增殖	[32]
<i>MELOE</i>	MELOE-1	39	与黑色素瘤的 T 细胞免疫识别有关	[34,36]
	MELOE-2	44		[35-36]
	MELOE-3	54		[33]
<i>LINC00961</i>	SPAAR	75	有助于成年期的生长发育, 并维持基础心血管功能, 从而降低心肌梗死的风险	[37-38]
<i>Rps4l</i>	RPS4XL	未披露	可抑制肺动脉平滑肌细胞增殖并减少其诱导的 PH 死亡	[40-41]
<i>LINC01420</i>	NoBody	68	参与 mRNA 稳态调节和降解过程	[42]
<i>MyolncR4</i>	LEMP	56	促进小鼠肌肉形成和再生, 并与多种线粒体蛋白相关	[44]
<i>LINC00948</i>	MLN	46	MLN 与 SR 膜中的 SERCA 形成稳定的复合物, 且 MLN 可直接影响 SR Ca ²⁺ 水平和最大运动表现	[45]
<i>TUNAR</i>	BNLN	48	降低了胰腺 β 细胞中内质网钙水平, 并提高了葡萄糖刺激的胰岛素分泌	[3]
<i>Gm11549</i>	NEMEP	61	和细胞膜上的 GLUT 相互作用并在中胚层分化过程中增加葡萄糖摄取, 维持细胞内的能量代谢	[48]
<i>MIR155HG</i>	miPEP155	17	调节炎性 DC 中 MHC II 类分子递呈, 显著改善小鼠 DC 驱动的自身炎症	[49]
<i>ASH1L-AS1</i>	APPLE	90	和 PABPC1 相互作用, 促进 mRNA 环化和翻译启动	[50]
<i>LINC00998</i>	SMIM30	59	通过 SERCA 起作用, 降低胞质钙水平, 从而维持细胞内钙稳态	[51]
<i>LINC00998</i>	MAV1	59	靶向线粒体上的线粒体抗病毒信号蛋白 MAVS, 以干扰其与 RNA 传感器蛋白 RIG-I 的相互作用	[46]
<i>LINC01013</i>	LINC01013ORF	未披露	与 TGF-β1 介导的成纤维细胞活化有关	[52]
<i>LINC00493</i>	SMIM26	95	SMIM26 通过其 N 末端与 AGK 和谷胱甘肽转运调节因子 SLC25A11 相互作用, 从而抑制 ccRCC 生长和转移性肺定植	[53]

续表 1

lncRNA 名称	lncRNA 编码微肽名称	氨基酸数量	微肽功能	参考文献
1810058124Rik	Stmp1	未披露	敲低该微肽可促进线粒体融合, 减少有氧糖酵解, 从而减轻炎症	[54]
LINC02381	LINC02381-aa	未披露	调节 GBM 的细胞增殖、侵袭、血管生成、免疫逃逸和耐药性	[55]
Linc-mipep	hmgn1a	87	敲除微肽后对染色质可及性产生影响	[56]
Linc-wrb	hmgn1b	93		
LINC00116	MOXI	56	在细胞核中调控基因转录, 且对肾纤维化具有调节作用	[57]
AC115619	AC115619-22aa	22	与 WTAP 结合, 阻碍 m ⁶ A 甲基转移酶复合物的组装, 从而抑制 HCC 进展	[58]

总的来说, 一些微肽在 HNSCC、CRC、食管癌和 BC 等多种癌症中都表现出特异性表达, 并发挥一定的促癌或抑癌作用。同时, 部分微肽在心血管疾病、mRNA 降解、肌肉发育等过程中也发挥着特定的调节作用。此外, 本文还归纳了不同细胞定位的微肽与其功能之间的联系。例如, 定位于内质网的微肽 APPLE 能够促进 mRNA 的环化和翻译启动; 定位于细胞膜的微肽 NEMEP 能够与细胞膜上的 GLUT 相互作用, 在中胚层分化过程中增加葡萄糖摄取; 而定位于细胞核的微肽 AC115619-22aa 则影响着 RNA 的稳定性、转录后调控及翻译过程等。综上所述, 这些微肽在细胞中所发挥的功能与其在亚细胞结构中的定位密切相关, 对微肽的亚细胞结构定位的研究有助于我们更准确、迅速地发现微肽在生命过程中所扮演的角色。本文亦介绍了微肽研究中常用的实验方法。随着高通量转录组测序、生物信息学技术和多种生物技术的革新和发展, 越来越多的实验方法被应用于微肽研究领域, 通过这些方法可筛选出更多具有编码能力的 lncRNA, 进而揭示其在疾病发生发展过程中的作用机制和功能。

微肽为肿瘤等多种疾病的预防和治疗提供了新的靶点和生物标志物, 其在生理病理过程中的调节作用也展现出在临床诊断和治疗方面的巨大应用潜力。目前, 在肿瘤治疗领域, 尽管微肽尚未真正应用于临床, 但多项研究均已经证实, 将抑癌性微肽直接注射到小鼠肿瘤模型中或者特异性敲除致癌微肽, 均可有效抑制肿瘤进展。例如, 在 BC 模型小鼠中进行的研究已证实, 微肽 CIP2A-BP-55aa 和 ASRPS-60aa 能够抑制肿瘤的肺转移和血管生成^[72]。此外, 自 100 年前胰岛素被发现以来, 多肽作为后

基因时代的热点, 已被广泛应用于药物开发领域。与化疗药物相比, 肽类药物具有特异性和活性高、细胞毒性小和免疫原性低等优点, 而内源性多肽为肽类药物的研发提供了丰富的天然资源宝库。微肽作为一类新兴的内源性多肽, 在许多生理病理过程中具有不可或缺的作用, 以微肽为先导化合物进行药物的设计、优化与成药性评价也是新药研发的重要方向。同时, 一些微肽如 RBRP、XBP1SBM、MELOE-1、MELOE-2 和 MELOE-3 在肿瘤组织中特异性表达, 且与肿瘤预后密切相关, 可作为新的肿瘤生物标志物, 应用于体液检测等早期诊断技术的开发, 为癌症的早期诊断、治疗监测及预后评估开辟了新的研究方向, 有望推动癌症研究和精准医疗的发展^[73]。

尽管研究表明微肽对多种疾病具有抑制或促进作用, 并有望成为疾病治疗的靶点, 但仍需进一步的实验研究来揭示其在疾病中的具体作用机制以及其在体内的分布和代谢情况。微肽到达肿瘤部位并通过细胞膜进入肿瘤细胞的能力仍有待阐明, 这些都是微肽研究亟需解决的问题^[74]。微肽所涉及的分子机制及信号通路不仅可为疾病治疗提供新的思路, 还有望打破传统的蛋白质编码机制, 从而完善和丰富中心法则, 这将是未来的主要研发方向。此外, 微肽在细胞中的分布和定位也有着显著的特异性, 并表现出与功能的相关性, 这也提示了微肽功能研究的方向和思路。需强调的是, 由于微肽的分子片段较短且表达丰度低^[72], 如何从基因组数以万计的 sORF 中准确识别编码蛋白的 sORF, 并将其与真正的转录“噪音”区分开来, 仍是一大挑战。验证 lncRNA 编码能力也是后续微肽功能研究的前提^[73]。因此, 开发用于识别翻译序列的新工具和新方法,

以及对现有算法和工具的改进, 将有助于完成对不同物种翻译组的表征。同时, 多种预测和实验手段的综合运用也将进一步提高微肽识别与验证的效率

及准确率。有理由相信, lncRNA 编码微肽相关研究的不断深入, 将会为生命医学领域中诸多悬而未决的难题提供新的解决思路。

【参考文献】

- [1] Beermann J, Piccoli M T, Viereck J, *et al.* Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches[J]. *Physiol Rev*, 2016, 96(4): 1297–1325.
- [2] Zhang Q, Wu E, Tang Y, *et al.* Deeply mining a universe of peptides encoded by long noncoding RNAs[J]. *Mol Cell Proteomics*, 2021, 20: 100109. DOI:10.1016/j.mcpro.2021.100109.
- [3] Li M, Shao F, Qian Q, *et al.* A putative long noncoding RNA-encoded micropeptide maintains cellular homeostasis in pancreatic β -cells[J]. *Mol Ther Nucleic Acids*, 2021, 26: 307–320.
- [4] Batista P J, Chang H Y. Long noncoding RNAs: cellular address codes in development and disease[J]. *Cell*, 2013, 152(6): 1298–1307.
- [5] Li J H, Liu S, Zhou H, *et al.* starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data[J]. *Nucleic Acids Research*, 2014, 42: D92–D97.
- [6] Jalali S, Bhartiya D, Lalwani M K, *et al.* Systematic transcriptome wide analysis of lncRNA-miRNA interactions[J]. *PLoS One*, 2013, 8(2): e53823. DOI:10.1371/journal.pone.0053823.
- [7] Geisler S, Collier J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts[J]. *Nat Rev Mol Cell Biol*, 2013, 14(11): 699–712.
- [8] Zhu S, Wang J Z, Chen D, *et al.* An oncopeptide regulates m6A recognition by the m6A reader IGF2BP1 and tumorigenesis[J]. *Nat Commun*, 2020, 11(1): 1685. DOI:10.1038/s41467-020-15403-9.
- [9] Housman G, Ulitsky I. Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs[J]. *Biochim Biophys Acta*, 2016, 1859(1): 31–40.
- [10] Maria A. The hidden world of non-canonical ORFs[J]. *Exp Cell Res*, 2020, 396(2): 112267. DOI:10.1016/j.yexcr.2020.112267.
- [11] Makarewich C A, Olson E N. Mining for micropeptides[J]. *Trends Cell Biol*, 2017, 27(9): 685–696.
- [12] Li M, Li X, Zhang Y, *et al.* Micropeptide MIAC inhibits HNSCC progression by interacting with aquaporin 2[J]. *J Am Chem Soc*, 2020, 142(14): 6708–6716.
- [13] Nelson C. Small but important[J]. *Lab Anim*, 2018, 47(8): 207. DOI:10.1038/s41684-018-0122-8.
- [14] Setrerrahmane S, Li M, Zoghbi A, *et al.* Cancer-related micropeptides encoded by ncRNAs: promising drug targets and prognostic biomarkers[J]. *Cancer Lett*, 2022, 547: 215723. DOI:10.1016/j.canlet.2022.215723.
- [15] Chen L L. Linking long noncoding RNA localization and function[J]. *Trends Biochem Sci*, 2016, 41(9): 761–772.
- [16] Kageyama Y, Kondo T, Hashimoto Y. Coding vs non-coding: translatability of short ORFs found in putative non-coding transcripts[J]. *Biochimie*, 2011, 93(11): 1981–1986.
- [17] Derrien T, Johnson R, Bussotti G, *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression[J]. *Genome Res*, 2012, 22(9): 1775–1789.
- [18] Mattick J S, Amaral P P, Carninci P, *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations[J]. *Nat Rev Mol Cell Biol*, 2023, 24(6): 430–447.
- [19] McHugh C A, Chen C K, Chow A, *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3[J]. *Nature*, 2015, 521(7551): 232–236.
- [20] Wu H, Yin Q F, Luo Z, *et al.* Unusual processing generates SPA lncRNAs that sequester multiple RNA binding proteins[J]. *Mol Cell*, 2016, 64(3): 534–548.
- [21] Dimitrova N, Zamudio J R, Jong R M, *et al.* lincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G₁/S checkpoint[J]. *Mol Cell*, 2014, 54(5): 777–790.
- [22] Rohrig H, Schmidt J, Miklashevichs E, *et al.* Soybean ENOD40 encodes two peptides that bind to sucrose synthase[J]. *Proc Natl Acad Sci USA*, 2002, 99(4): 1915–1920.
- [23] Van de Sande K, Pawlowski K, Czaja I, *et al.* Modification of phytohormone response by a peptide encoded by ENOD40 of legumes and a nonlegume[J]. *Science*, 1996, 273(5273): 370–373.
- [24] Crappé J, Crieckinge W V, Menschaert G. Little things make big things happen: a summary of micropeptide encoding genes[J]. *EuPA Open Proteom*, 2014, 3: 128–137.
- [25] Wang J, Zhu S, Meng N, *et al.* ncRNA-encoded peptides or proteins and cancer[J]. *Mol Ther*, 2019, 27(10): 1718–1725.
- [26] Wu S, Zhang L, Deng J, *et al.* A novel micropeptide encoded by Y-linked LINC00278 links cigarette smoking and AR signaling in male esophageal squamous cell carcinoma[J]. *Cancer Res*, 2020, 80(13): 2790–2803.

- [27] Xu B, Yang H J, Zhang X P, *et al.* Research progresses in roles of lncRNA and its relationships with breast cancer[J]. *Cancer Cell Int*, 2018, 18(1): 179. DOI:10.1186/s12935-018-0674-0.
- [28] Tomar D, Yadav A S, Kumar D, *et al.* Non-coding RNAs as potential therapeutic targets in breast cancer[J]. *Biophys Acta Gene Regul Mech*, 2020, 1863(4): 194378. DOI:10.1016/j.bbagrm.2019.04.005.
- [29] Wu S, Guo B, Zhang L, *et al.* A micropeptide XBP1SBM encoded by lncRNA promotes angiogenesis and metastasis of TNBC via XBP1s pathway[J]. *Oncogene*, 2022, 41(15): 2163–2172.
- [30] Wang Y, Wu S, Zhu X, *et al.* LncRNA-encoded polypeptide ASRPS inhibits triple-negative breast cancer angiogenesis[J]. *J Exp Med*, 2020, 217(3): jem.20190950. DOI:10.1084/jem.20190950.
- [31] Guo B, Wu S, Zhu X, *et al.* Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression[J]. *EMBO J*, 2020, 39(1): e102190. DOI:10.15252/embj.2019102190.
- [32] Zhang M, Zhao K, Xu X, *et al.* A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma[J]. *Nat Commun*, 2018, 9(1): 4475. DOI:10.1038/s41467-018-06862-2.
- [33] Charpentier M, Croyal M, Carbonnelle D, *et al.* IRES-dependent translation of the long non coding RNA meloe in melanoma cells produces the most immunogenic MELOE antigens[J]. *Oncotarget*, 2016, 7(37): 59704–59713.
- [34] Godet Y, Moreau-Aubry A, Guilloux Y, *et al.* MELOE-1 is a new antigen overexpressed in melanomas and involved in adoptive T cell transfer efficiency[J]. *J Exp Med*, 2008, 205(11): 2673–2682.
- [35] Godet Y, Moreau-Aubry A, Mompelat D, *et al.* An additional ORF on meloe cDNA encodes a new melanoma antigen, MELOE-2, recognized by melanoma-specific T cells in the HLA-A2 context[J]. *Cancer Immunol Immunother*, 2010, 59(3): 431–439.
- [36] Carbonnelle D, Vignard V, Sehedic D, *et al.* The melanoma antigens MELOE-1 and MELOE-2 are translated from a bona fide polycistronic mRNA containing functional IRES sequences[J]. *PLoS One*, 2013, 8(9): e75233. DOI:10.1371/journal.pone.0075233.
- [37] Spencer H L, Sanders R, Boulberdaa M, *et al.* The LINC00961 transcript and its encoded micropeptide, small regulatory polypeptide of amino acid response, regulate endothelial cell function[J]. *Cardiovasc Res*, 2020, 116(12): 1981–1994.
- [38] Spiroski A M, Sanders R, Meloni M, *et al.* The influence of the LINC00961/SPAAR locus loss on murine development, myocardial dynamics, and cardiac response to myocardial infarction[J]. *Int J Mol Sci*, 2021, 22(2): 969. DOI:10.3390/ijms22020969.
- [39] Li Y, Zhang J, Sun H, *et al.* lnc-Rps4l-encoded peptide RPS4XL regulates RPS6 phosphorylation and inhibits the proliferation of PASMCS caused by hypoxia[J]. *Mol Ther*, 2021, 29(4): 1411–1424.
- [40] Hu L, Wang J, Huang H, *et al.* YTHDF1 regulates pulmonary hypertension through translational control of MAGED1[J]. *Am J Respir Crit Care Med*, 2021, 203(9): 1158–1172.
- [41] Liu Y, Zhang H, Li Y, *et al.* Long noncoding RNA Rps4l mediates the proliferation of hypoxic pulmonary artery smooth muscle cells[J]. *Hypertension*, 2020, 76(4): 1124–1133.
- [42] D'Lima N G, Ma J, Winkler L, *et al.* A human microprotein that interacts with the mRNA decapping complex[J]. *Nat Chem Biol*, 2017, 13(2): 174–180.
- [43] Zhu M, Liu J, Xiao J, *et al.* Lnc-mg is a long non-coding RNA that promotes myogenesis[J]. *Nat Commun*, 2017, 8: 14718. DOI:10.1038/ncomms14718.
- [44] Wang L, Fan J, Han L, *et al.* The micropeptide LEMP plays an evolutionarily conserved role in myogenesis[J]. *Cell Death Dis*, 2020, 11(5): 357. DOI:10.1038/s41419-020-2570-5.
- [45] Anderson D M, Anderson K M, Chang C L, *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance[J]. *Cell*, 2015, 160(4): 595–606.
- [46] Shi T T, Huang Y, Li Y, *et al.* MAV11, an endoplasmic reticulum-localized microprotein, suppresses antiviral innate immune response by targeting MAVS on mitochondrion[J]. *Sci Adv*, 2023, 9(35): eadg7053. DOI:10.1126/sciadv.adg7053.
- [47] Aspden J L, Eyre-Walker Y C, Phillips R J, *et al.* Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq[J]. *Elife*, 2014, 3: e03528. DOI:10.7554/eLife.03528.
- [48] Fu H, Wang T, Kong X, *et al.* A nodal enhanced micropeptide NEMEP regulates glucose uptake during mesendoderm differentiation of embryonic stem cells[J]. *Nat Commun*, 2022, 13(1): 3984. DOI:10.1038/s41467-022-31762-x.
- [49] Niu L, Lou F, Sun Y, *et al.* A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation[J]. *Sci Adv*, 2020, 6(21): eaaz2059. DOI:10.1126/sciadv.aaz2059.
- [50] Sun L, Wang W, Han C, *et al.* The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation[J]. *Mol Cell*, 2021, 81(21): 4493–4508.
- [51] Yang J E, Zhong W J, Li J F, *et al.* LINC00998-encoded micropeptide SMIM30 promotes the G₁/S transition of cell cycle by regulating cytosolic calcium level[J]. *Mol Oncol*, 2023, 17(5): 901–916.
- [52] Quaipe N M, Chothani S, Schulz J F, *et al.* LINC01013 is a determinant of fibroblast activation and encodes a novel fibroblast-activating micropeptide[J]. *J Cardiovasc Transl Res*, 2023, 16(1): 77–85.
- [53] Meng K, Lu S, Li Y Y, *et al.* LINC00493-encoded microprotein

- SMIM26 exerts anti-metastatic activity in renal cell carcinoma[J]. *EMBO Rep*, 2023, 24(6): e56282. DOI:10.15252/embr.202256282.
- [54] Zheng X, Wang M, Liu S, *et al.* A lncRNA-encoded mitochondrial micropeptide exacerbates microglia-mediated neuroinflammation in retinal ischemia/reperfusion injury[J]. *Cell Death Dis*, 2023, 14(2): 126. DOI:10.1038/s41419-023-05617-2.
- [55] Jiang L, Yang J, Xu Q, *et al.* Machine learning for the micropeptide encoded by LINC02381 regulates ferroptosis through the glucose transporter SLC2A10 in glioblastoma[J]. *BMC Cancer*, 2022, 22(1): 882. DOI:10.1186/s12885-022-09972-9.
- [56] Tomini V A, Miao L, Lee H J, *et al.* linc-mipep and linc-wrb encode micropeptides that regulate chromatin accessibility in vertebrate-specific neural cells[J]. *Elife*, 2023, 12: e82249. DOI:10.7554/eLife.82249.
- [57] Li J, Qu X, Guan C, *et al.* Mitochondrial micropeptide MOXI promotes fibrotic gene transcription by translocation to the nucleus and bridging N-acetyltransferase 14 with transcription factor c-Jun[J]. *Kidney Int*, 2023, 103(5): 886–902.
- [58] Zhang Q, Wei T, Yan L, *et al.* Hypoxia-responsive lncRNA AC115619 encodes a micropeptide that suppresses m⁶A modifications and hepatocellular carcinoma progression[J]. *Cancer Res*, 2023, 83(15): 2496–2512.
- [59] Jöchl C, Rederstorff M, Hertel J, *et al.* Small ncRNA transcriptome analysis from *Aspergillus fumigatus* suggests a novel mechanism for regulation of protein synthesis[J]. *Nucleic Acids Res*, 2008, 36(8): 2677–2689.
- [60] Choi S W, Kim H W, Nam J W. The small peptide world in long noncoding RNAs[J]. *Brief Bioinform*, 2019, 20(5): 1853–1864.
- [61] Peeters M K R, Menschaert G. The hunt for sORFs: a multidisciplinary strategy[J]. *Exp Cell Res*, 2020, 391(1): 111923. DOI:10.1016/j.yexcr.2020.111923.
- [62] Hanada K, Akiyama K, Sakurai T, *et al.* sORF finder: a program package to identify small open reading frames with high coding potential[J]. *Bioinformatics*, 2010, 26(3): 399–400.
- [63] Min X J, Butler G, Storms R, *et al.* OrfPredictor: predicting protein-coding regions in EST-derived sequences[J]. *Nucleic Acids Res*, 2005, 33(2): W677–W680.
- [64] Lin M F, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions[J]. *Bioinformatics*, 2011, 27(13): i275–i282.
- [65] Guo J C, Fang S S, Wu Y, *et al.* CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition[J]. *Nucleic Acids Res*, 2019, 47(W1): W516–W522.
- [66] Pueyo J I, Magny G, Couso J P. New peptides under the s(ORF)ace of the genome[J]. *Trends Biochem Sci*, 2016, 41(8): 665–678.
- [67] Li F, Zhong X, Zhao Z, *et al.* Comprehensive analysis of ncRNA involvement in brain microglia immunology[J]. *Clin Immunol*, 2022, 241: 109075. DOI:10.1016/j.clim.2022.109075.
- [68] Slavoff S A, Mitchell A J, Schwaid A G, *et al.* Peptidomic discovery of short open reading frame-encoded peptides in human cells[J]. *Nat Chem Biol*, 2013, 9(1): 59–64.
- [69] Orr M W, Mao Y, Storz G, *et al.* Alternative ORFs and small ORFs: shedding light on the dark proteome[J]. *Nucleic Acids Res*, 2020, 48(3): 1029–1042.
- [70] Yeasmin F, Yada T, Akimitsu N. Micropeptides encoded in transcripts previously identified as long noncoding RNAs: a new chapter in transcriptomics and proteomics[J]. *Front Genet*, 2018, 9: 144. DOI:10.3389/fgene.2018.00144.
- [71] Brunet M A, Leblanc S, Roucou X. Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs[J]. *Exp Cell Res*, 2020, 393(1): 112057. DOI:10.1016/j.yexcr.2020.112057.
- [72] Deng J, Xu W, Jie Y, *et al.* Subcellular localization and relevant mechanisms of human cancer-related micropeptides[J]. *FASEB J*, 2023, 37(12): e23270. DOI:10.1096/fj.202301019RR.
- [73] 王心怡, 张祯, 石成渝, 等. 长链非编码 RNA 编码小肽的生物医学意义及研究策略[J]. *浙江大学学报(医学版)*, 2023, 52(4): 397–405.
- [74] Ye M, Zhang J, Wei M, *et al.* Emerging role of long noncoding RNA-encoded micropeptides in cancer[J]. *Cancer Cell Int*, 2020, 20: 506. DOI:10.1186/s12935-020-01589-x.



【专家介绍】王颖：中国药科大学副研究员，硕士生导师，博士。长期致力于探索非编码 RNA 及其编码的新型微肽在疾病发生发展中的作用机制，并利用高通量微肽筛选与验证技术平台发现疾病相关的新型微肽分子，开发原创多肽类新药，为疾病的诊疗提供新的策略，曾参与完成 2 个多肽新药的临床前开发并获新药临床批件 2 件。以第一作者在 *Signal Transduct Target Ther* (IF: 38.104)、*J Am Chem Soc* (IF: 16.383)、*Acta Pharm Sin B* (IF: 14.903)、*Mol Ther Nucleic Acids* (IF: 10.183)、*Cell Death Dis* (IF: 9.685) 和 *Oncogene* (IF: 8.756) 等国际一流学术期刊发表多篇论文；主持国家自然科学基金、博士后面上基金；申请中国发明专利 2 项，授权 1 项；获中国产学研合作创新成果奖二等奖、第六届江苏医药科技进步奖二等奖。