

肿瘤与非经典蛋白：从“暗蛋白质组”到诊疗新策略

范子鸿¹, 王苏杰^{1,2}, 何庆瑜¹, 张静³

(1. 暨南大学生命科学技术学院生命健康工程研究院, 教育部肿瘤分子生物学重点实验室, 广东 广州 510632;

2. 中山大学肿瘤防治中心分子诊断科, 华南恶性肿瘤防治全国重点实验室, 广东 广州 510060;

3. 暨南大学附属第一医院, 广东 广州 510632)

摘要: 非经典蛋白是指源于传统上被归类为“非编码区”或位于“暗基因组区域”的翻译产物, 其正日益成为肿瘤生物学及精准医学领域研究的新焦点。随着核糖体图谱、翻译组全长测序、蛋白质从头测序技术的快速发展, 大量先前被忽视的非编码区来源的暗蛋白和在经典编码区内通过非典型机制产生的蛋白变体, 已被证实广泛参与肿瘤的代谢重编程、信号传导、免疫逃逸机制及临床耐药性等关键环节。本文系统综述了非经典蛋白的发现历程、在肿瘤发生发展中的分子作用机制、作为肿瘤特异性抗原的潜在价值, 以及当前临床应用面临的挑战与未来发展方向。

关键词: 非经典蛋白; 暗蛋白质组; 肿瘤标志物; 肿瘤新抗原

中图分类号: O657.63

文献标志码: A

文章编号: 1004-2997(2026)03-0380-18

DOI: 10.7538/zpxb.2025.0139

CSTR: 32365.14.zpxb.2025.0139

Tumors and Non-Canonical Proteins: From the “Dark Proteome” to Novel Diagnostic and Therapeutic Strategies

FAN Zi-hong¹, WANG Su-jie^{1,2}, HE Qing-yu¹, ZHANG Jing³

(1. MOE Key Laboratory of Tumor Molecular Biology, Institute of Life and Health Engineering, College of Life

Science and Technology, Jinan University, Guangzhou 510632, China; 2. State Key Laboratory of Oncology in South

China, Department of Molecular Diagnostics, Sun Yat-sen University Cancer Center, Guangzhou 510060, China;

3. The First Affiliated Hospital, Jinan University, Guangzhou 510632, China)

Abstract: Non-canonical proteins are translation products that arise from genomic regions historically classified as “non-coding” or situated within the so-called “dark genome”. They also encompass proteins or variants produced through atypical translational mechanisms within classical protein-coding genes, such as alternative reading frames (aORFs), upstream open reading frames (uORFs), non-AUG initiation, and ribosomal frameshifting. For a long time, these translational products were largely overlooked because they tend to be short, expressed at low abundance, and lack recognizable conserved domains, thereby evading detection by standard gene-prediction algorithms. Over the past decade, methodological advances have dramatically altered this view. High-resolution ribosome profiling, long-read and full-length translome sequencing, single-cell transcriptomic and translomic approaches, improvements in high-sensitivity mass spectrometry and de novo peptide sequencing, and integrated proteogenomic pipelines have collectively enabled robust identification

and confident peptide assignment for numerous dark-proteome members and non-canonical isoforms. Experimental validation, ranging from targeted proteomics and immunopeptidomics to functional perturbation assays, now supports the biological existence and functional relevance of numerous such peptides. A growing body of evidence indicates that non-canonical proteins are not mere translational noise but can exert substantial biological effects: they rewire cellular metabolism, modulate canonical signalling cascades, influence proteostasis and stress responses, remodel the tumour immune microenvironment, and contribute to resistance against chemotherapy, radiotherapy and targeted agents. Importantly, many non-canonical products exhibit tumour-restricted or tumour-enriched expression patterns and include sequence elements absent from normal tissues. These properties make them compelling candidates for tumour-specific biomarkers, neoantigens for personalized immunotherapy, and novel molecular targets for precision interventions. Despite this promise, major challenges remain for their routine clinical translation. These include heterogeneous and incomplete annotation criteria, variable detection sensitivity and reproducibility across platforms, high false-positive rates in peptide identification, a lack of scalable functional assays to assign mechanistic roles, and unresolved questions around immunogenicity and safety of targeting these molecules. To move the field forward, coordinated efforts are advocated to establish community standards for annotation and reporting, to integrate complementary multi-omics datasets, to develop high-throughput functional screening pipelines, and to design early-phase clinical studies prioritizing safety and translational feasibility. This review synthesized the technological milestones and current mechanistic insights into non-canonical proteins in cancer, evaluated their translational potential as biomarkers and therapeutic targets, and outlined strategic priorities to accelerate their responsible integration into precision oncology practice.

Key words: noncanonical proteins; dark proteome; tumor markers; tumor neoantigens

随着人类基因组计划(human genome project)的推进与完成,生命科学研究正式迈入后基因组时代,蛋白质组学探索不断向纵深发展^[1]。随着蛋白质组学技术的持续发展与进步,蛋白质组学能够更直接地揭示肿瘤发生发展的分子机制^[2]。在过去20年间,蛋白质组学研究经历了从蛋白质鉴定、定量分析到翻译后修饰解析的跨越式进步。然而,在经典蛋白之外,生物界还存在大量被称为“暗蛋白质(dark proteins)”的神秘组分^[3]。早期的“暗蛋白质组(dark proteome)”主要指“缺失蛋白”,即基因组预测但未被质谱或抗体证实的蛋白^[4];而如今的“暗蛋白质组”已拓展为更宏大的框架,还包含由非经典开放阅读框(ncORFs)编码的功能微肽、难以检测的蛋白变体以及结构和功能尚不明确的蛋白种类,其检测难度往往源于这些分子的理化特性与低丰度等因素^[5]。这一概念的演变,体现了学界对人类基因组编码能力复杂性和蛋白质组多样性的认知不断深化^[3-4]。

暗蛋白构成了基因组注释缺口的重要部分,越来越多的研究^[3,6]表明,这些暗蛋白深度参与肿瘤的发生发展、转移乃至耐药的过程,并在神经退行性疾病等其他重大病理过程中扮演着重要角色。

本文将以肿瘤为核心场景,系统梳理暗蛋白发现技术的革新、功能机制的解析及临床转化的路径;回顾质谱、生物信息学及转录组学如何协同突破传统检测盲区;深入探讨暗蛋白在肿瘤信号网络、代谢重编程及免疫微环境中的多方位作用;并重点评估其作为下一代治疗靶点和生物标志物的临床潜力;最后,讨论暗蛋白研究面临的挑战与未来方向。

1 非经典蛋白概述

1.1 非经典蛋白来源与分类

在肿瘤生物学领域中,蛋白质研究长期集中于由经典编码基因经转录、翻译生成的产物。人类基因组中大约只有2%的基因是编码基因,绝大多数转录本缺乏编码开放阅读框,被视为非

编码 RNA(ncRNA),包括微小 RNA(microRNA)、长链非编码 RNA(lncRNA)和环状 RNA(circRNA)^[7-10]。然而,近年来随着基因组学和蛋白质组学技术的飞速发展,暗蛋白受到研究者广泛关注。这些蛋白质并非来源于传统的蛋白质编码基因,而是来源于基因组的“非编码”区域或异常翻译事件,它们在肿瘤的发生、发展及免疫应答中扮演着意想不到的关键角色,为癌症诊断和治疗提供了全新的思路和靶点。根据生物合成起源,肿瘤中的非经典蛋白主要可分为非编

码区来源的微蛋白和经典编码区内通过非典型机制产生的蛋白变体两大类,具体信息列于表 1。非编码区来源的微蛋白主要包括 5'UTR/3'UTR/CDS 内重叠 ORF 产生的暗蛋白、lncRNA/lincRNA/伪基因转录本中短开放读码框(sORF)编码的暗蛋白、circRNA 驱动的 ORF 编码蛋白等 3 类;而经典编码区内通过非典型机制产生的蛋白变体则主要指经典 CDS 基因通过非典型翻译机制(如非-AUG 起始、N 端延伸)产生的变体蛋白。

表 1 非经典蛋白的分类

Table 1 Classification of non-classical proteins

类别 Type	来源 Origin	特点 Feature	代表性蛋白 Representative protein	参考文献 Reference
非编码区来源的微蛋白	5'UTR/3'UTR/CDS内重叠ORF产生的暗蛋白	翻译起始及框架非典型(可能位于5'UTR、3'UTR或CDS内部重叠阅读框),长度较短	如MPEP(MYC-uORF peptide)、ASNSD1-uORF microprotein	[11-12]
	lncRNA/lincRNA/伪基因转录本中sORF编码的暗蛋白	原本标注为非编码RNA(如lncRNA、伪基因),多数暗蛋白产物长度<100 aa,部分研究 ^[14] 也将100~150 aa以内的翻译产物纳入讨论范畴	如NBASP、HOXB-AS3	[13-14]
	circRNA驱动的ORF编码蛋白	翻译机制非常非典型(无5'cap、环状结构、IRES驱动)	如FNDC3B、circPIAS1-108aa	[15-18]
经典编码区内通过非典型机制产生的蛋白变体	经典CDS基因通过非典型翻译机制(如非-AUG起始、N端延伸)产生的变体蛋白	来自传统蛋白编码基因,但翻译起始位置、阅读框、起始密码子(如非-AUG)等发生改变,从而产生新的蛋白变体	如PTEN N-terminal extension proteoform	[19-21]

1.1.1 CDS/5'UTR/3'UTR 内重叠 ORF 产生的暗蛋白 这一类暗蛋白起源于传统 mRNA 注释以外的区域,它们的翻译起始方式与经典蛋白不同,如 5'非翻译区(5'UTR)上的上游开放读码框(uORF)、3'非翻译区(3'UTR)中的下游 ORF(dORF)或在 CDS 内部但处于不同阅读框的重叠 ORF(altORF/intORF)^[22-23]。近年来,通过核糖体测序分析、质谱和 CRISPR 筛选等技术发现,这些 ORF 虽短,通常只有几十至数百个氨基酸,但却能编码具有生物活性的“微蛋白”^[24-26]。有研究发现,在儿童高危型髓母细胞瘤中,位于 ASNSD1 基因 5'UTR 的 uORF 所翻译出的微蛋白“ASNSD1-uORF microprotein”是 MYC 驱动型肿瘤细胞生存的依赖因子^[12]。此外,还有研究发现,一种由 MYC mRNA 5'-UTR 内的 ORF 编码的分泌性微蛋白 MPEP 可作为 TRKB 受体酪氨酸激酶的激动配体发挥作用,促进胶质母细胞瘤干细胞生长,独立于 MYC 蛋白功能^[11]。此类产物常常逃避免疫监视,可能作为新的肿瘤抗原,同时其翻

译起始位置多样且与主 CDS 蛋白的功能不同或相辅相成。近年来,uORF、dORF 及 CDS 内重叠 ORF 在癌症中的作用受到学界重视,并为肿瘤诊断、治疗和新抗原发现提供了新视角。

1.1.2 lncRNA/lincRNA/伪基因转录本中 sORF 编码的暗蛋白 过去被标注为“非编码”的转录本,如 lncRNA、长间隔非编码 RNA(lincRNA)、伪基因转录本,近年来被发现含有能够翻译的 sORF,这些 sORF 可编码长度通常为低于 101~150 个氨基酸的微蛋白^[27-30]。这些转录本最初被注释为非编码 RNA 而被认为不具有蛋白质编码能力,但随着翻译组学和相关多组学技术的快速发展,越来越多的研究揭示它们具有实际的翻译活性,并且其编码的暗蛋白在各种生物过程中发挥功能性作用,特别是在癌症发生和进展中表现出重要的生物学功能。此类微蛋白可能参与 RNA 剪接、翻译调控、代谢通路、信号复合体的组装等生物学过程。例如,在神经母细胞瘤中,由 lncRNA FAM201A 编码的小蛋白 NBASP 被证明

通过与 FABP5 结合并抑制 MAPK 通路, 从而抑制肿瘤进展^[13]。在结直肠癌中, 由 lncRNA HOXB-AS3 编码的 HOXB-AS3 peptide 通过与异质核糖核蛋白 A1(hnRNP A1) 直接结合来干扰 hnRNP A1 调控的剪接机制, 抑制肿瘤细胞增殖、迁移与肿瘤形成^[14]。类似地, 其他由 lncRNA 编码的微蛋白在结直肠癌、肾细胞癌中也被报道具有促癌或抑癌作用。这类发现提示着“非编码”转录本的编码潜力和生物学功能正被学界重新评估。

1.1.3 circRNA 驱动的 ORF 编码蛋白 circRNA 是通过反向剪接(back-splicing)形成的闭合环状结构 RNA, 不具有典型的 5'cap 和 3'poly(A) 尾巴^[22]。在传统认知中, circRNA 多被视为 miRNA 海绵或蛋白结合分子, 而近年来发现, 其中一些 circRNA 含有可被翻译的 ORF, 并能通过 IRES (内部核糖体进入位点)、m6A 修饰驱动或滚环翻译(rolling-circle)机制产生微蛋白^[22,31]。这些由 circRNA 翻译出的蛋白在肿瘤中逐渐显示出功能潜力, 可能介入信号传导、代谢调控、免疫逃逸或新抗原产生^[31]。有研究指出, 由 circPIAS1 编码的 108 aa 蛋白可通过影响 STAT1 的 SUMO 化促进免疫诱导铁死亡逃逸^[18]。还有研究指出, 在肿瘤细胞中, FNDC3B 的表达升高与细胞增殖、侵袭能力增强及患者不良预后相关^[15-17]。该类暗蛋白的出现强调了翻译机制的高度非典型性, 也扩展了蛋白质组学边界, 且 circRNA 编码的蛋白可能成为肿瘤生物学和免疫治疗中的新靶点。

1.1.4 经典 CDS 基因通过非典型翻译机制产生的变体蛋白 即便是传统注释的蛋白编码基因(CDS)也并非总是只产生一种标准蛋白形式。越来越多的证据表明, 这些基因能够开启非典型翻译机制: 例如, 采用非-AUG 起始密码子(如 CUG、GUG)形成 N 端延伸或截短的异构蛋白形式(proteoforms); 这些变体蛋白在定位、稳定性、互作网络或功能上与经典形式截然不同^[21,32-34]。在癌症背景下, 这类变体蛋白被发现具有重要的生物学意义, 比如, 由 PTEN 基因非-AUG 起始编码生成的 N 端延伸型蛋白, 其定位和功能可能与经典 PTEN 蛋白不同, 从而为 PI3K/AKT 通路带来新的调控维度^[19-21]。尽管这类机制尚未得到系统研究, 但它们拓展了蛋白质组的复杂性,

并提示我们应更全面地审视“经典”基因的翻译产物多样性。

2 非经典蛋白检测技术革新: 照亮“暗蛋白质组”

2.1 传统方法的局限与暗蛋白质组发现困境

近年来, 许多研究报道了通过核糖体测序分析等技术发现大量的非经典开放读码框在转录本中被翻译^[35], 暗示“暗蛋白质组”远比传统注释的蛋白编码基因所反映的要更丰富。然而, 当研究者试图用传统鸟枪法质谱(shotgun MS)法去验证这些翻译产物时, 发现可被检测和鉴定出的微蛋白比例非常低。有研究指出, 尽管 Ribo-seq 显示大量翻译事件发生, 但在酵母细胞中通过 MS 仅鉴定到少量非经典蛋白^[36]。

导致这一困境的原因是多方面的: 其一, 现有蛋白质组学方法的设计主要面向典型蛋白而非“微蛋白”这一类别; 另一方面, 这些新蛋白所具有的特殊生物学特征使其难以在传统蛋白质组学流程中被捕捉。这种“方法学-数据库-生物特性”三重制约, 使得暗蛋白的研究陷入瓶颈。

2.1.1 传统蛋白质组学的局限性 传统蛋白质组学流程中, MS/MS 获得的肽段数据通常通过与传统的“经典蛋白”数据库匹配进行鉴定, 这类数据库中只包含由已有注释基因衍生的蛋白序列。然而, 传统蛋白质组学技术与数据库框架并未针对这些暗蛋白量身定制, 许多暗蛋白在传统数据库中未被注释, 因此这些信号极可能被视为噪音或伪检出予以过滤。尽管这些数据库在持续更新, 但其收录的蛋白质序列仍仅限于经过实验验证或通过同源性预测并被注释为编码基因的区域。那些位于非编码区域的潜在 ORFs, 或因 RNA 剪接、翻译重编程而产生的新型蛋白质变体, 几乎被完全排除在数据库之外。质谱分析软件(如 MaxQuant、Spectronaut)通常会过滤掉与任何数据库序列不匹配的谱图, 从而导致大量来源于非经典蛋白的质谱信号被当作噪音丢弃, 最终造成系统性的检测偏差^[36]。此外, 在非经典蛋白鉴定中, 错误发现率(false discovery rate, FDR)控制也因数据库构建、伪库设计(decoy)偏差等问题而变得更加复杂。有研究指出, 当用非经典蛋白与典型蛋白一并控制整体 FDR 时, 非经典蛋白子集的实际 FDR 往往被低估^[36]。因

此,若数据库内不纳入 ncORF-产物或微蛋白条目,在质谱数据处理中,这些候选项极可能在“搜库”阶段就被排除,从而造成目标蛋白发现率低下。

2.1.2 非经典蛋白质性质的特殊性 除数据库问题外,暗蛋白独特的理化性质和丰度特征也为其被传统质谱方法识别构成了重大障碍。首先,由于这些蛋白通常长度极短(几十至百余氨基酸),对应可生成的胰蛋白酶酶切肽段数量少、肽段覆盖度低,从而降低了 MS 识别的概率;传统鸟枪法质谱若无专门的富集和优化策略,往往很难对其实现有效覆盖^[36]。Papadopoulos 等^[37]在酵母相关研究中指出,相比经典蛋白,非经典蛋白更难产生可观的肽段,从而导致检测率显著偏低。这些蛋白通常表达量极低,部分还可能被迅速降解,导致细胞内稳态水平较低,即使通过开放搜索策略推测出某些非经典肽段的存在,但对其进行后续的体外验证也较为困难。而且,在液相色谱-质谱分离或离子化过程中,其可能因基质抑制、前体离子信号弱、噪声干扰相对显著的问题进一步影响检测限。因此,从理化层面而言,暗蛋白属于“质谱弱信号”群体,传统流程缺乏专门适配方案,导致相关质谱信号丢失严重。

长期以来,蛋白质研究主要集中于经典编码基因,而对非编码区域的翻译潜力认知不足,许多非经典 ORFs 被默认为“转录噪音”或“翻译噪音”,其生物学意义长期被低估^[38]。数据库注释缺失、搜库流程过滤、理化特征不利于检测、蛋白丰度极低、降解速率快、检测信号弱等因素,共同造成了“我们知道有翻译事件发生,却无法广泛检测到产物”的悖论。只有通过优化数据库、富集短肽、优化酶切条件和提高检测灵敏度等方法改进质谱策略,并结合多组学(如 Ribo-seq+MS+免疫肽组)联用方法,才能突破这一发现困境。

2.2 多组学技术革新突破检测瓶颈

依赖标准数据库分析“暗蛋白质组”的策略,极大地限制了对非经典蛋白质的发现与鉴定。近年来,多组学技术的交叉融合与创新,为系统性揭示这些“隐藏的蛋白质宇宙”提供了强大的工具,正在逐步突破长期存在的检测瓶颈。

2.2.1 核糖体图谱 核糖体图谱(Ribo-seq)技术通过高通量测序捕获被核糖体保护的 mRNA 片

段,从而精确绘制全基因组范围内的翻译活性位点^[39]。Ribo-Seq 的序列片段被称为核糖体足迹(RFs),长度通常为 30 nt,这些片段受到核糖体的保护,免受核酸酶的消化^[40]。Ribo-seq 数据可用于研究翻译控制的各个方面,如核糖体暂停、密码子使用、选择性剪接和无意义介导的衰变^[41-43]。其优点在于不仅可以直接监测翻译的过程,无需预先知晓蛋白质编码潜力,即可直接鉴定正在被翻译的 ORFs,包括位于注释基因之外的非经典 ORFs;还可以高分辨率定位起始密码子,能够准确识别非 AUG 起始等非经典翻译事件,进而揭示新的翻译起始位点。此外,Ribo-seq 技术还能提供翻译动态信息,同时评估翻译效率,为区分功能性翻译与随机噪音提供可靠依据,已成为发现新型微肽和非经典蛋白的“雷达”,是构建全新注释数据库的关键数据源头^[44]。Zheng 等^[26]在结直肠癌中结合 Ribo-seq 与 CRISPR 的多组学策略,鉴定到多个“隐匿 ORF”(cryptic ORFs)在癌细胞中可被翻译,且其中 1 个暗蛋白 SMIMP 可通过与 SMC1A 相互作用增强对细胞周期抑制基因的表观遗传抑制,从而促进结直肠癌细胞的增殖与肿瘤生长,提示这些隐匿 ORF 编码产物可能在肿瘤发生发展中具有致癌性功能。同样,Camarena 等^[45]在肝细胞癌中利用 Ribo-seq 和免疫肽组学(immunopeptidomics)技术鉴定到许多源自 ncORFs 的肿瘤特异抗原。

2.2.2 翻译组全长测序 翻译组全长测序(ribosome-nascent chain complex, RNC-seq)是一种与 Ribo-seq 互补的技术,它通过蔗糖梯度离心富集与核糖体结合的新生肽链复合物,并对其对应的 mRNA 进行全长测序^[46]。该方法可以富集低丰度的转录本,进而有效富集传统转录组测序中容易忽略的、正在翻译的低丰度 mRNA,弥补传统转录组测序对低丰度转录本的检出盲区。同时,因其全长测序的特性,该技术有助于识别由可变剪接、基因融合等产生的新型转录本及其翻译产物。RNC-seq 可对 Ribo-seq 所鉴定的翻译事件提供更直接的支持与验证,从而提高非经典翻译事件判断的可靠性^[47-48]。

2.2.3 蛋白质基因组学 蛋白质基因组学(proteogenomics)通常被定义为将样本的基因组和转录组信息与质谱驱动蛋白质组学数据进行整合的一类方法学^[49-50],其核心思想是利用样

本特异性的核酸序列构建定制化的蛋白质序列数据库,从而在质谱数据搜索中识别变异肽段、拼接异构体、起始位点变异等先前未注释的开放阅读框,以此补充基因或转录层面的信息盲点^[50-51]。与传统“仅依赖公共蛋白数据库”的蛋白质组学相比,蛋白质基因组学在研究非经典蛋白方面具有独特且关键的优势^[52-53]。蛋白质基因组学通过将样本的 RNA-seq、长读长转录组和 Ribo-seq 转译为候选肽序列并加入搜索数据库,使研究者能够在质谱搜索中直接检索源自 lncRNA、5'/3'UTR、内含子等在传统数据库中未被注释的肽段,从而大幅提高对低丰度或未注释翻译事件的检测灵敏度和特异性^[25]。此外,蛋白质基因组学流程可以把多组学证据联合起来,为候选非经典肽提供“多证据”支持,显著降低假阳性率并增强对真实翻译事件的置信度^[54]。近年来,随着质谱技术的不断发展和成熟,蛋白质基因组学在疾病机制研究中的作用日益凸显,特别是在肿瘤发生发展、肿瘤抗原研发、肿瘤分型分子标志物筛选及精准医疗领域^[55],通过将免疫肽组学与肿瘤特异性蛋白质基因组数据库联合,其已成为发现肿瘤特异性新抗原(neoantigens)与优选免疫治疗靶点的有力工具^[56]。此外,蛋白质基因组学还在传染性疾病、心血管疾病、药效评价、炎症、新药开发等多个领域得到应用^[57]。例如,在约5%的非小细胞肺癌(NSCLC)病例中,ALK基因重排会驱动肿瘤的形成^[58]。这一发现促成了ALK抑制剂的开发,包括克唑替尼和塞瑞替尼,并获得了美国食品药品监督管理局(FDA)批准,可用于ALK基因重排阳性的肺癌患者,这不仅改变了众多患者的生存质量,还降低了近年来全球非小细胞肺癌的死亡率^[59]。

2.2.4 从头测序技术 在蛋白质组学研究中,液相色谱-串联质谱(LC-MS/MS)是常用的蛋白质鉴定与定量技术,但该技术往往依赖已有的蛋白质数据库^[60-61],如何在复杂的样品中准确检测和定量蛋白质类肿瘤标志物是其面临的挑战^[62]。从头测序(de novo)技术可直接从串联质谱(MS/MS)谱图推断肽段的氨基酸序列,在不依赖已知数据库的情况下,直接从质谱数据中解码出肽段的氨基酸序列,从而发现与已知序列均不匹配的、完全新颖的蛋白质序列,进而提高未知肽段的识别率,这种方法尤其适用于研究由遗传变

异、RNA编辑等产生的独特蛋白变体。De novo技术为研究人员发现肿瘤新抗原提供了“数据库以外”的路径,突破了已知蛋白质组注释的局限,拓展了肿瘤特异抗原的发现范围。Luo等^[63]采用de novo方法对肝癌样本进行分析,发现与数据库搜索相比,该方法识别了更多突变基因对应的肽段,显示了其在新抗原发现中的数据库“补缺”潜力,有望加速临床癌症疫苗的开发。然而,由于在质谱检测过程中,酶切、噪声处理可能会出现离子丢失等问题,导致质谱质量的波动,这给de novo带来了极大挑战。此外,de novo技术对质谱图质量要求极高,且推导过程计算难度大、复杂度高、结果容易出错^[64]。但随着算法的不断优化(如基于深度学习模型),其准确性和可靠性正在显著提升,并成为蛋白质组学研究中不可或缺的补充手段。随着蛋白质组学领域的不断发展,de novo技术的进步对于揭示蛋白质的结构和功能变得愈发关键^[65]。

综上,核糖体图谱和翻译组测序为发现翻译事件提供了“线索”,蛋白质基因组学利用这些线索构建“新地图”并指导质谱仪在这张新地图上进行“精准导航”,而从头测序则作为最终的“探险家”去确认那些地图上未曾标记的“新大陆”。这些多组学技术的协同应用,正在推动蛋白质组学进入一个全新的发现时代。

2.3 生物信息学工具升级

非经典蛋白的系统性发现高度依赖于生物信息学工具的重大升级,主要体现在数据库资源的丰富与预测算法的精准化两方面。传统的、静态的参考数据库正被动态的、可定制的数据库生成流程所补充和替代。例如,sORFs.org、OpenProt等综合性数据库系统地收录了来自多种预测工具和Ribo-seq研究的ncORFs数据,为研究者提供了宝贵的基础资源^[66-67]。更重要的是,诸如ProteoGenerator、CustomProDBJ等工具允许用户直接整合样本特异的基因组和转录组(如RNA-seq)数据,自动构建包含所有可能的经典和非经典ORF的定制化蛋白质序列数据库,为后续的蛋白质基因组学分析奠定了坚实基础^[68]。此外,基于深度学习的从头肽测序模型DeepNovo能够不依赖数据库,仅根据质谱数据推断出肽段的氨基酸序列,从而识别出未被数据库记录的肽段,为暗蛋白研究带来希望^[64]。

在预测算法层面,早期的计算工具主要依赖进化保守性序列和固有 ORF 特征进行预测。如今,基于机器学习和深度学习的新一代算法(如 deepribo、ORFfinder 及其进阶应用)能够更智能地整合多组学特征(特别是 Ribo-seq 的读段分布、周期性和翻译效率),显著提高了对具有翻译潜力的非经典 ORF,尤其是短 ORF 的预测精度,并有效降低了假阳性率^[69]。PyPropel 是旨在简化蛋白质序列数据处理的 Python 软件包,能够高效处理、注释及整合多来源的蛋白质(包括未充分注释的)序列与特征,特别适用于机器学习研究;它能提供广泛的功能,以促进蛋白质序列数据的预处理、结构和功能描绘以及后处理,并且支持整合来自多个来源的蛋白质特征^[70]。CPC2 是基于序列特征、进化保守性等指标,评估转录本(尤其是非编码 RNA)的编码潜力,可快速、准确地评估 RNA 转录本的编码能力;CPC2

模型具有物种中立性,使其有利于研究不断增长的非模式生物转录组,甚至是注释不佳或缺乏基因组组装的生物体转录组^[71]。PhyloCSF 是一种利用进化保守方法检测蛋白质编码区域的成熟方法,其输入多序列比对(MSA)和相关物种的系统发育树,计算基于密码子替换率和密码子频率的评分,以此反映该区域编码或非编码的可能性^[72]。IRESpy/IRESfinder 等工具可以预测内部核糖体进入位点(IRES),这是许多 circRNA 等非经典 mRNA 翻译起始的关键元件^[73]。OpenProt 收录多物种的非经典或替代 ORF 及其预测的蛋白序列、结构和功能注释,支持定制化数据库构建^[74]。这些算法与定制数据库的紧密结合,共同构成了从海量数据中筛选和验证非经典蛋白的强大计算引擎,极大地加速了这一前沿领域的探索进程。部分生物信息处理工具的具体信息列于表 2。

表 2 生物信息分析工具描述

Table 2 Description of bioinformatics analysis tools

名称 Name	网址 Website	参考文献 Reference	功能描述 Function description
DeepNovo	https://github.com/nh2tran/DeepNovo	[64]	直接从串联质谱(MS/MS)数据中进行从头测序和数据库搜索,以鉴定多肽、重建蛋白质序列
sORFs.org	http://www.sorfs.org	[66]	鉴定、查询和评估sORF的编码潜力
DeepRibo	https://github.com/Biobix/DeepRibo	[69]	从Ribo-seq数据中精确预测正在被核糖体活跃翻译的ORF
ORFfinder	https://www.ncbi.nlm.nih.gov/orffinder/	[69]	从任意核酸序列中查找所有可能的ORF
PyPropel	https://github.com/2003100127/pypropel	[70]	在原核生物基因组中高精度预测蛋白质编码基因
CPC2	http://cpc2.cbi.pku.edu.cn	[71]	快速、准确地从转录本序列中区分蛋白质编码RNA与非编码RNA
PhyloCSF/PhyloCSF++	https://github.com/cpockrandt/PhyloCSFpp	[72]	评估转录本是否具有蛋白质编码潜能
IRESfinder	https://github.com/xiaofengsong/IRESfinder	[73]	识别mRNA序列中内部核糖体进入位点(IRES)
OpenProt	https://www.openprot.org/	[74]	系统收录并整合了来自多物种的非经典ORF及其翻译产物的实验与预测证据

3 非经典蛋白在肿瘤中的多维调控机制

随着转录组学、核糖体足迹测序(Ribo-seq)和高灵敏度质谱等技术的发展,越来越多的证据表明,非经典蛋白并非单纯的“翻译噪音”,而是以多层次、多维度方式参与调控肿瘤的发生与演进,进而影响肿瘤的侵袭性、转移能力及耐药性等关键表型。目前已有研究报道的非经典蛋白在癌症中主要参与的过程示于图 1。

3.1 驱动肿瘤恶性表型

非经典蛋白在肿瘤中兼具抑癌和促癌的功效,

lncRNAs 可通过多种机制调控基因表达^[75]。一方面,诸多由 lncRNA 或 circRNA 编码的肽被证实为新型致癌驱动因子,可直接增强肿瘤细胞的增殖和存活能力。例如,由 lncRNA 翻译产生的 UBAP1-AST6 蛋白主要定位于细胞核,在 A549 肺癌细胞中高表达,其过表达可显著促进细胞增殖和集落形成,从而推动肿瘤的发生发展^[55]。在中枢神经系统肿瘤中,非经典蛋白同样表现出显著的致癌驱动作用。研究表明,来源于 circRNA 的翻译产物 C-E-Cad 在胶质母细胞

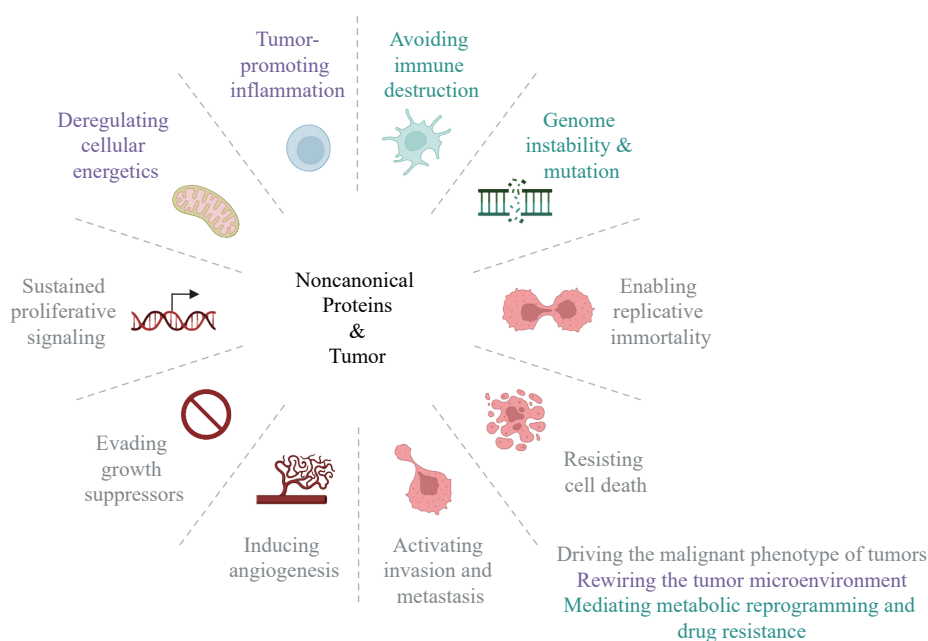


图1 非经典蛋白在癌症中参与的过程

Fig. 1 Processes involving non-canonical proteins in cancer

瘤中高表达,可通过激活 EGFR-STAT3 信号通路维持胶质瘤干细胞的肿瘤发生能力^[76];此外,由 circ-SMO 翻译产生的 SMO-193aa 蛋白也可通过增强 Hedgehog 信号通路活性,促进胶质母细胞瘤的生长和侵袭^[77]。另有研究表明,部分非经典蛋白可发挥抑癌作用,典型代表为 HOXB-AS3 编码的微肽,其在结直肠癌组织中显著下调,能够抑制肿瘤细胞的增殖、侵袭与转移,并通过调控代谢通路抑制肿瘤生长,其低表达通常与不良预后密切相关^[14]。此外, LINC00961 编码的微肽 SPAR 可通过抑制 mTORC1 信号通路限制肿瘤细胞生长^[78]; circLINC-PINT 翻译产生的 PINT87aa 可结合 PAF1 复合体,抑制多种致癌基因的转录延伸,从而抑制胶质母细胞瘤细胞增殖与肿瘤形成^[79]。这些研究共同表明,非经典蛋白可作为肿瘤恶性表型的重要调控因子。

3.2 介导代谢重编程与耐药

代谢重编程是指细胞通过调整代谢途径以满足能量与生物合成的需求,从而促进增殖并提高对外界胁迫的耐受性^[80]。肿瘤代谢重编程是肿瘤细胞适应快速增殖和复杂微环境的重要机制,非经典蛋白在这一过程中发挥着不可忽视的调控作用。肿瘤细胞往往通过增强糖酵解、重塑线粒体功能或改变脂质代谢途径,以满足能量和生物合成的需求^[81],而部分来源于非经典

ORFs 的微肽正是这些代谢通路的关键调节因子。例如, HOXB-AS3 编码的 53-aa 微肽可通过干预 PKM 的可变剪接,抑制糖酵解相关代谢重编程^[14]。此外,部分由环状 RNA 翻译产生的非经典蛋白也被证实参与肿瘤代谢相关信号通路并影响治疗耐受。例如, circAKT3 编码的 AKT3-174aa 微肽可通过竞争性抑制 PDK1-介导的 AKT Thr308 磷酸化,显著降低 AKT 信号通路活性,从而抑制胶质母细胞瘤细胞的增殖并降低其放射耐受性^[82]。类似地, circFBXW7 编码的 FBXW7-185aa 可促进 c-Myc 蛋白降解,而 c-Myc 是驱动肿瘤代谢重编程和治疗耐受的重要转录因子,该微肽的缺失与胶质瘤恶性的进展密切相关^[83]。值得注意的是,非经典蛋白不仅在肿瘤细胞内部发挥作用,还可通过外泌体介导的细胞间通讯影响代谢状态。例如,肿瘤相关巨噬细胞 (tumor-associated macrophages, TAMs) 可通过外泌体向乳腺癌细胞传递 lncRNA-HISLA,从而提高肿瘤细胞的需氧糖酵解水平,并赋予其对化疗诱导凋亡的显著耐受性;阻断该外泌体-HISLA 通路可明显削弱肿瘤细胞的代谢优势并恢复其对化疗药物的敏感性^[84]。

综上,非经典蛋白通过直接调控肿瘤细胞的代谢通路,或通过肿瘤微环境中的细胞间通讯网络,参与肿瘤代谢重编程与治疗耐受的形成。这类

研究提示,非经典蛋白及其相关翻译产物在肿瘤代谢调控和耐药形成中具有重要地位,可能成为克服肿瘤耐药和开发新型治疗策略的重要切入点。

3.3 重塑肿瘤微环境

除直接作用于肿瘤细胞外,非经典蛋白及其衍生微肽能够通过分泌或外泌体等形式释放至胞外,并参与肿瘤微环境(tumor microenvironment, TME)的动态重塑。TME是肿瘤生态系统的核心,是肿瘤细胞与免疫细胞、成纤维细胞及其他基质成分相互作用的关键生态位,其状态直接影响肿瘤进展、免疫监视和治疗反应^[85]。肿瘤细胞及肿瘤相关基质细胞可通过分泌包含蛋白、肽、核酸和脂质的外泌体实现长距离通讯,从而介导免疫抑制和免疫逃逸^[86-87]。外泌体中含有大量 ncRNA,作为 ncRNA 主要类型之一的外泌体 miRNA,已被证实是多种癌症的新生物标志物^[88]。肿瘤相关巨噬细胞来源的外泌体通过分泌 miRNA 介导癌细胞与巨噬细胞之间的通讯,从而影响肿瘤的增殖和迁移^[89]。例如,外泌体中的信号分子可诱导肿瘤浸润性巨噬细胞向免疫抑制性的 M2 型极化状态转变,抑制 CD8⁺ T 细胞的细胞毒性活性,促进肿瘤免疫耐受,这一机

制被认为是肿瘤进展的重要推动因素^[87,90-91]。此外,非经典蛋白及微肽还可通过影响抗原呈递和免疫识别,直接参与 TME 的免疫调控。大量免疫肽组学研究表明,许多来源于非编码区或替代开放阅读框(altORFs)的非经典肽段能够被加工并通过 MHC-I 类分子呈递于肿瘤细胞表面^[92-93]。这些非经典抗原不仅扩展了肿瘤特异性抗原的范畴,更可能通过激活或调节 T 细胞反应,直接改变 TME 中的免疫细胞组成与功能状态,进而影响整体免疫微环境的平衡^[94]。因此,非经典蛋白及其衍生物在调控免疫细胞、影响抗原呈递、重塑免疫微环境等方面扮演着关键角色,其研究为理解 TME 的动态复杂性及开发靶向微环境的治疗策略提供了新的视角与潜在切入点。

综上所述,非经典蛋白在肿瘤生物学中展现出极强的功能多样性,部分目前已报道的具有代表性的肿瘤相关非经典蛋白(参与肿瘤发生发展的非经典蛋白)列于表 3。

4 非经典蛋白临床转换路径

非经典蛋白的发现不仅具有重要的理论意义,更开拓了极具潜力的临床转化路径。鉴于其

表 3 参与肿瘤发生发展的非经典蛋白示例

Table 3 Examples of non-classical proteins involved in tumorigenesis and tumor progression

癌种 Cancer type	代表性蛋白 Representative protein	功能 Function	参考文献 Reference	来源 Origin
胶质母细胞瘤	MPEP(MYC-uORF peptide)	调控MYC表达与代谢	[11]	5'UTR/3'UTR/CDS内重叠ORF产生的暗蛋白
髓母细胞瘤	ASNSD1-uORF microprotein	影响翻译效率	[12]	
肺癌	UBAP1-AST6	定位于细胞核,促进细胞增殖	[55]	lncRNA/lincRNA/伪基因转录本中 sORF编码的暗蛋白
结直肠癌	HOXB-AS3	抑制糖酵解代谢重编程(抑癌)	[14]	
肝细胞癌	MRPIP	抑制线粒体翻译,从而抑制肿瘤生长	[95]	
多种肿瘤	SPAR	抑制mTORC1信号通路(抑癌)	[78]	
乳腺癌	HISLA	外泌体传递,增强肿瘤细胞需氧糖酵解,诱导耐药	[96]	
神经母细胞瘤	NBASP	参与基因组稳定性调节	[13]	
肝细胞癌	FNDC3B	促进上皮-间质转化(EMT)	[15-17]	circRNA驱动的ORF编码蛋白
胶质母细胞瘤	PINT87aa	结合PAF1复合体,抑制癌基因转录	[79]	
黑色素瘤	circPIAS1-108aa	调控免疫检查点表达,影响T细胞杀伤活性	[18]	
肝细胞癌	FNDC3B-219aa	促进上皮-间质转化(EMT)	[97]	
肝细胞癌	PTEN N-terminal extension proteoform	改变PTEN定位与磷酸酶活性	[19-21]	经典CDS基因通过非典型翻译机制(如非-AUG起始、N端延伸)产生的变体蛋白

肿瘤特异性表达模式和多样的生物学功能,非经典蛋白已成为发现新型生物标志物、免疫治疗靶点和精准干预策略的宝贵资源。

4.1 肿瘤标志物

由于非经典蛋白常由肿瘤特异性基因组变异或表观遗传重编程(如异常转录翻译等)所诱导表达,所以它们在体液中的存在或含量变化具有作为新型诊断、预后或疗效监测生物标志物的巨大潜力。相较于传统标志物,非经典蛋白可能表现出更高的肿瘤特异性和早期出现的特征。已有研究表明,多种长链非编码 RNA 存在于肺癌组织中,参与肿瘤的进展、侵袭、转移和增殖过程。例如, GAS5 已被证明可通过抑制血管生成,成为肺癌的潜在治疗靶点^[98]。另有研究发现,在非小细胞肺癌(NSCLC)患者血清中,外泌体 MALAT1 呈高表达,该分子可通过抑制细胞凋亡和缩短细胞周期,加速肿瘤的迁移和增殖^[99];因此,外泌体 MALAT1 可能作为非侵入性生物标志物用于 NSCLC 的诊断,或成为 NSCLC 的潜在治疗靶点。通过质谱或多重抗体检测技术对这些循环微肽进行定量分析,可为实现肿瘤的无创早期诊断和动态监控提供新方案。

4.2 肿瘤新抗原

肿瘤新抗原是源于肿瘤细胞基因突变、能被免疫系统识别为“非己”的蛋白质片段,是激活 T 细胞免疫应答的关键。非经典蛋白是肿瘤新抗原的新来源,这些蛋白源于基因组中的“暗区”,对免疫系统而言是全新的“非己”抗原,因此能有效打破免疫耐受,激发强烈的抗肿瘤 T 细胞反应。非经典蛋白的发现还极大地扩展了新抗原的“库容”。由于它们的合成通常具有肿瘤特异性,因此产生的肽段是近乎完美的免疫靶点,具有高度特异性和极低的自身免疫风险^[100]。另外,由于非经典蛋白的表达高度局限于肿瘤组织,以其为靶点的免疫疗法有望大幅降低对身体的毒副作用。通过整合外显子组测序、Ribo-seq 和质谱数据,可以系统性地鉴定出由非经典 ORF 编码、并经 HLA 递呈的个体化新抗原肽^[101]。例如, Tran 等^[102]在 1 例转移性结直肠癌患者中回输识别 KRAS G12D 突变的 CD8⁺TIL,使患者的 7 个肺转移灶出现完全或接近完全退缩,证明了新抗原特异性 T 细胞治疗在实体瘤中的潜力。此外, Sahin 等^[103]采用患者特异性突变构建了多

表位 mRNA 疫苗,疫苗输注后成功诱导了肿瘤特异性 T 细胞反应,部分患者实现了长期无复发生存。这类研究为以肿瘤新抗原为靶点的免疫策略提供了重要的生物学与临床证据基础。

非经典蛋白的发现打破了对基因组和蛋白质组的传统认知边界,这不仅丰富了肿瘤蛋白质组的复杂性,更开辟了癌症生物标志物挖掘和免疫治疗方案创新的全新方向。随着对这些非经典蛋白的深入探索和系统解析,将深化对肿瘤生物学机制的理解,并最终推动更精准、更高效的癌症诊疗策略的建立与应用^[104]。

4.3 靶向治疗

除免疫治疗外,非经典蛋白本身也可作为直接靶向干预的分子靶点。对于功能确证的致癌性微肽,可采用小分子抑制剂、干扰抗体或降解剂(如 PROTAC)的策略阻断其功能或诱导降解。反之,对于具有抑癌功能的微肽,则可考虑通过 mRNA 疗法或肽类替代疗法将其导入肿瘤细胞内部,重建抑癌信号通路,从而抑制肿瘤生长^[105]。鉴于许多微肽介导肿瘤关键的代谢过程,靶向这些微肽有望逆转肿瘤的代谢重编程,有效克服肿瘤细胞的耐药性。另一方面,非编码 RNA 在这些复杂的代谢网络中也发挥着关键的调控作用,靶向非编码 RNA 以调控其翻译产物或下游代谢网络显著影响肿瘤的代谢状态和生长。有研究发现,肿瘤组织中 5 种 lncRNA 编码的多肽表达水平显著上调,强调了作为有价值癌症生物标志物的潜力^[106]。此外, Humberto 等^[107]在黑色素瘤和肺癌样本免疫肽组中鉴定出 54 种独特的 circRNA 衍生肽,表明这些肽段在免疫治疗中的应用潜力。总体而言,针对非经典蛋白的靶向治疗策略正成为肿瘤精准治疗领域中充满前景的新方向。

5 总结与展望

非经典蛋白的发现是肿瘤生物学领域的一次范式转变,极大地拓展了对癌症蛋白质组复杂性的认知边界。大量证据表明,源自非编码区的微肽等非经典蛋白绝非“转录噪音”,而是积极参与了驱动恶性表型演进、介导肿瘤代谢重编程、重塑肿瘤免疫微环境等核心的生物学过程,构成了一个全新的肿瘤调控维度。多组学技术与生物信息学分析方法的交叉融合,特别是蛋白

质基因组学框架的应用,为系统揭示这座“暗物质”宝藏提供了强大的技术引擎。

从治疗和临床转化的角度看,非经典蛋白与经典蛋白在来源、特性及开发策略上存在本质性差异。经典蛋白来源于已明确注释的蛋白质编码基因,是经过长期研究的已知靶点,其生物学功能及与疾病的关联已较为明确。例如,热休克蛋白(HSPs)作为经典分子伴侣,在肿瘤中的促生存作用已被广泛阐明,针对HSP90的抑制剂研发已持续多年^[108]。经典蛋白的成药路径总体上是优化迭代式的,例如,针对“不可成药”的经典蛋白,可通过改进药物形式(如开发抗体偶联药物或采用PROTAC等新兴技术)进行降解来克服“不可成药”挑战^[109]。尽管经典蛋白的转化路径已非常成熟,但仍面临耐药性及部分靶点难以干预等问题,未来研究方向主要在于开发新一代作用机制的药物(如PROTAC)以及攻克现有疗法的耐药性^[110]。相比之下,非经典蛋白源于基因组“暗物质”,主要来自基因组非编码区(如lncRNA、circRNA或altORF),这些翻译事件通常与特定肿瘤类型高度相关,并不出现在传统注释的蛋白编码基因中,因此具有更强的肿瘤特异性和低背景毒性。这使得它们在作为靶点时具备独特优势,能够绕开由经典蛋白介导的耐药通路。例如,浙江大学团队从肝癌中鉴定出1个来源于lncRNA的微肽MRPIP,其能通过抑制线粒体翻译从而抑制肿瘤,是一个全新的干预环节^[111]。非经典蛋白的成药路径是探索性的,由于其分子质量小、结构独特,可能没有独立的三维折叠域或保守结合口袋,与经典蛋白的成药靶点不同^[28],这使得传统的小分子或抗体策略难以直接应用。因此,非经典蛋白的开发策略通常更具探索性,例如可采用肽模仿分子、合成功能片段、免疫调控或基于结构功能域的设计方法进行特异性干预。有研究针对上述微肽MRPIP,基于其功能和结构域合成了1个由20个氨基酸构成的微肽药物前体分子PMHR,这是一种全新的药物形式^[111]。总而言之,经典蛋白的药物开发是在已知框架内进行深度优化,而非经典蛋白则致力于开辟全新的靶点领域和药物形态,两者在转化医学领域面临不同的机遇与挑战。

尽管如此,该领域的研究仍处于高速发展的早期阶段,还面临着诸多机遇与挑战。首先,需

要进行技术体系的深化升级,开发更高灵敏度、更高通量的翻译组与蛋白质组学技术,特别是原位检测技术,以精准定位低丰度微肽的时空表达动态,并解析其三维空间结构。其次,当前对绝大多数非经典蛋白的功能仍知之甚少,需利用大规模的CRISPR筛选、蛋白质相互作用组学等技术,系统性地阐明其在肿瘤信号调控网络中的功能节点作用。将非经典蛋白从基础发现推向临床应用是最终目标。这不仅需要大规模临床队列验证其作为生物标志物的可靠性,同时还需要建立标准化、自动化、个体化的新抗原筛选流程,以降低研发与应用成本。此外,开发靶向微肽的创新药物将是未来的一个重要方向。与此同时,建立专属于非经典蛋白的整合数据库与知识库,统一注释标准,促进数据共享与数据挖掘,对于加速领域发展至关重要。

总之,对非经典蛋白的探索正在揭开肿瘤生物学全新的一页。这些研究不仅有望带来一批全新的诊断标志物、治疗靶点和免疫治疗策略,推动肿瘤精准医疗的跨越式发展,更将从根本上深化对生命遗传信息流动与调控的理解。

参考文献:

- [1] VENTER J C, ADAMS M D, MYERS E W, LI P W, MURAL R J, SUTTON G G, SMITH H O, YANDELL M, EVANS C A, HOLT R A, GOCAYNE J D, AMANATIDES P, BALLEW R M, HUSON D H, WORTMAN J R, ZHANG Q, KODIRA C D, ZHENG X H, CHEN L, SKUPSKI M, SUBRAMANIAN G, THOMAS P D, ZHANG J, GABOR MIKLOS G L, NELSON C, BRODER S, CLARK A G, NADEAU J, MCKUSICK V A, ZINDER N, LEVINE A J, ROBERTS R J, SIMON M, SLAYMAN C, HUNKAPILLER M, BOLANOS R, DELCHER A, DEW I, FASULO D, FLANIGAN M, FLOREA L, HALPERN A, HANNENHALLI S, KRAVITZ S, LEVY S, MOBARRY C, REINERT K, REMINGTON K, ABU-THREIDEH J, BEASLEY E, BIDDICK K, BONAZZI V, BRANDON R, CARGILL M, CHANDRAMOULISWARAN I, CHARLAB R, CHATURVEDI K, DENG Z, DI FRANCESCO V, DUNN P, EILBECK K, EVANGELISTA C, GABRIELIAN A E, GAN W, GE W, GONG F, GU Z, GUAN P, HEIMAN T J, HIGGINS M E, JI R R, KE Z, KETCHUM K A, LAI Z, LEI Y, LI Z, LI J, LIANG Y, LIN X, LU F, MERKULOV G V, MILSHINA N,

- MOORE H M, NAIK A K, NARAYAN V A, NEELAM B, NUSSKERN D, RUSCH D B, SALZBERG S, SHAO W, SHUE B, SUN J, WANG Z, WANG A, WANG X, WANG J, WEI M, WIDES R, XIAO C, YAN C, YAO A, YE J, ZHAN M, ZHANG W, ZHANG H, ZHAO Q, ZHENG L, ZHONG F, ZHONG W, ZHU S, ZHAO S, GILBERT D, BAUMHUETER S, SPIER G, CARTER C, CRAVCHIK A, WOODAGE T, ALI F, AN H, AWE A, BALDWIN D, BADEN H, BARNSTEAD M, BARROW I, BEESON K, BUSAM D, CARVER A, CENTER A, CHENG M L, CURRY L, DANAHER S, DAVENPORT L, DESILETS R, DIETZ S, DODSON K, DOUP L, FERRIERA S, GARG N, GLUECKSMANN A, HART B, HAYNES J, HAYNES C, HEINER C, HLADUN S, HOSTIN D, HOUCK J, HOWLAND T, IBEGWAM C, JOHNSON J, KALUSH F, KLINE L, KODURU S, LOVE A, MANN F, MAY D, MCCAWLEY S, MCINTOSH T, MCMULLEN I, MOY M, MOY L, MURPHY B, NELSON K, PFANNKOCHE C, PRATTS E, PURI V, QURESHI H, REARDON M, RODRIGUEZ R, ROGERS Y H, ROMBLAD D, RUHFEL B, SCOTT R, SITTER C, SMALLWOOD M, STEWART E, STRONG R, SUH E, THOMAS R, TINT N N, TSE S, VECH C, WANG G, WETTER J, WILLIAMS S, WILLIAMS M, WINDSOR S, WINN-DEEN E, WOLFE K, ZAVERI J, ZAVERI K, ABRIL J F, GUIGÓ R, CAMPBELL M J, SJOLANDER K V, KARLAK B, KEJARIWAL A, MI H, LAZAREVA B, HATTON T, NARECHANIA A, DIEMER K, MURUGANUJAN A, GUO N, SATO S, BAFNA V, ISTRAIL S, LIPPERT R, SCHWARTZ R, WALENZ B, YOOSSEPH S, ALLEN D, BASU A, BAXENDALE J, BLICK L, CAMINHA M, CARNESSTINE J, CAULK P, CHIANG Y H, COYNE M, DAHLKE C, DESLATTES MAYS A, DOMBROSKI M, DONNELLY M, ELY D, ESPARHAM S, FOSLER C, GIRE H, GLANOWSKI S, GLASSER K, GLODEK A, GOROKHOV M, GRAHAM K, GROPMAN B, HARRIS M, HEIL J, HENDERSON S, HOOVER J, JENNINGS D, JORDAN C, JORDAN J, KASHA J, KAGAN L, KRAFT C, LEVITSKY A, LEWIS M, LIU X, LOPEZ J, MA D, MAJOROS W, MCDANIEL J, MURPHY S, NEWMAN M, NGUYEN T, NGUYEN N, NODELL M, PAN S, PECK J, PETERSON M, ROWE W, SANDERS R, SCOTT J, SIMPSON M, SMITH T, SPRAGUE A, STOCKWELL T, TURNER R, VENTER E, WANG M, WEN M, WU D, WU M, XIA A, ZANDIEH A, ZHU X. The sequence of the human genome[J]. *Science*, 2001, 291(5 507): 1 304-1 351.
- [2] WILHELM M, SCHLEGL J, HAHNE H, GHOLAMI A M, LIEBERENZ M, SAVITSKI M M, ZIEGLER E, BUTZMANN L, GESSULAT S, MARX H, MATHIESON T, LEMEER S, SCHNATBAUM K, REIMER U, WENSCHUH H, MOLLENHAUER M, SLOTTA-HUSPENINA J, BOESE J H, BANTSCHIEFF M, GERSTMAIR A, FAERBER F, KUSTER B. Mass-spectrometry-based draft of the human proteome[J]. *Nature*, 2014, 509(7 502): 582-587.
- [3] DEUTSCH E W, KOK L W, MUDGE J M, RUIZ-ORERA J, FIERRO-MONTI I, SUN Z, ABELIN J G, ALBA M M, ASPDEN J L, BAZZINI A A, BRUFORD E A, BRUNET M A, CALVIELLO L, CARR S A, CARVUNIS A R, CHOTHANI S, CLAUWAERT J, DEAN K, FARIDI P, FRANKISH A, HUBNER N, INGOLIA N T, MAGRANE M, MARTIN M J, MARTINEZ T F, MENSCHAERT G, OHLER U, ORCHARD S, RACKHAM O, ROUCOU X, SLAVOFF S A, VALEN E, WACHOLDER A, WEISSMAN J S, WU W, XIE Z, CHOUDHARY J, BASSANI-STERNBERG M, VIZCAÍNO J A, TERNETTE N, MORITZ R L, PRENSNER J R, van HEESCH S. High-quality peptide evidence for annotating non-canonical open reading frames as human proteins[J]. *bioRxiv*, 2025. DOI:10.1101/2024.09.09.612016.
- [4] ZHOU L, WONG L, BIN GOH W W. Understanding missing proteins: a functional perspective[J]. *Drug Discovery Today*, 2018, 23(3): 644-651.
- [5] 何崔彤, 张瑶, 姜颖, 徐平. 小蛋白质富集法鉴定酿酒酵母“漏检蛋白”[J]. *生物工程学报*, 2018, 34(11): 1 860-1 869.
- HE Cuitong, ZHANG Yao, JIANG Ying, XU Ping. Identification of missing proteins in *Saccharomyces cerevisiae* by small protein-based enrichment[J]. *Sheng Wu Gong Cheng Xue Bao*, 2018, 34(11): 1 860-1 869(in Chinese).
- [6] KESNER J S, CHEN Z, SHI P, APARICIO A O, MURPHY M R, GUO Y, TREHAN A, LIPPONEN J E, RECINOS Y, MYEKU N, WU X. Noncoding translation mitigation[J]. *Nature*, 2023, 617(7 960): 395-402.
- [7] MIANO V, CODINO A, PANDOLFINI L, BARBIERI I. The non-coding epitranscriptome in cancer[J]. *Briefings in Functional Genomics*, 2021, 20(2): 94-105.
- [8] PANNI S, LOVERING R C, PORRAS P, ORCHARD S. Non-coding RNA regulatory networks[J]. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 2020, 1 863(6): 194 417.
- [9] WANG J, ZHU S, MENG N, HE Y, LU R, YAN G R.

- ncRNA-encoded peptides or proteins and cancer[J]. *Molecular Therapy*, 2019, 27(10): 1 718-1 725.
- [10] WANG Y, HUANG T, SUN X, WANG Y. Identification of a potential prognostic lncRNA-miRNA-mRNA signature in endometrial cancer based on the competing endogenous RNA network[J]. *Journal of Cellular Biochemistry*, 2019, 120(11): 18 845-18 853.
- [11] LI F, YANG K, GAO X, ZHANG M, GU D, WU X, LU C, WU Q, DIXIT D, GIMPLE R C, YOU Y, MACK S C, SHI Y, KANG T, AGNIHOTRI S A, TAYLOR M D, RICH J N, ZHANG N, WANG X. A peptide encoded by upstream open reading frame of MYC binds to tropomyosin receptor kinase B and promotes glioblastoma growth in mice[J]. *Science Translational Medicine*, 2024, 16(767): eadk9524.
- [12] HOFMAN D A, RUIZ-ORERA J, YANNUZZI I, MURUGESAN R, BROWN A, CLAUSER K R, CONDURAT A L, van DINTER J T, ENGELS S A G, GOODALE A, van der LUGT J, ABID T, WANG L, ZHOU K N, VOGELZANG J, LIGON K L, PHOENIX T N, ROTH J A, ROOT D E, HUBNER N, GOLUB T R, BANDOPADHAYAY P, van HEESCH S, PRENSNER J R. Translation of non-canonical open reading frames as a cancer cell survival mechanism in childhood medulloblastoma[J]. *Molecular Cell*, 2024, 84(2): 261-276. e18.
- [13] YE M, GAO R, CHEN S, BAI J, CHEN J, LU F, GU D, SHI X, YU P, TIAN Y, TANG Q, DONG K. FAM201A encodes small protein NBASP to inhibit neuroblastoma progression *via* inactivating MAPK pathway mediated by FABP5[J]. *Communications Biology*, 2023, 6(1): 714.
- [14] HUANG J, CHEN M, CHEN D, GAO X, ZHU S, HUANG H, HU M, ZHU H, YAN G. A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth[J]. *Molecular Cell*, 2017, 68(1): 171-184. e6.
- [15] PAN Z, CAI J, LIN J, ZHOU H, PENG J, LIANG J, XIA L, YIN Q, ZOU B, ZHENG J, QIAO L, ZHANG L. A novel protein encoded by circFNDC3B inhibits tumor progression and EMT through regulating Snail in colon cancer[J]. *Molecular Cancer*, 2020, 19(1): 71.
- [16] WANG X, HUANG Y, LI S, ZHANG H. Integrated machine learning methods identify FNDC3B as a potential prognostic biomarker and correlated with immune infiltrates in glioma[J]. *Frontiers in Immunology*, 2022, 13: 1 027 154.
- [17] 江金鑫, 马和, 刘娜, 赵瑞欢, 蒋成兰, 何越峰. 亚砷酸钠调控 A549 细胞 FNDC3B 表达的分子机制及其促癌功能[J]. *环境与职业医学*, 2025, 42(6): 691-698.
- JIANG Jinyun, MA He, LIU Na, ZHAO Ruihuan, JIANG Chenglan, HE Yuefeng. Molecular mechanisms and cancer-promoting roles of sodium arsenite in regulating FNDC3B expression in A549 cells[J]. *Journal of Environmental and Occupational Medicine*, 2025, 42(6): 691-698(in Chinese).
- [18] ZANG X, HE X, XIAO C, LIN Q, WANG M, LIU C, KONG L, CHEN Z, XIA Y. Circular RNA-encoded oncogenic PIAS1 variant blocks immunogenic ferroptosis by modulating the balance between SUMOylation and phosphorylation of STAT1[J]. *Molecular Cancer*, 2024, 23(1): 207.
- [19] LIANG H, HE S, YANG J, JIA X, WANG P, CHEN X, ZHANG Z, ZOU X, McNUTT M A, SHEN W, YIN Y. PTEN α , a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism[J]. *Cell Metabolism*, 2014, 19(5): 836-848.
- [20] LIANG H, CHEN X, YIN Q, RUAN D, ZHAO X, ZHANG C, McNUTT M A, YIN Y. PTEN β is an alternatively translated isoform of PTEN that regulates rDNA transcription[J]. *Nature Communications*, 2017, 8: 14 771.
- [21] PANCSA R, ANDREEV D E, DEAN K. The implication of non-AUG-initiated *N*-terminally extended proteoforms in cancer[J]. *RNA Biology*, 2025, 22(1): 1-18.
- [22] DENG X, YU Y, JIN Y. Non-canonical translation in cancer: significance and therapeutic potential of non-canonical ORFs, m6A-modification, and circular RNAs[J]. *Cell Death Discovery*, 2024, 10(1): 412.
- [23] GE A, CHAN C, YANG X, GE A, CHAN C, YANG X. Exploring the dark matter of human proteome: the emerging role of non-canonical open reading frame (ncORF) in cancer diagnosis, biology, and therapy[J]. *Cancers*, 2024, 16(15): 2 660.
- [24] SAMANDI S, ROY A V, DELCOURT V, LUCIER J F, GAGNON J, BEAUDOIN M C, VANDERPERRE B, BRETON M A, MOTARD J, JACQUES J F, BRUNELLE M, GAGNON-ARSENAULT I, FOURNIER I, OUANGRAOUA A, HUNTING D J, COHEN A A, LANDRY C R, SCOTT M S, ROUCOU X. Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins[J]. *eLife*, 2017, 6: e27860.
- [25] WRIGHT B W, YI Z, WEISSMAN J S, CHEN J. The dark proteome: translation from noncanonical open reading frames[J]. *Trends in Cell Biology*, 2022, 32(3): 243-258.

- [26] ZHENG C, WEI Y, ZHANG P, LIN K, HE D, TENG H, MANYAM G, ZHANG Z, LIU W, LEE H R L, TANG X, HE W, ISLAM N, JAIN A, CHIU Y, CAO S, DIAO Y, MEYER-GAUEN S, HÖÖK M, MALOVAN-NAYA A, LI W, HU M, WANG W, XU H, KOPETZ S, CHEN Y. CRISPR-Cas9-based functional interrogation of unconventional translome reveals human cancer dependency on cryptic non-canonical open reading frames[J]. *Nature Structural & Molecular Biology*, 2023, 30(12): 1 878-1 892.
- [27] ZHANG Y. LncRNA-encoded peptides in cancer[J]. *Journal of Hematology & Oncology*, 2024, 17(1): 66.
- [28] YE M, ZHANG J, WEI M, LIU B, DONG K. Emerging role of long noncoding RNA-encoded micropeptides in cancer[J]. *Cancer Cell International*, 2020, 20(1): 506.
- [29] MENG K, LI Y, YUAN X, SHEN H, HU L, LIU D, SHI F, ZHENG D, SHI X, WEN N, CAO Y, PAN Y, HE Q, ZHANG C. The cryptic lncRNA-encoded microprotein TPM3P9 drives oncogenic RNA splicing and tumorigenesis[J]. *Signal Transduction and Targeted Therapy*, 2025, 10: 43.
- [30] HU H, HAN L, FU J, HE X, TAN J, CHEN Q, HAN J, HE Q. LINC00982-encoded protein PRDM16-DT regulates CHEK2 splicing to suppress colorectal cancer metastasis and chemoresistance[J]. *Theranostics*, 2024, 14(8): 3 317-3 338.
- [31] ZHANG T, LI Z, LI J, PENG Y. Small open reading frame-encoded microproteins in cancer: identification, biological functions and clinical significance[J]. *Molecular Cancer*, 2025, 24(1): 105.
- [32] IVANOV I P, FIRTH A E, MICHEL A M, ATKINS J F, BARANOV P V. Identification of evolutionarily conserved non-AUG-initiated *N*-terminal extensions in human coding sequences[J]. *Nucleic Acids Research*, 2011, 39(10): 4 220-4 234.
- [33] KEARSE M G, WILUSZ J E. Non-AUG translation: a new start for protein synthesis in eukaryotes[J]. *Genes & Development*, 2017, 31(17): 1 717-1 731.
- [34] ANDREEV D E, LOUGHRAN G, FEDOROVA A D, MIKHAYLOVA M S, SHATSKY I N, BARANOV P V. Non-AUG translation initiation in mammals[J]. *Genome Biology*, 2022, 23(1): 111.
- [35] INGOLIA N T, BRAR G A, ROUSKIN S, McGEACHY A M, WEISSMAN J S. The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments[J]. *Nature Protocols*, 2012, 7(8): 1 534-1 550.
- [36] WACHOLDER A, CARVUNIS A R. Biological factors and statistical limitations prevent detection of most noncanonical proteins by mass spectrometry[J]. *PLoS Biology*, 2023, 21(12): e3002409.
- [37] PAPAPOPOULOS C, ARBES H, CORNU D, CHEVROLLIER N, BLANCHET S, ROGINSKI P, RABIER C, ATIA S, LESPINET O, NAMY O, LOPES A. The ribosome profiling landscape of yeast reveals a high diversity in pervasive translation[J]. *Genome Biology*, 2024, 25(1): 268.
- [38] RUIZ-ORERA J, HÜBNER N. The non-canonical proteome: a novel contributor to cancer proliferation[J]. *Cell Research*, 2025, 35(3): 155-156.
- [39] EASTMAN G, SMIRCICH P, SOTELO-SILVEIRA J R. Following ribosome footprints to understand translation at a genome wide level[J]. *Computational and Structural Biotechnology Journal*, 2018, 16: 167-176.
- [40] INGOLIA N T, GHAEMMAGHAMI S, NEWMAN J R S, WEISSMAN J S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling[J]. *Science*, 2009, 324(5 924): 218-223.
- [41] LI G W, OH E, WEISSMAN J S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria[J]. *Nature*, 2012, 484(7 395): 538-541.
- [42] JUNTAWONG P, GIRKE T, BAZIN J, BAILEY-SERRES J. Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis[J]. *PNAS*, 2014, 111(1): E203-E212.
- [43] SMITH J E, BAKER K E. Nonsense-mediated RNA decay—a switch and dial for regulating gene expression[J]. *BioEssays*, 2015, 37(6): 612-623.
- [44] CALVIELLO L, OHLER U. Beyond read-counts: ribo-seq data analysis to understand the functions of the transcriptome[J]. *Trends in Genetics*, 2017, 33(10): 728-744.
- [45] CAMARENA M E, THEUNISSEN P, RUIZ M, RUIZ-ORERA J, CALVO-SERRA B, CASTELO R, CASTRO C, SAROBE P, FORTES P, PERERA-BEL J, ALBÀ M M. Microproteins encoded by noncanonical ORFs are a major source of tumor-specific antigens in a liver cancer patient meta-cohort[J]. *Science Advances*, 2024, 10(28): eadn3628.
- [46] WANG T, CUI Y, JIN J, GUO J, WANG G, YIN X, HE Q Y, ZHANG G. Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific[J]. *Nucleic Acids Research*, 2013, 41(9): 4 743-4 754.
- [47] KOZLOVA A, SARYGINA E, ILGISONIS E, TARBEEVA S, PONOMARENKO E. The translome map:

- RNC-seq vs. ribo-seq for profiling of HBE, A549, and MCF-7 cell lines[J]. *International Journal of Molecular Sciences*, 2024, 25(20): 10 970.
- [48] WU C, LU X, LU S, WANG H, LI D, ZHAO J, JIN J, SUN Z, HE Q Y, CHEN Y, ZHANG G. Efficient detection of the alternative spliced human proteome using translome sequencing[J]. *Frontiers in Molecular Biosciences*, 2022, 9: 895 746.
- [49] RAJINIKANTH N, CHAUHAN R, PRABAKARAN S. Harnessing noncanonical proteins for next-generation drug discovery and diagnosis[J]. *WIREs Mechanisms of Disease*, 2025, 17(3): e70001.
- [50] ERADY C, BOXALL A, PUNTAMBEKAR S, SUHAS JAGANNATHAN N, CHAUHAN R, CHONG D, MEENA N, KULKARNI A, KASABE B, PRATHIVADI BHAYANKARAM K, UMRANIA Y, ANDREANI A, NEL J, WAYLAND M T, PINA C, LILLEY K S, PRABAKARAN S. Pan-cancer analysis of transcripts encoding novel open-reading frames (nORFs) and their potential biological functions[J]. *npj Genomic Medicine*, 2021, 6(1): 4.
- [51] SHEYINKMAN G M, SHORTREED M R, CESNIK A J, SMITH L M. Proteogenomics: integrating next-generation sequencing and mass spectrometry to characterize human proteomic variation[J]. *Annual Review of Analytical Chemistry*, 2016, 9(1): 521-545.
- [52] PRABAKARAN S, HEMBERG M, CHAUHAN R, WINTER D, TWEEDIE-CULLEN R Y, DITTRICH C, HONG E, GUNAWARDENA J, STEEN H, KREIMAN G, STEEN J A. Quantitative profiling of peptides from RNAs classified as noncoding[J]. *Nature Communications*, 2014, 5: 5 429.
- [53] SLAVOFF S A, HEO J, BUDNIK B A, HANAKAHI L A, SAGHATELIAN A. A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining[J]. *Journal of Biological Chemistry*, 2014, 289(16): 10 950-10 957.
- [54] PRENSNER J R, ABELIN J G, KOK L W, CLAUSER K R, MUDGE J M, RUIZ-ORERA J, BASSANISTERBERG M, DEUTSCH E W, van HEESCH S. What can Ribo-seq and proteomics tell us about the non-canonical proteome?[J]. *bioRxiv*, 2023. DOI:10.1101/2023.05.16.541049.
- [55] LU S, ZHANG J, LIAN X, SUN L, MENG K, CHEN Y, SUN Z, YIN X, LI Y, ZHAO J, WANG T, ZHANG G, HE Q. A hidden human proteome encoded by 'non-coding' genes[J]. *Nucleic Acids Research*, 2019, 47(15): 8 111-8 125.
- [56] MERTINS P, MANI D R, RUGGLES K V, GILLETTE M A, CLAUSER K R, WANG P, WANG X, QIAO J W, CAO S, PETRALIA F, KAWALER E, MUNDT F, KRUG K, TU Z, LEI J T, GATZA M L, WILKERSON M, PEROU C M, YELLAPANTULA V, HUANG K L, LIN C, McLELLAN M D, YAN P, DAVIES S R, TOWNSEND R R, SKATES S J, WANG J, ZHANG B, KINSINGER C R, MESRI M, RODRIGUEZ H, DING L, PAULOVICH A G, FENYŐ D, ELLIS M J, CARR S A. Proteogenomics connects somatic mutations to signalling in breast cancer[J]. *Nature*, 2016, 534(7 605): 55-62.
- [57] ZHANG B, WHITEAKER J R, HOOFNAGLE A N, BAIRD G S, RODLAND K D, PAULOVICH A G. Clinical potential of mass spectrometry-based proteogenomics[J]. *Nature Reviews Clinical Oncology*, 2019, 16(4): 256-268.
- [58] SODA M, CHOI Y L, ENOMOTO M, TAKADA S, YAMASHITA Y, ISHIKAWA S, FUJIWARA S I, WATANABE H, KURASHINA K, HATANAKA H, BANDO M, OHNO S, ISHIKAWA Y, ABURATANI H, NIKI T, SOHARA Y, SUGIYAMA Y, MANO H. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer[J]. *Nature*, 2007, 448(7 153): 561-566.
- [59] HOWLADER N, FORJAZ G, MOORADIAN M J, MEZA R, KONG C Y, CRONIN K A, MARIOTTO A B, LOWY D R, FEUER E J. The effect of advances in lung-cancer treatment on population mortality[J]. *The New England Journal of Medicine*, 2020, 383(7): 640-649.
- [60] JEMAL M. High-throughput quantitative bioanalysis by LC/MS/MS[J]. *Biomedical Chromatography*, 2000, 14(6): 422-429.
- [61] McCORMACK A L, SCHIELTZ D M, GOODE B, YANG S, BARNES G, DRUBIN D, YATES J R. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level[J]. *Analytical Chemistry*, 1997, 69(4): 767-776.
- [62] NEUBERT H, SHUFORD C M, OLAH T V, GAROFOLO F, SCHULTZ G A, JONES B R, AMARAVADI L, LATERZA O F, XU K, ACKERMANN B L. Protein biomarker quantification by immunoaffinity liquid chromatography-tandem mass spectrometry: current state and future vision[J]. *Clinical Chemistry*, 2020, 66(2): 282-301.
- [63] LUO S, PENG H, SHI Y, CAI J, ZHANG S, SHAO N, LI J. Integration of proteomics profiling data to facilitate discovery of cancer neoantigens: a survey[J]. *Brief-*

- ings in *Bioinformatics*, 2025, 26(2): bbaf087.
- [64] TRAN N H, ZHANG X, XIN L, SHAN B, LI M. *De novo* peptide sequencing by deep learning[J]. *PNAS*, 2017, 114(31): 8 247-8 252.
- [65] KARUNRATANAKUL K, TANG H Y, SPEICHER D W, CHUANGSUWANICH E, SRISWASDI S. Uncovering thousands of new peptides with sequence-mask-search hybrid *de novo* peptide sequencing framework[J]. *Molecular & Cellular Proteomics*, 2019, 18(12): 2 478-2 491.
- [66] OLEXIOUK V, CRAPPÉ J, VERBRUGGEN S, VERHEGEN K, MARTENS L, MENSCHAERT G. sORFs.org: a repository of small ORFs identified by ribosome profiling[J]. *Nucleic Acids Research*, 2016, 44(D1): D324-D329.
- [67] BRUNET M A, BRUNELLE M, LUCIER J F, DELCOURT V, LEVESQUE M, GRENIER F, SAMANDI S, LEBLANC S, AGUILAR J D, DUFOUR P, JACQUES J F, FOURNIER I, OUANGRAOUA A, SCOTT M S, BOISVERT F M, ROUCOU X. OpenProt: a more comprehensive guide to explore eukaryotic coding potential and proteomes[J]. *Nucleic Acids Research*, 2019, 47(D1): D403-D410.
- [68] WANG X, ZHANG B. customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search[J]. *Bioinformatics*, 2013, 29(24): 3 235-3 237.
- [69] ZHANG S, HU H, ZHOU J, HE X, JIANG T, ZENG J. Analysis of ribosome stalling and translation elongation dynamics by deep learning[J]. *Cell Systems*, 2017, 5(3): 212-220. e6.
- [70] SUN J, RU J, CRIBBS A P, XIONG D. PyPropel: a Python-based tool for efficiently processing and characterising protein data[J]. *BMC Bioinformatics*, 2025, 26(1): 70.
- [71] KANG Y, YANG D, KONG L, HOU M, MENG Y, WEI L, GAO G. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features[J]. *Nucleic Acids Research*, 2017, 45(W1): W12-W16.
- [72] POCKRANDT C, STEINEGGER M, SALZBERG S L. PhyloCSF++: a fast and user-friendly implementation of PhyloCSF with annotation tools[J]. *Bioinformatics*, 2022, 38(5): 1 440-1 442.
- [73] YANG T, WANG C, TSAI H C, LIU C. Human IRES Atlas: an integrative platform for studying IRES-driven translational regulation in humans[J]. *Database*, 2021, 2021: baab025.
- [74] LEBLANC S, YALA F, PROVENCHER N, LUCIER J F, LEVESQUE M, LAPOINTE X, JACQUES J F, FOURNIER I, SALZET M, OUANGRAOUA A, SCOTT M S, BOISVERT F M, BRUNET M A, ROUCOU X. OpenProt 2.0 builds a path to the functional characterization of alternative proteins[J]. *Nucleic Acids Research*, 2024, 52(D1): D522-D528.
- [75] ENGREITZ J M, HAINES J E, PEREZ E M, MUNDSON G, CHEN J, KANE M, McDONEL P E, GUTTMAN M, LANDER E S. Local regulation of gene expression by lncRNA promoters, transcription and splicing[J]. *Nature*, 2016, 539(7 629): 452-455.
- [76] GAO X, XIA X, LI F, ZHANG M, ZHOU H, WU X, ZHONG J, ZHAO Z, ZHAO K, LIU D, XIAO F, XU Q, JIANG T, LI B, CHENG S Y, ZHANG N. Circular RNA-encoded oncogenic E-cadherin variant promotes glioblastoma tumorigenicity through activation of EGFR-STAT3 signalling[J]. *Nature Cell Biology*, 2021, 23(3): 278-291.
- [77] WU X, XIAO S, ZHANG M, YANG L, ZHONG J, LI B, LI F, XIA X, LI X, ZHOU H, LIU D, HUANG N, YANG X, XIAO F, ZHANG N. A novel protein encoded by circular SMO RNA is essential for Hedgehog signaling activation and glioblastoma tumorigenicity[J]. *Genome Biology*, 2021, 22(1): 33.
- [78] MATSUMOTO A, PASUT A, MATSUMOTO M, YAMASHITA R, FUNG J, MONTELEONE E, SAGHATELIAN A, NAKAYAMA K I, CLOHESSY J G, PANDOLFI P P. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide[J]. *Nature*, 2017, 541(7 636): 228-232.
- [79] ZHANG M, ZHAO K, XU X, YANG Y, YAN S, WEI P, LIU H, XU J, XIAO F, ZHOU H, YANG X, HUANG N, LIU J, HE K, XIE K, ZHANG G, HUANG S, ZHANG N. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma[J]. *Nature Communications*, 2018, 9(1): 4 475.
- [80] LU Y, LU D, LI C, CHEN L, LU Y, LU D, LI C, CHEN L. Exploring immune cell infiltration and small molecule compounds for ulcerative colitis treatment[J]. *Genes*, 2024, 15(12): 1 548.
- [81] DELL'ANTONE P. Energy metabolism in cancer cells: how to explain the Warburg and Crabtree effects?[J]. *Medical Hypotheses*, 2012, 79(3): 388-392.
- [82] XIA X, LI X, LI F, WU X, ZHANG M, ZHOU H, HUANG N, YANG X, XIAO F, LIU D, YANG L, ZHANG N. A novel tumor suppressor protein encoded by circular AKT3 RNA inhibits glioblastoma tumorigenicity by competing with active phosphoinositide-

- dependent Kinase-1[J]. *Molecular Cancer*, 2019, 18(1): 131.
- [83] YANG Y, GAO X, ZHANG M, YAN S, SUN C, XIAO F, HUANG N, YANG X, ZHAO K, ZHOU H, HUANG S, XIE B, ZHANG N. Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis[J]. *Journal of the National Cancer Institute*, 2018, 110(3): 304-315.
- [84] CHEN F, CHEN J, YANG L, LIU J, ZHANG X, ZHANG Y, TU Q, YIN D, LIN D, WONG P P, HUANG D, XING Y, ZHAO J, LI M, LIU Q, SU F, SU S, SONG E. Extracellular vesicle-packaged HIF-1 α -stabilizing lncRNA from tumour-associated macrophages regulates aerobic glycolysis of breast cancer cells[J]. *Nature Cell Biology*, 2019, 21(4): 498-510.
- [85] KLOOSTERMAN D J, AKKARI L. Macrophages at the interface of the co-evolving cancer ecosystem[J]. *Cell*, 2023, 186(8): 1 627-1 651.
- [86] LI Y, YOU J, ZOU Z, SUN G, SHI Y, SUN Y, XU S, ZHANG X. Decoding the tumor microenvironment: exosome-mediated macrophage polarization and therapeutic frontiers[J]. *International Journal of Biological Sciences*, 2025, 21(9): 4 187-4 214.
- [87] CHEN J, LIN Z, LIU L, ZHANG R, GENG Y, FAN M, ZHU W, LU M, LU L, JIA H, ZHANG J, QIN L X. GOLM1 exacerbates CD8⁺ T cell suppression in hepatocellular carcinoma by promoting exosomal PD-L1 transport into tumor-associated macrophages[J]. *Signal Transduction and Targeted Therapy*, 2021, 6(1): 397.
- [88] JIN X, CHEN Y, CHEN H, FEI S, CHEN D, CAI X, LIU L, LIN B, SU H, ZHAO L, SU M, PAN H, SHEN L, XIE D, XIE C. Evaluation of tumor-derived exosomal miRNA as potential diagnostic biomarkers for early-stage non-small cell lung cancer using next-generation sequencing[J]. *Clinical Cancer Research*, 2017, 23(17): 5 311-5 319.
- [89] SU T, ZHANG P, ZHAO F, ZHANG S. Exosomal microRNAs mediating crosstalk between cancer cells with cancer-associated fibroblasts and tumor-associated macrophages in the tumor microenvironment[J]. *Frontiers in Oncology*, 2021, 11: 631 703.
- [90] SUN Z, XU Y, SHAO B, DANG P, HU S, SUN H, CHEN C, WANG C, LIU J, LIU Y, HU J. Exosomal circPOLQ promotes macrophage M2 polarization *via* activating IL-10/STAT3 axis in a colorectal cancer model[J]. *Journal for Immunotherapy of Cancer*, 2024, 12(5): e008491.
- [91] WANG F, LI B, WEI Y, ZHAO Y, WANG L, ZHANG P, YANG J, HE W, CHEN H, JIAO Z, LI Y. Tumor-derived exosomes induce PD1⁺ macrophage population in human gastric cancer that promotes disease progression[J]. *Oncogenesis*, 2018, 7: 41.
- [92] BEDRAN G, GASSER H C, WEKE K, WANG T, BEDRAN D, LAIRD A, BATTAIL C, ZANZOTTO F M, PESQUITA C, AXELSON H, RAJAN A, HARRISON D J, PALKOWSKI A, PAWLIK M, PARYS M, O'NEILL J R, BRENNAN P M, SYMEONIDES S N, GOODLETT D R, LITCHFIELD K, FAHRAEUS R, HUPP T R, KOTE S, ALFARO J A. The immunopeptidome from a genomic perspective: establishing the noncanonical landscape of MHC class I-associated peptides[J]. *Cancer Immunology Research*, 2023, 11(6): 747-762.
- [93] BIAN Y, LI J, CAO J, WANG S, MA C. Advances in tumor-derived exosomal non-coding RNAs regulating M2 macrophage polarization: molecular mechanisms and signaling pathway[J]. *Cancer Medicine*, 2025, 14(23): e71421.
- [94] LIU C, LIU D, ZHENG X, GUAN J, ZHOU X, ZHANG H, WANG S, LI Q, LU G, HE J, MA C. Munc13-4 mediates tumor immune evasion by regulating the sorting and secretion of PD-L1 *via* exosomes[J]. *Nature Communications*, 2025, 16: 9 080.
- [95] GIDENER T, DIERKHISING R A, MARA K C, THERNEAU T M, VENKATESH S K, EHMAN R L, YIN M, ALLEN A M. Change in serial liver stiffness measurement by magnetic resonance elastography and outcomes in NAFLD[J]. *Hepatology*, 2023, 77(1): 268-274.
- [96] KHAMESPANAH F, MARX M, CROCHET D B, POKHAREL U R, FRONCZEK F R, MAVERICK A W, BELLER M. Oxalate production *via* oxidation of ascorbate rather than reduction of carbon dioxide[J]. *Nature Communications*, 2021, 12(1): 1 997.
- [97] XIONG P, ZHANG F, ZHANG X, WANG S, LIU H, SUN B, ZHANG J, SUN Y, MA R, BANDO Y, ZHOU C, LIU Z, SASAKI T, WANG G. Strain engineering of two-dimensional multilayered heterostructures for beyond-lithium-based rechargeable batteries[J]. *Nature Communications*, 2020, 11: 3 297.
- [98] CHENG Y, DAI X, YANG T, ZHANG N, LIU Z, JIANG Y. Low long noncoding RNA growth arrest-specific transcript 5 expression in the exosomes of lung cancer cells promotes tumor angiogenesis[J]. *Journal of Oncology*, 2019, 2019: 2 476 175.
- [99] ZHANG R, XIA Y, WANG Z, ZHENG J, CHEN Y, LI X, WANG Y, MING H. Serum long non coding RNA MALAT-1 protected by exosomes is up-regulated and promotes cell proliferation and migration in non-small

- cell lung cancer[J]. *Biochemical and Biophysical Research Communications*, 2017, 490(2): 406-414.
- [100] ZHANG T, CELIKER B, SHAO Y, GAI J, HILL M, WANG C, ZHENG L. Comparison of shared class I HLA-bound noncanonical neoepitopes between normal and neoplastic tissues of pancreatic adenocarcinoma[J]. *Clinical Cancer Research*, 2025, 31(10): 1956-1965.
- [101] LAUMONT C M, VINCENT K, HESNARD L, AUDEMARD É, BONNEIL É, LAVERDURE J P, GENDRON P, COURCELLES M, HARDY M P, CÔTÉ C, DURETTE C, ST-PIERRE C, BENHAMMADI M, LANOIX J, VOBECKY S, HADDAD E, LEMIEUX S, THIBAUT P, PERREAULT C. Noncoding regions are the main source of targetable tumor-specific antigens[J]. *Science Translational Medicine*, 2018, 10(470): eaau5516.
- [102] TRAN E, ROBBINS P F, LU Y C, PRICKETT T D, GARTNER J J, JIA L, PASETTO A, ZHENG Z, RAY S, GROH E M, KRILEY I R, ROSENBERG S A. T-cell transfer therapy targeting mutant KRAS in cancer[J]. *The New England Journal of Medicine*, 2016, 375(23): 2255-2262.
- [103] SAHIN U, DERHOVANESSIAN E, MILLER M, KLOKE B P, SIMON P, LÖWER M, BUKUR V, TADMOR A D, LUXEMBURGER U, SCHRÖRS B, OMOKOKO T, VORMEHR M, ALBRECHT C, PARUZYNSKI A, KUHN A N, BUCK J, HEESCH S, SCHREEB K H, MÜLLER F, ORTSEIFER I, VOGLER I, GODEHARDT E, ATTIG S, RAE R, BREITKREUZ A, TOLLIVER C, SUCHAN M, MARTIC G, HOHBERGER A, SORN P, DIEKMANN J, CIESLA J, WAKSMANN O, BRÜCK A K, WITT M, ZILLGEN M, ROTHERMEL A, KASEMANN B, LANGER D, BOLTE S, DIKEN M, KREITER S, NEMECEK R, GEBHARDT C, GRABBE S, HÖLLER C, UTIKAL J, HUBER C, LOQUAI C, TÜRECI Ö. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer[J]. *Nature*, 2017, 547(7662): 222-226.
- [104] CAI Y, LI D, LV D, YU J, MA Y, JIANG T, DING N, LIU Z, LI Y, XU J. MHC-I-presented non-canonical antigens expand the cancer immunotherapy targets in acute myeloid leukemia[J]. *Scientific Data*, 2024, 11: 831.
- [105] YARMARKOVICH M, MARSHALL Q F, WARRINGTON J M, PREMARATNE R, FARREL A, GROFF D, LI W, Di MARCO M, RUNBECK E, TRUONG H, TOOR J S, TRIPATHI S, NGUYEN S, SHEN H, NOEL T, CHURCH N L, WEINER A, KENDSERSKY N, MARTINEZ D, WEISBERG R, CHRISTIE M, EISENLOHR L, BOSSE K R, DIMITROV D S, STEVANOVIC S, SGOURAKIS N G, KIEFEL B R, MARIS J M. Targeting of intracellular oncoproteins with peptide-centric CARs[J]. *Nature*, 2023, 623(7988): 820-827.
- [106] CHAKRABORTY S, ANDRIEUX G, MAHMUDUL HASAN A M, AHMED M, HOSEN M I, RAHMAN T, HOSSAIN M A, BOERRIES M. Harnessing the tissue and plasma lncRNA-peptidome to discover peptide-based cancer biomarkers[J]. *Scientific Reports*, 2019, 9(1): 12322.
- [107] FERREIRA H J, STEVENSON B J, PAK H, YU F, ALMEIDA OLIVEIRA J, HUBER F, TAILLANDIER-COINDARD M, MICHAUX J, RICART-ALTIMIRAS E, KRAEMER A I, KANDALAFI L E, SPEISER D E, NESVIZHSHKII A I, MÜLLER M, BASSANI-STERNBERG M. Immunopeptidomics-based identification of naturally presented non-canonical circRNA-derived peptides[J]. *Nature Communications*, 2024, 15(1): 2357.
- [108] ZUO W F, PANG Q, ZHU X, YANG Q Q, ZHAO Q, HE G, HAN B, HUANG W. Heat shock proteins as hallmarks of cancer: insights from molecular mechanisms to therapeutic strategies[J]. *Journal of Hematology & Oncology*, 2024, 17(1): 81.
- [109] MOON S, LEE B H. Chemically induced cellular proteolysis: an emerging therapeutic strategy for undruggable targets[J]. *Molecules and Cells*, 2018, 41(11): 933-942.
- [110] OMAR E A, RAJESH R, DAS P K, PAL R, PURAWARGA MATADA G S, MAJI L. Next-generation cancer therapeutics: PROTACs and the role of heterocyclic warheads in targeting resistance[J]. *European Journal of Medicinal Chemistry*, 2025, 281: 117034.
- [111] ZHU L, LIU F, SHI C, SU X, TAN M, XIE S, YU M, ZOU S, TAN Y, XIE S, LIU J, YAN Q, NECULAI D, SUN Q, LIU W, DING Y, FU X, SHAO J, LI X, DING K, YUAN Y, ZHOU T, LIN A. Hepatic micropeptide modulates mitochondrial RNA processing machinery in hepatocellular carcinoma[J]. *Molecular Cell*, 2025, 85(12): 2303-2319. e7.

(收稿日期: 2025-11-01; 修回日期: 2026-01-13)