

[**专家简介**] 何昆仑，解放军总医院医学创新研究部某中心主任，主任医师、教授，主要从事医疗大数据、医学人工智能、智能卫勤装备和系统等研究。担任智能医学工程国家级重点实验室主任，承担重大项目等27项，制定国家行业标准29项，授权发明专利103项，获国家科技进步二等奖1项，省部级一、二等奖5项。担任中国研究型医院学会副会长、北京医师协会副会长等。



[**专家简介**] 石金龙，解放军总医院医学创新研究部某中心副主任，博士，高级工程师，硕士研究生导师，主要从事医学大数据与人工智能、生物信息处理与精准医学等应用研究。任中国医院协会信息专委会(CHIMA)、中国生物工程学会计算生物学与生物信息学专委会、中国生物医学工程学会临床医学工程专委会委员等职务。承担国家自然科学基金、国家重点研发计划、国家重大科研仪器研制等项目及子课题10余项，授权国家发明专利20余项。曾获军队优秀专业技术人才岗位津贴、解放军总医院科技进步一等奖1项、军队科技进步三等奖2项等。

生物大数据：现状、挑战与未来发展

王霈虹，何昆仑，石金龙
解放军总医院医学创新研究部，医疗大数据应用技术国家工程研究中心，解放军总医院医学工程实验室，国家药监局人工智能医疗器械研究与评价重点实验室，北京　100853

摘要：生物大数据通常指包括基因组、转录组、蛋白质组、表观遗传组、代谢组、微生物组等在内的各类生命组学数据，其数字化记录了以双链DNA为遗传信息源头的多层次生命活动，并反映了生命系统对环境变化的响应。生物大数据的爆发性增长得益于高通量实验技术的快速发展和广泛应用。同时，生物、数学、计算机等多学科技术的深度融合应用，极大提升了生物大数据的分析处理能力，推动了现代生物医学的创新与临床应用。本文主要对构成生物大数据的各类组学数据及其在生物医学中的应用进行综述，同时对其未来发展进行展望。

关键词：计算生物学；大数据；精准医学；多组学；机器学习
中图分类号：R394.8　　**文献标志码：**A　　**文章编号：**2095-5227(2025)01-0022-06
DOI：10.12435/j.issn.2095-5227.24072504
引用本文：王霈虹，何昆仑，石金龙.生物大数据：现状、挑战与未来发展［J］.解放军医学院学报，2025，46（1）：22-27.

Big data in biology: Status quos, challenges, and future development

WANG Peihong, HE Kunlun, SHI Jinlong
Medical Innovation Research Department of PLA General Hospital, National Engineering Research Center for Medical Big Data Application Technology, Medical Engineering Laboratory of PLA General Hospital, Key Laboratory of Artificial Intelligence Medical Device Research and Evaluation of the National Medical Products Administration, Beijing 100853, China
Corresponding author: HE Kunlun. Email: kunlunhe@301hospital.com.cn　SHI Jinlong. Email: shijinlong@301hospital.com.cn

Abstract: Big data in biology generally refers to omics data generated by multiple high-throughput screening technologies, encompassing genomics, transcriptomics, proteomics, epigenomics, metabolomics, microbiomics, et al. Essentially, it can be regarded as the digital record of multi-layered biological activities derived from genetic information flow, and the responses of life

收稿日期：2024-07-25
第一作者：王霈虹，博士，助理研究员。Email: wangpeihong@301hospital.com.cn
通信作者：何昆仑，博士，主任医师，教授，博士生导师。Email: kunlunhe@301hospital.com.cn
石金龙，博士，高级工程师。Email: shijinlong@301hospital.com.cn

systems to environmental changes. The explosive growth of omics data can be attributed to the rapid advances and widespread adoption of high-throughput sequencing technologies. Simultaneously, the deep integration and application of techniques from biology, mathematics, and computer science have enhanced the analytical and processing capabilities of biological big data. This integration has brought revolutionary impacts on the exploration, innovation, and clinical applications in modern biomedicine. This article primarily provides a review of the various omics data constituting biological big data and their applications in biomedicine, while also outlining future development directions.

Keywords: computational biology; big data; precision medicine; multiomics; machine learning

Cited as: Wang PH, He KL, Shi JL. Big data in biology: Status quos, challenges, and future development[J]. Acad J Chin PLA Med Sch, 2025, 46(1): 22-27.

人类基因组计划的完成标志着生物大数据时代的开启[1]。高通量测序技术的快速发展和多组学领域的拓展，推动了数据规模的爆发性增长，并形成了涵盖数据生成、处理分析和应用场景的完整技术体系，使得生物大数据的内涵和应用得以持续拓展[2]。生物大数据以遗传信息为核心，揭示了生命复杂性的多维特征，推动了疾病相关遗传变异或信号通路的精准辨识，为解析重大和罕见疾病的分子机制提供了基础[3-5]。本文综述了以基因组学、转录组学、蛋白质组学、表观基因组学、代谢组学和微生物组学为核心的生物大数据的特征及其在生物医学研究中的重要应用，并探讨了生物大数据可能面临的困难与挑战，还展望了基于新兴组学技术及人工智能驱动的生物大数据未来发展趋势，以期为相关领域研究者提供参考与启示。

1 生物大数据特征、类型及其应用

生物大数据指在生物学研究中生成和使用的复杂或大型数据集，其主要特征可归纳为5个方面，即大量(Volume)、高速(Velocity)、多样性(Variety)、低价值密度(Value)和真实性(Veracity)。一次高通量测序即可在短时间内产生数百GB至TB级的数据。这些数据类型丰富，体量庞大，但其中真正有价值的信息仅占很小比例，需要通过复杂的数据挖掘和分析加以提取和利用。以基因组学、转录组学、表观基因组学、蛋白质组学、代谢组学和微生物组学为代表的生物大数据，能够反映生命系统对环境及内在状态变化的动态响应，在医学等领域展现出广泛而深远的应用前景。

1.1 基因组学

基因组学研究涵盖生物体所有DNA序列及其相互关系，旨在揭示这些序列对生物体的影响，并为个体或群体的进化解析及变异识别提供基础。作为解读与疾病相关基因变异、识别复杂疾病潜在治疗靶点的核心方法，基因组学在遗传病诊断、疾病风险预测、产前诊断和新生儿筛查等领域广泛应用[6-7]。全基因组关联研究(genome - wide association study，GWAS)是基因组学的典型应用之一。GWAS通过检测人类基因组中存在的序列变异，识别基因型与表型关联，确定疾病和特征的风险位点，可揭示个体的遗传易感性并预测患病风险[6]。然而，GWAS识别的变异并非总是与疾病具有直接的生物学相关性，因此可以进一步使用因果推理确定风险位点的靶基因，如孟德尔随机化(Mendelian randomization，MR)[8]或数量性状基因位点(quantitative trait locus，QTL)分析[9]等。通过基因组测序，还可以识别致病突变和染色体异常，进而进行精准的疾病诊断或产前筛查等。

1.2 转录组学

转录组学数据涉及多种RNA类别，包括信使RNA、非编码RNA、环状RNA等，这些小分子在细胞生理和病理过程中具有重要功能。通过高通量测序或高密度生物芯片技术，转录组学能够定量特定条件下基因的表达水平，为疾病诊断、疾病机制解析以及精准治疗提供重要信息。记录各类RNA分子在特定时间和条件下的表达变化，可以识别与疾病状态相关的基因表达模式，并开发用于疾病诊断和预后的标志物[10]。通过构建共表达网络和调控网络，可以揭示基因间的相互作用或转录因子与靶基因之间的复杂调控关系，揭示疾病的分子调控机制[11]。基于转录组学数据，还可以实现对不同类型癌症的精确分类，指导个性化治疗。以肝细胞癌为例，基于转录组分析揭示了3种亚型，其临床结果、突变情况、免疫微环境和失调途径显著多样化[12]。

1.3 蛋白质组学

蛋白质组学相比转录组学，能够更全面、精确地反映与环境变化或疾病进展直接相关的动态信息，在分子水平上揭示蛋白质调控机制与表型之间的复杂关系具有重要价值[13]。当前获取蛋白质组学数据的方法包括高通量质谱、蛋白质芯片和反相蛋白质微阵列等。蛋白质组学研究极大地推动了对健

康和疾病分子过程的理解，尤其是人类蛋白质组计划发布了90.4%的完整高质量蛋白质组蓝图，为疾病的理解、诊断和治疗提供了宝贵资源[14]。通过蛋白质组学分析，可以识别与疾病相关的特异性蛋白质或蛋白质修饰，揭示疾病的分子机制。特异性蛋白质作为生物标志物可应用于疾病的早期诊断、治疗反应监测及疾病进展评估中。通过实时监测体内特定蛋白质浓度的变化，可以追踪患者疾病进展、识别疾病复发的迹象，评估患者对治疗的反应，甚至预测治疗后的预后，从而提升个体化治疗的精准性和有效性[15]。此外，蛋白质的翻译后修饰(如磷酸化、乙酰化、甲基化等)对细胞内信号转导、蛋白质转运和酶活性具有重要影响，对修饰位点的分析有助于深入理解蛋白质功能及疾病机制[16]。由于大多数药物靶点是蛋白质，因此蛋白质组学的进展加速了新药的发现和开发。如在慢性肾功能不全的研究中，发现了65种蛋白质风险位点，其中14种蛋白质可以作为药物靶点[17]。

1.4　表观遗传组学

表观遗传组学主要研究如何通过可遗传的修饰影响基因表达和细胞过程。常见的表观遗传组学数据包括DNA甲基化数据和组蛋白修饰数据。表观遗传过程的失调被认为是人类疾病发生和发展的关键因素，许多研究表明表观遗传修饰与多种疾病有关，如在炎症性肠病的发病机制中，免疫活性基因的甲基化与基因表达之间存在显著相关性[18]。表观组学在精准医学中的应用已渗透到多个领域，不仅揭示了基因表达调控的复杂性，还为个性化医疗、靶向治疗、疾病预测和预防提供了新的研究方向和治疗策略。癌症中常出现DNA甲基化异常，组蛋白修饰的变化也与多种肿瘤的进展密切相关。表观遗传组学数据的分析可帮助识别潜在生物标志物，用于疾病的诊断和治疗[5]。表观遗传学的应用还促进了基因组编辑工具的发展，如可通过调节表观遗传标记来实现基因沉默，从而提高基因编辑的准确性和持久性[19]。

1.5　代谢组学

代谢组学通过非靶向和靶向的研究方式，对生物体内所有代谢物或特定类别代谢物进行定量分析，以反映细胞生理的动态变化。常见的代谢组学研究方法包括核磁共振法和质谱法等。当代谢物水平或比例异常时，可能成为疾病的诱因。相比其他组学，代谢组学更接近表型，反映了基因、转录和蛋白质调控的最终功能输出。代谢组学与其他组学之间存在着紧密的关联，如研究发现纤维肌痛患者的肠道细菌与氨基酸水平之间存在相关性[20]；在肥厚型心肌病中，肌酸激酶和线粒体蛋白的表达与酰基肉毒碱和乙酰辅酶A的表达一致[21]。因此，代谢组学不仅能独立分析代谢物的变化，还通过与其他组学结合，深化对复杂生物过程的理解，推动精准医学的应用。通过识别特异性代谢物、监测代谢状态变化，代谢组学分析为疾病诊断、治疗、预后预测和个性化健康管理提供了重要的工具和策略。

1.6　肠道微生物组学

肠道微生物作为一种重要的环境因素，通过与宿主的相互作用影响宿主表型，如影响宿主的新陈代谢和免疫系统的调节等[22]。尽管微生物表型受宿主遗传因素的影响，但其组成在个体间表现出高度的独特性和多样性，并且更容易受到环境因素，如饮食习惯、生活方式和药物使用等的影响[23]。这种多样性和动态性使得微生物组在疾病发生和临床变异中扮演着重要的角色[8, 24]。借助16S rRNA测序、宏基因组测序和宏转录组测序等先进技术，研究者能够获取丰富的微生物组数据，从而解析微生物群的组成、基因功能和表达模式，揭示其如何影响特定疾病的发生和发展[25]。越来越多的研究表明，许多疾病都与肠道微生物组的改变密切相关[20, 26]。许多微生物可作为疾病进程的生物标志物，为疾病的预防、诊断和治疗提供新的思路和方法[25]。如拟杆菌与厚壁菌比例失衡，可作为预测肥胖的潜在风险标志物，而其他菌群的特定变化，可辅助疾病的早期发现和分类诊断等。结合微生物标志物和临床指标，可提高癌症筛查的灵敏度和特异性。粪便微生物移植的应用已被用于治疗多种疾病，显示出极高的治愈率[22]。

2　生物大数据面临的困难与挑战

当前，生物大数据的生成速率与规模已达到前所未有的高度，年均产出量以PB为单位，并呈现指数式增长[27]。面对这一态势，众多数据库和知识库如雨后春笋相继涌现。其中，以美国国家生物技术信息中心(National Center for Biotechnology Information，NCBI)、欧洲生物信息研究所(EMBL-European Bioinformatics Institute，EMBL-EBI)、日本的国际核酸序列数据库联盟(International Nucleotide Sequence Database Collaboration，INSDC)和中国的国家生物信息中心

(China National Center for Bioinformation，CNCB)为首的数据中心，形成了全球性的数据枢纽。此外，如千人基因组计划、ENCODE(Encyclopedia of DNA Elements)、TCGA(The Cancer Genome Atlas)及GTEx(Genotype-Tissue Expression)等大型研究项目也极大地丰富了生物大数据资源的多样性(表1)。这些数据资源为生物医学研究、临床诊断和药物开发提供了强大的数据资源与技术支撑。

海量生物数据涌现的同时带来了诸多挑战。首先，不同平台生成的数据在格式、标准和质量上存在显著差异，使得异构数据的高效处理和有效整合成为亟待解决的难题。其次，如何从生物大数据中提取高质量、有价值的知识是另一大挑战。传统研究多集中于单一组学，限制了对生命系统复杂机制的解析。随着研究逐渐向多模态数据整合分析方向发展，结合临床数据、组学数据和病理影像数据的综合分析方法也应运而生[28-29]。这种多模态整合策略能够更全面地理解疾病机制，但同样面临着数据异构性和噪声干扰的问题。同时，处理多模态数据的计算复杂度远超单组学分析，对计算资源和算法的要求显著提高。从多模态、多来源的数据中构建统一的表示和映射关系，在尽量保留关键信息的基础上，进行不同类型数据的有效整合分析，同时对潜在关联进行推断，最终对结果进行可视化和解释，这一过程的每个环节都面临着独特且复杂的挑战。此外，生物大数据通常具有高维特性，样本量相对较少时会进一步增加数据整合和分析的复杂性，对样本的数量、质量以及算法的设计与优化提出了更高要求。

3 生物大数据未来发展趋势

生物大数据的飞速发展为生物医学和精准医学带来了前所未有的机遇，但同时也伴随着诸多挑战。随着单细胞组学、时空组学等新兴技术的崛起，以及人工智能与生物信息学的深度融合，未来生物大数据将进一步推动疾病机制的解析、精准医疗的落地和全球健康研究的创新发展。

3.1 新兴技术驱动精细研究

单细胞组学、时空组学等新兴技术的发展对生物大数据的获取、处理和应用产生了深远的影响。传统组学研究主要基于大量细胞的平均数据，这在一定程度上忽略了个体细胞间的差异，难以解析细胞异质性。而单细胞测序技术突破了这一限制，在解析细胞异质性、描绘细胞类型和状态的多样性、追踪细胞发育和分化、深入研究疾病的细胞和分子机制等方面展现了巨大的潜力[4, 30]。如通过单细胞技术，研究人员可以识别肿瘤内部的异质性，发现罕见突变和亚群体，为癌症的精准治疗提供了重要依据[30]。空间转录组学则进一步突破了传统转录组学在空间信息上的缺失，能够精确地保留并解析基因表达的空间位置信息。该技术不仅为组织或器官的基因表达提供了详细的空间

表1 生物大数据资源平台及特点

Tab. 1 Biological big data resource platforms and their characteristics

全称	简称	主要特点及网址
美国国家生物技术信息中心 National Center for Biotechnology Information	NCBI	提供全面的生物学数据库(包括GenBank、GEO等)、文献资源库(PubMed)和工具支持(BLAST)等，用于基因组学、蛋白质组学、进化生物学等研究。www.ncbi.nlm.nih.gov
欧洲生物信息学研究所 European Bioinformatics Institute	EBI	维护的数据库包括ENA(欧洲核酸档案馆)、UniProt(蛋白质数据库)、ArrayExpress(基因表达数据)、Ensembl(基因组浏览器)等。www.ebi.ac.uk
日本DNA数据库 DNA Data Bank of Japan	DDBJ	提供如DRA、GEA和JGA等数据资源。www.ddbj.nig.ac.jp
国家生物信息中心 China National Center for Bioinformation	CNCB	中国的国家生物信息中心，负责管理和共享国家基因组数据。其核心平台NGDC包括Genome Warehouse(基因组数据库)、Gene Expression Nebula(基因表达数据库)、Variation Database(变异数据库)等。www.big.ac.cn
千人基因组计划 1 000 Genomes Project	1 000 G	建立了一个广泛的人类基因组变异图谱。www.internationalgenome.org
DNA元件百科全书项目 Encyclopedia of DNA Elements	ENCODE	旨在识别人类基因组中的所有功能元件，包括编码区和非编码区，提供全基因组范围内的表观遗传、转录组和染色质相互作用等数据。www.encodeproject.org
癌症基因组图谱项目 The Cancer Genome Atlas	TCGA	为癌症类型构建全面的基因组图谱，提供了各种癌症的基因组测序、表达、拷贝数变异、甲基化等多组学数据。portal.gdc.cancer.gov
基因型-组织表达项目 Genotype-Tissue Expression	GTEx	包含不同人类组织中的基因表达数据，允许研究基因型与表型之间的关联。www.gtexportal.org

图谱，还为揭示细胞在组织微环境中的相互作用提供了新视角，有助于我们深入理解微环境对细胞行为和功能所施加的影响。如在肿瘤和其他疾病研究中，空间转录组学技术可以揭示不同肿瘤微环境中的空间格局，以及肿瘤细胞与免疫细胞、基质细胞之间的相互作用[31]。

此外，高通量染色质构象捕获技术(high-throughput chromosome conformation capture，Hi-C)及其衍生技术突破了传统的局限性，能够揭示基因调控的空间机制、提供三维基因组信息、揭示基因调控的空间机制、检测结构变异等，在疾病机制和基因调控研究中具有重要应用价值[32-33]。

3.2 数据标准化助力整合互通

随着生物大数据的持续增长，多源异构性数据的整合已经成为一大挑战。在未来，进一步加强数据标准化将成为必然趋势，这不仅有助于提升全球范围内的数据互操作性，还能推动更加高效的分析与跨领域协作。数据标准化确保了不同来源数据的一致性和可比性，从而简化跨平台、跨机构的整合分析。这不仅提高了数据分析的效率，还增强了数据的质量和可靠性。例如，全球范围内的遗传数据共享平台(如GA4GH)采用统一的基因组数据格式和标准，推动了人类基因组研究的国际合作[34]。这类标准化平台使得各国研究者能够无缝协作，快速对比不同人群的遗传背景，揭示遗传变异与疾病之间的关联。数据标准化的另一个趋势是实现多模态数据的统一整合。随着多模态数据在生物医学研究中的重要性不断提升，构建不同组学数据间的标准化整合框架将成为研究的焦点。

3.3 多模态融合促进机制解析

生命系统的复杂性制约了单一组学技术在疾病研究领域的全面应用。基因组变异、转录组表达、蛋白质组互作等多维信息与疾病状态密切相关。此外，疾病的发生发展往往还交织着环境因素与个体遗传背景的双重影响，这进一步增加了疾病机制的复杂性。因此跨层次、多维度的数据整合分析日益重要。将同一个体的多组学数据进行融合关联，能够更全面、深入地挖掘遗传信息传递过程的深层次特征，有助于解析复杂生命活动的机制，揭示疾病的潜在致病变化及其背后的因果关系，挖掘更关键的功能并提供更多的临床见解[3-4]。

借助多组学整合分析的方法，研究人员已能够深入剖析多种复杂疾病，如II型糖尿病、骨关节炎、阿尔茨海默病、系统性红斑狼疮等[35-36]。此外，影像数据与组学、临床数据的整合，能够更好地阐明疾病在分子、组织、器官层面的表现。如通过整合MRI影像和基因组数据，识别出与胶质母细胞瘤预后相关的影像特征，进而改善肿瘤的精准分型和治疗方案的制定[28]。融合影像组学与基因组数据特征，不仅可提供乳腺癌的诊断价值，还可以提供图像与生物学特征之间的预测或预后关联，从而推动了乳腺癌精准医疗的发展[29]。

3.4 大规模队列支撑精准医学研究

大规模队列研究涵盖数千至数百万样本，包含不同性别、年龄、种族和地理区域的样本，带来了强大的统计能力，使得研究者能够可靠地检测复杂的生物学特征、微小的基因效应及环境因素的影响，提升了结果的普适性和代表性。此外，大队列研究为精确建模和深入分析提供了坚实的数据基础。通过大规模数据的积累，研究者能够识别更广泛的疾病模式和健康状态的变化。这些研究通常通过长期随访，动态收集个体的健康数据，允许研究者跟踪疾病的发生、进展及预后。通过持续获取和分析临床、组学、生活方式等多维数据，可以深入分析从健康状态到疾病状态的转变过程，进而识别早期生物标志物和风险因素。

目前，多个国家已经启动了大规模的生物数据队列研究项目，如英国的UK Biobank、美国的All of Us Research Program和中国的国家精准医学计划等，这些项目通过收集大规模的基因组、表观遗传组及临床表型数据，为复杂疾病研究和个性化治疗探索提供了宝贵的生物大数据资源。

3.5 人工智能拓展数据应用

人工智能技术的飞速发展，催生并持续涌现出一大批新算法新模型，为挖掘生物大数据、解读疾病发生进展规律提供了有效工具。这些工具在助力深入解析数据中蕴含的多种生命活动特征和多种分子间作用机制发挥了重要作用。人工智能与生物大数据的深度融合，在疾病诊断、药物开发、个性化治疗等方面取得了显著进展，如图神经网络在药物靶标预测中展现了强大的潜力，通过将化学分子结构作为图进行建模，准确捕捉分子间的相互作用，已被用于预测药物的药理活性和不良反应[37]。在基因编辑方面，基于机器学习的算法可精确预测CRISPR-Cas9的编辑效率和脱靶风险，为基因疗法的个性化设计提供支持[38]。此外，人工智能驱动的蛋白质结构预测，如AlphaFold的成功，极大加速了药物靶标和治疗机制的发现，已经成为结构

生物学和药物开发中的关键工具[39]。

未来，多维度生物大数据的积累和自然语言处理领域的发展，进一步为利用人工智能技术来解释生物语言铺平了道路，生物大模型也将迎来快速发展机遇。如阿里云发布了首个联合DNA、RNA、蛋白质的生物大模型——LucaOne[40]。该大模型有效地学习了分子生物学的中心法则，在基因到蛋白质的映射以及探索基本的生物过程等方面有较高的应用潜力。此外，Evo的出现突破了传统模型单模态功能的局限性，具备了更强大的能力，其不仅能够执行从分子水平到全基因组规模的预测任务和生成设计，还能处理编码序列与非编码序列之间的复杂共进化关系[41]。Evo还支持设计包括功能性CRISPR-Cas等复杂生物体系，极大地拓展了其在基因组编辑和生物合成等领域的应用潜力，有望成为重要的科学工具。

作者贡献　王霈虹：撰写论文；何昆仑、石金龙：审阅和修改。

利益冲突　所有作者声明无利益冲突。

参考文献

1　Collins FS，Green ED，Guttmacher AE，et al. A vision for the future of genomics research［J］. Nature，2003，422（6934）：835-847.
2　Dai XF，Shen L. Advances and trends in omics technology development［J］. Front Med（Lausanne），2022，9：911861.
3　Mukherjee A，Abraham S，Singh A，et al. From data to cure：a comprehensive exploration of multi-omics data analysis for targeted therapies［J/OL］. https://doi. org/10.1007/s12033-024-01133-6.
4　Chen CY，Wang J，Pan DH，et al. Applications of multi-omics analysis in human diseases［J］. MedComm（2020），2023，4（4）：e315.
5　Liu H，Ma HM，Li Y，et al. Advances in epigenetic modifications and cervical cancer research［J］. Biochim Biophys Acta Rev Cancer，2023，1878（3）：188894.
6　Tam V，Patel N，Turcotte M，et al. Benefits and limitations of genome-wide association studies［J］. Nat Rev Genet，2019，20（8）：467-484.
7　González-Del Pozo M，Fernández-Suárez E，Bravo-Gil N，et al. A comprehensive WGS-based pipeline for the identification of new candidate genes in inherited retinal dystrophies［J］. NPJ Genom Med，2022，7（1）：17.
8　Liu XM，Tong X，Zou YQ，et al. Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome［J］. Nat Genet，2022，54（1）：52-61.
9　Lin L，Zhang XR，Fan JL，et al. Natural variation in BnaA07. MKK9 confers resistance to Sclerotinia stem rot in oilseed rape［J］. Nat Commun，2024，15（1）：5059.
10　Mehta SL，Dempsey RJ，Vemuganti R. Role of circular RNAs in brain development and CNS diseases［J］. Prog Neurobiol，2020，186：101746.
11　Das S，Rai SN. Predicting the effect of miRNA on gene regulation to foster translational multi-omics research-a review on the role of super-enhancers［J］. Noncoding RNA，2024，10（4）：45.
12　Hu DX，Zhang ZT，Liu XY，et al. Generalizable transcriptome-based tumor malignant level evaluation and molecular subtyping towards precision oncology［J］. J Transl Med，2024，22（1）：512.
13　Ohlsson M，Hellmark T，Bengtsson AA，et al. Proteomic data analysis for differential profiling of the autoimmune diseases SLE，RA，SS，and ANCA-associated vasculitis［J］. J Proteome Res，2021，20（2）：1252-1260.
14　Adhikari S，Nice EC，Deutsch EW，et al. A high-stringency blueprint of the human proteome［J］. Nat Commun，2020，11（1）：5301.
15　Breuer EY，Murph MM. The role of proteomics in the diagnosis and treatment of women’s cancers：current trends in technology and future opportunities［J/OL］. https://doi. org/10.1155/2011/373584.
16　Bhardwaj S，Bulluss M，D’Aubeterre A，et al. Integrating the analysis of human biopsies using post-translational modifications proteomics［J］. Protein Sci，2024，33（4）：e4979.
17　Dubin RF，Deo R，Ren Y，et al. Proteomics of CKD progression in the chronic renal insufficiency cohort［J］. Nat Commun，2023，14（1）：6340.
18　Kalla R，Adams AT，Nowak JK，et al. Analysis of systemic epigenetic alterations in inflammatory bowel disease：defining geographical，genetic and immune-inflammatory influences on the circulating methylome［J］. J Crohns Colitis，2023，17（2）：170-184.
19　Nuñez JK，Chen J，Pommier GC，et al. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing［J］. Cell，2021，184（9）：2503-2519.
20　Clos-Garcia M，Andrés-Marin N，Fernández-Eulate G，et al. Gut microbiome and serum metabolome analyses identify molecular biomarkers and altered glutamate metabolism in fibromyalgia［J］. EBioMedicine，2019，46：499-511.
21　Previs MJ，O’Leary TS，Morley MP，et al. Defects in the proteome and metabolome in human hypertrophic cardiomyopathy［J］. Circ Heart Fail，2022，15（6）：e009521.
22　Loh JS，Mak WQ，Tan LKS，et al. Microbiota-gut-brain axis and its therapeutic applications in neurodegenerative diseases［J］. Signal Transduct Target Ther，2024，9（1）：37.
23　Goodrich JK，Waters JL，Poole AC，et al. Human genetics shape the gut microbiome［J］. Cell，2014，159（4）：789-799.
24　Marchesi JR，Adams DH，Fava F，et al. The gut microbiota and host health：a new clinical frontier［J］. Gut，2016，65（2）：330-339.
25　Cammarota G，Ianiro G，Ahern A，et al. Gut microbiome，big data and machine learning to promote precision medicine for cancer［J］. Nat Rev Gastroenterol Hepatol，2020，17（10）：635-648.
26　Yue B，Luo XP，Yu ZL，et al. Inflammatory bowel disease：a potential result from the collusion between gut microbiota and mucosal immune system［J］. Microorganisms，2019，7（10）：440.
27　Zou D，Ma LN，Yu J，et al. Biological databases for human research［J］. Genomics Proteomics Bioinformatics，2015，13（1）：55-63.
28　Zinn PO，Mahajan B，Sathyan P，et al. Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme［J］. PLoS One，2011，6（10）：e25451.

（下转第36页）