·综述·

# 高通量测序仪与生物信息学分析平台在bulk测序中的集成与应用

薛柔[a,b,c,d] 晏菲[e] 印明柱[a,b,c,d] 李鑫[a,b,c,d*]

（重庆大学附属三峡医院 a. 临床研究中心；b. 医学病理中心；c. 肿瘤早期诊治中心；d. 转化医学研究中心；e. 重庆市老年疾病临床医学研究中心，重庆 404000）

【摘要】 高通量测序技术即下一代测序(next-generation sequencing, NGS)技术的快速发展推动了基因组学、转录组学的进步，但其数据规模的指数级增长对数据分析能力提出了严峻挑战，促使生物信息学分析平台成为高通量测序数据解析的核心工具。本文综述了bulk测序中高通量测序仪与生物信息学分析平台的集成与应用，重点讨论了数据处理流程整合、算法优化及自动化处理在提升研究效率和准确性中的重要作用。首先介绍了不同高通量测序技术的基本原理、仪器特征及其在bulk测序中的适用场景，分析了NGS数据预处理与质量控制、序列比对和变异检测算法的演进，以及自动化分析流程与测序系统之间的协同关系。在此基础上，分别从科研场景中的集成模式、临床检测中的流程化集成模式以及智能化与云端平台驱动的新型集成模式3个方面，结合人群基因组研究、肿瘤精准医疗及病原体快速检测等实际案例，探讨了高通量测序仪与生物信息学分析平台协同应用在提升数据处理效率、分析准确性及结果转化能力方面的价值。最后，展望了人工智能算法、云计算平台、实时分析及多模态数据整合在bulk测序生物信息学分析中的发展前景。

【关键词】 高通量测序仪；生物信息学分析平台；数据整合；精准医学；自动化处理

【中图分类号】 R730.43 【文献标识码】 A 【DOI】 10.11876/mimt202604018

# Integration and application of high-throughput sequencers and bioinformatics analysis platforms in bulk sequencing

*Xue Rou[a,b,c,d], Yan Fei[e], Yin Mingzhu[a,b,c,d], Li Xin[a,b,c,d*]*

*[a. Clinical Research Center (CRC); b. Medical Pathology Center (MPC); c. Cancer Early Detection and Treatment Center (CEDTC); d. Translational Medicine Research Center (TMRC); e. Chongqing Municipality Clinical Research Center for Geriatric Diseases, Chongqing University Three Gorges Hospital, Chongqing 404000, China]*

【Abstract】 The rapid advancement of high-throughput sequencing technology, also known as next-generation sequencing (NGS), has significantly propelled research in genomics and transcriptomics. However, the exponential growth of data poses substantial challenges to data analysis capabilities, effectively positioning bioinformatics analysis platforms as core tools for the interpretation of high-throughput sequencing datasets. This review explores the integration and application of high-throughput sequencers and bioinformatics analysis platforms in bulk sequencing, focusing on the critical roles of data processing workflow integration, algorithm optimization, and automated processing in enhancing research efficiency and accuracy. Firstly, it introduces the fundamental principles and instrument characteristics of various high-throughput sequencing technologies and their applicable scenarios in bulk sequencing, and analyzes NGS data preprocessing and quality control, the evolution of sequence alignment and

收稿日期：2026-03-02

基金项目：重庆市科卫联合医学科研项目青年（医师）科学家项目（2025QNKX004）；新重庆青年创新人才项目（CSTB2025YITP-QCRCX0005）

第一作者：薛柔，女，硕士研究生，助理研究员，研究方向：分子病理；晏菲，男，博士研究生，副研究员，研究方向：生物力学与康复工程

通信作者：李鑫，女，博士研究生，研究员，研究方向：复杂疾病微环境调控机制研究

variant detection algorithms, and the coordination between automated analysis workflows and sequencing systems. On this basis, the integration modes are discussed from three perspectives: integration in research settings, workflow–based integration in clinical testing, and novel integration driven by intelligent and cloud platforms. By incorporating practical examples from population genomics research, precision oncology, and rapid pathogen detection, it explores the value of integrating high–throughput sequencers with bioinformatics analysis platforms in improving data processing efficiency, analytical accuracy, and the translational potential of results. Finally, the future prospects of artificial intelligence algorithms, cloud computing platforms, real–time analysis, and multimodal data integration in bioinformatics analysis for bulk sequencing are discussed.

**【Key words】** High–throughput sequencer; Bioinformatics analysis platform; Data integration; Precision medicine; Automated processing

人类基因组计划的完成标志着生命科学进入后基因组时代，研究重点逐渐由“基因序列解析”转向基因功能研究，这对高通量、高精度数据获取技术提出了更高要求[1]。在此背景下，高通量测序技术，亦称下一代测序（next–generation sequencing, NGS）技术，逐步发展成为现代生物医学研究的重要技术支撑[2]。按照测序对象和样本分析层级，NGS可分为bulk测序、单细胞测序等不同类型。其中，bulk测序主要以组织或混合细胞群体为研究对象，获得样本整体或群体平均水平的分子特征，是目前应用较为广泛的测序模式。与传统Sanger测序相比，NGS依托大规模并行测序策略显著提升了数据产出效率，降低了全基因组测序成本，并推动了bulk测序在大样本基因组学研究中的广泛应用。测序能力的提升不仅加快了基因组图谱的构建进程，也使基于大样本的数据分析逐渐成为生命科学研究的重要方式，是近年来分子生物学领域最具代表性的技术突破之一[3]。

随着测序技术的持续发展，NGS平台形成多技术路线并行的格局。以Illumina为代表的第二代测序技术凭借较高的准确度和通量，成为当前科研和临床检测的主要平台；以PacBio和Oxford Nanopore Technologies（ONT）为代表的第三代测序技术依托长读长特征，在结构变异检测、基因组组装及复杂重复序列解析等方面表现出良好的应用潜力[3]。在bulk测序应用中，测序仪性能的不断提升，使研究者能在更高分辨率下开展基因组研究，拓展研究深度。然而，测序能力的提升也带来了数据规模的快速增长。研究[4]表明，基因组数据的增长速度已超过多个传统数据密集型领域，预计将在未来全球数据中占据显著份额。大规模数据在提供研究资源的同时，也对计算能力、存储架构及数据管理提出了更高要求，逐渐成为后基因组研究中需要重点关注的问题[5]。

在此背景下，生物信息学逐渐发展为生命科学研究的重要组成部分。生物信息学分析平台通常由计算与存储资源、分析软件、数据库及流程管理工具构成，主要承担测序数据的质量控制、序列比对、变异检测、定量分析和功能注释等任务。其通过整合标准化分析流程、统计学方法与高性能计算技术，实现了从原始测序数据到生物学解释的系统转化[6]。多组学整合分析的发展使研究者能够从多个分子层面理解疾病机制，但也对算法性能和计算资源提出了更高要求[7]。NGS与生物信息学的结合为精准医学的发展提供了重要支撑。精准医学强调基于个体遗传特征制定诊疗策略，NGS能识别致病突变和分子分型信息，从而为个体化治疗提供参考。在肿瘤研究中，大规模基因组测序已成为靶向治疗与免疫治疗研究的重要技术基础[8]。目前，面向bulk测序的高通量测序仪与生物信息学分析平台的高效协同仍面临多重挑战，如数据标准化程度不足、分析流程缺乏统一规范、计算资源需求持续增加等。因此，系统梳理高通量测序仪与生物信息学分析平台在bulk测序中的集成模式及其应用进展，对促进技术体系优化具有重要的现实意义。

## 1 主流NGS平台及仪器特征

NGS技术的快速发展伴随着测序仪器系统的持续优化，近年来，不同测序平台在信号检测方式、阵列构建策略及系统集成度等方面形成了差异化技

术路线，构建起多层次的测序设备体系[2]。在短读长测序领域，Illumina和华大智造（MGI）平台已形成较为成熟的仪器产品体系，并在科研和临床检测中广泛应用。Illumina系列测序仪基于边合成边测序（sequencing by synthesis, SBS）原理，通过可逆终止子控制DNA链的逐碱基延伸，并利用高分辨率光学成像系统捕获荧光信号以完成碱基判读[9]。同时，其流动槽采用高密度阵列设计，使大量DNA分子在固相表面同步扩增并被连续读取，从而实现大规模并行测序。这种将测序化学与精密光学系统相结合的工程设计是高通量数据获取的基础。MGI平台采用基于DNA纳米球的阵列构建方式，结合联合探针锚定聚合（combinatorial probe–anchor synthesis, cPAS）技术进行碱基识别。该策略通过滚环扩增形成高度均一的DNA纳米球，可降低扩增偏倚，提升阵列密度，从而支持高通量测序[10]。有研究[11]对华大智造BGISEQ–500平台和Illumina HiSeq 2500平台进行了比较，结果显示二者在数据生成方面大体相似，且在核DNA水平、序列GC含量、双链DNA损伤率等多个测序参数上无显著差异，表明BGISEQ–500在基因组测序中具备较为稳定的表现，并可能在古基因组学等特殊应用场景中成为一种有价值的替代平台。

系统架构方面，短读长测序仪普遍呈现出高度集成化的发展趋势，通过将模板制备、测序反应、信号采集及初步数据处理等关键环节整合于自动化流程中，一定程度上减少了人工操作带来的实验偏差，提升了运行稳定性[3]。同时，成像速度和数据输出能力的持续优化使单次运行的数据产出显著提高，为大规模队列研究提供了设备支撑。得益于较高的测序准确度和成熟的工程体系，短读长测序仪已成为群体基因组研究、转录组分析及临床检测的重要技术基础。然而，受限于50 ~ 300 bp的读长范围，此类平台在解析复杂基因组结构方面存在一定局限，对大型结构变异、高度重复序列及染色体重排的识别能力相对不足[3]。

第三代测序平台以单分子测序为主要特征，在读长方面实现了显著提升。PacBio开发的单分子实时测序（single–molecule real–time sequencing, SMRT）技术通过实时监测DNA聚合酶催化反应记录碱基掺入信号，可获得远超短读长测序平台的连续序列信息[12]。较长的读取长度使SMRT在全长转录本测序、复杂结构变异检测及高质量基因组组装等研究中展现出重要价值。ONT推出的纳米孔测序技术通过监测DNA分子穿过生物纳米孔时引起的电流变化进行碱基识别，可在无需扩增的情况下直接读取超长DNA片段[13]。此外，纳米孔测序设备在体积与部署方式上具有一定灵活性，为现场快速检测和病原体监测提供了新的技术路径。

总体来看，不同NGS平台在仪器架构和信号检测策略上形成了互补格局，见图1。从技术性能看，短读长测序平台通常具有单碱基准确度高、通量大、成本相对可控、分析流程成熟等优势，更适用于bulk RNA–seq、全外显子组测序、靶向panel检测及大规模队列分析，但其读长较短，对复杂重复序列、大型结构变异、染色体重排及长片段单倍型解析能力有限。长读长测序平台能够提供更连续的序列信息，在结构变异识别、复杂基因组区域解析、全长转录本分析及基因组组装中具有优势，但面临数据成本、错误模型、测序通量及计算处理复杂度等方面的限制。研究[14]表明，将不同读长数据进行联合分析可进一步提升基因组解析能力，该策略正逐渐成为复杂基因组研究的重要方向。因此，平台选择不宜仅依据通量或读长单一指标判断，而应结合研究目的、样本规模、目标变异类型、测序深度需求、成本预算及下游分析能力综合评估。随着国产测序设备的持续发展，部分新兴平台逐步进入科研应用场景，在丰富技术选择的同时，也为生物医学研究提供了更多可能。

## 2 NGS数据特征与生物信息学分析需求

### 2.1 数据预处理与质量控制体系

NGS数据的质量直接影响下游分析结果的可靠性，因此系统性的质量控制和预处理是生物信息学分析流程中的关键环节。与传统低通量测序相比，NGS在提升数据产出的同时引入了多来源技术误差，包括碱基识别错误、聚合酶链式反应（polymerase chain reaction, PCR）扩增偏倚、接头污染及测序深度不均等，可能干扰变异检测和定量分析结果[15]。

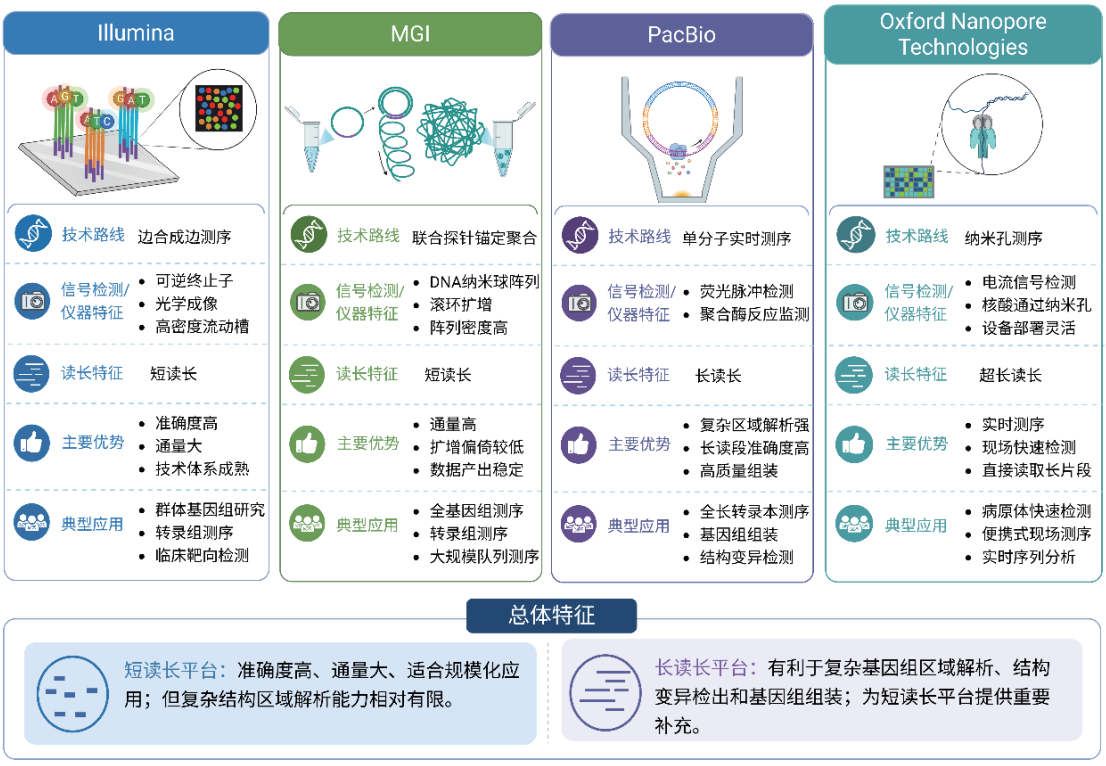



**图 1** 主流 NGS 平台及仪器特征

注：NGS 为下一代测序。

短读长测序通常依赖 Phred 质量评分评估碱基判读准确性，该指标以对数形式反映碱基识别错误概率，被广泛用于过滤低可信度序列。研究[16]表明，在下游分析前去除低质量读段能够显著降低假阳性变异率，进而提高基因组分析的准确性。接头序列残留是影响比对效率的重要因素之一，序列修剪可有效减少错误匹配并提升比对质量。除了单碱基质量问题，测序过程中的系统性偏倚同样值得关注。例如，PCR 扩增可能导致序列重复比例升高，而不同 GC 含量［核酸序列中鸟嘌呤（guanine，G）和胞嘧啶（cytosine，C）所占的比例］区域在扩增和测序效率上的差异可能造成局部覆盖度不足，从而影响变异检测灵敏度。研究[15]指出，标记或去除重复序列及评估重复序列覆盖度分布，有助于减少技术噪声对分析结果的干扰。单分子长读长测序平台误差特征与短读长测序平台不同，早期 PacBio 测序的原始错误率相对较高，但错误多呈随机分布，可通过增加测序深度或构建共识序列提升准确度[12]。随后出现的环形一致性测序（circular consensus sequencing，CCS）策略进一步提高了单分子测序精度，可同时兼顾长度和准确性[17]。因此，针对不同测序技术建立差异化质量控制策略，已成为生物信息学分析平台设计的重要考量。

### 2.2 序列比对和变异检测的算法演进

序列比对是 NGS 数据分析的核心步骤之一，目标是将大量测序读段准确映射到参考基因组，为后续变异检测和功能解析提供基础。随着测序通量的快速提升，传统基于动态规划的比对方法已难以满足大规模数据处理的需求，兼顾计算效率和比对精度的算法设计成为生物信息学研究的重要方向。Burrows–Wheeler 变换（Burrows–Wheeler transform，BWT）算法的引入推动了短读长比对算法的发展，该策略构建的索引结构能够在压缩存储空间的同时实现快速字符串匹配，帮助研究者在可接受的计算资源范围内完成海量读段的基因组定位[18]。随后发展的比对算法在处理错配和小片段插入缺失方面不断优化，进一步提升了比对准确性，逐渐成为群体基因组研究的基础技术之一[19-20]。在完成序列定位后，变异检测算法通过整合碱基质量值、测序深度及等位基因比例等信息，对潜在突变进行概率

评估。研究[21]表明，引入贝叶斯模型能够有效区分真实变异与测序误差，降低假阳性率，从而提高检测灵敏度。随着分析框架的不断完善，基于局部重组装的策略被进一步用于复杂区域的变异识别，有助于提高插入缺失突变的检测准确性[16]。

长读长测序技术的兴起对传统算法体系提出了新的挑战。由于单分子测序读段更长且错误模式不同，短读长比对策略往往难以直接适配。为解决这一问题，容错率高且支持大跨度比对的新型算法逐渐建立，帮助研究者准确识别结构变异并解析复杂基因组区域[22-23]。研究[17,24]指出，长读长数据在重复序列跨越能力方面优势明显，为基因组结构解析提供了新的技术路径。值得关注的是，随着测序深度和样本规模持续增加，计算效率逐渐成为算法设计的重要考量因素。通过并行计算和数据结构优化，现代比对和变异检测流程能够在保证准确性的同时提升运行速度，以适应大规模基因组研究的需求。

由此可见，短读长与长读长数据在读长和误差模式上存在差异，这决定了其分析流程设计各有侧重。对于以表达定量、常见小变异识别及大样本队列比较为目标的bulk测序任务，生物信息学分析平台更关注统一质控、覆盖度评估、批次效应控制及结果可重复性；而对于结构变异、复杂重复区域及转录本结构解析等任务，则需要依赖长读段比对、错误校正及复杂变异识别算法。因此，生物信息学分析平台的构建不宜采用单一固定流程，而应根据测序平台、研究终点及目标变异类型进行差异化配置。

### 2.3 自动化分析流程与测序系统的协同

随着NGS逐步进入规模化、常态化应用阶段，单纯依赖人工干预的数据分析模式已难以满足效率和可重复性要求。在此背景下，自动化分析流程逐渐成为连接测序仪和生物信息学分析平台的技术纽带，以实现测序数据由原始信号向可解释结果的高效转化。现代测序系统通常配备基础数据处理模块，可在测序运行过程中完成碱基识别、质量评分及初步数据格式转换，为后续分析提供标准化输入。这种由仪器端生成结构化数据的模式，在一定程度上降低了后续分析流程的复杂性，提升了跨平台数据处理的一致性[3]。

在此基础上，流程管理系统的引入进一步推动了分析自动化的发展。通过预先定义分析步骤和参数配置，自动化流程能够在减少人为操作误差的同时提升结果稳定性。研究[25]指出，基于工作流的计算框架不仅有助于规范复杂数据处理过程，还可以提升研究的可重复性。此外，实验流程与计算流程之间的衔接正在不断加强。例如，在临床测序场景中，从样本建库、上机测序到变异报告生成的各环节正逐步实现流程化管理，以缩短检测周期。这种端到端的整合模式是推动测序技术走向临床应用的重要基础[26]。随着测序规模的持续扩大，自动化分析对计算资源调度提出了更高要求。部分平台通过任务并行，使大规模数据处理能够在更短的时间内完成，提升整体运行效率[5]。可以认为，自动化不仅是提升分析速度的技术手段，也是实现测序系统稳定运行的重要保障。这种从仪器数据生成到计算分析输出的流程整合，不仅提升了研究效率，也为NGS在科研、临床与公共健康领域的进一步应用奠定了技术基础，见图2。

## 3 高通量测序仪与生物信息学分析平台集成的应用模式

### 3.1 科研场景中的集成模式

高通量测序仪与生物信息学分析平台的深度融合正在重塑生命科学研究范式，使研究由传统小规模假设驱动模式向以数据为中心的系统性探索模式转变[3]。随着测序成本的持续下降和数据产出能力的提升，大规模队列研究已成为解析遗传结构和疾病机制的重要手段。在人群基因组研究中，测序仪提供稳定的数据来源，标准化生物信息学分析平台则支持跨样本数据整合与比较。例如，多国合作开展的国际千人基因组计划（1000 Genomes Project, 1KGP）通过统一数据处理流程，实现人类遗传变异的系统描绘，为疾病关联研究提供了关键资源[27]。在另一项研究[28]中，英国生物样本库完成了490 640名参与者的全基因组测序，并将测序数据与长期积累的表型及健康相关信息相结合，为大规模遗传变异识别和疾病关联分析提供了重要资源。该项目体现了高通量测序、规模化计算分析及多源

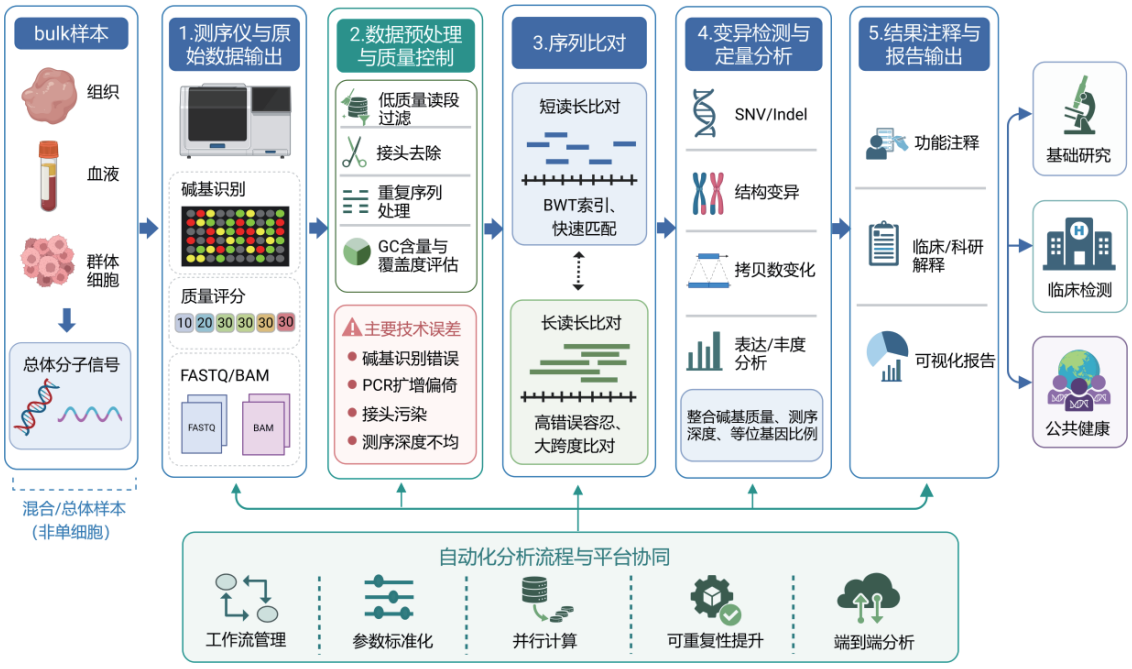



**图 2** 测序仪与生物信息学分析平台的自动化协同流程

注：FASTQ为包含核酸序列及其质量评分的文本文件格式；BAM为二进制序列比对文件格式；GC为鸟嘌呤和胞嘧啶；PCR为聚合酶链式反应；BWT为Burrows–Wheeler变换；SNV为单核苷酸变异；Indel为插入/缺失变异。

数据整合在人群基因组学研究中的协同应用。上述模式表明，只有在测序能力和计算框架同步提升的条件下，群体尺度的研究方能有效开展。

围绕疾病队列和组织样本开展的大规模bulk测序研究同样凸显了高通量测序仪和生物信息学分析平台协同的重要性。此类研究通常需要在队列水平系统刻画突变谱、表达谱及相关分子分型特征，而样本规模扩大和研究设计复杂化也使测序批次增加、样本来源与数据来源差异及技术噪声累积等问题更加突出。已有的大规模癌症基因组研究[29]表明，建立标准化、可重复的分析体系，在质控标准、分析参数、批次效应评估及结果注释等环节保持一致，有助于提升数据稳定性和生物学解释能力，推动研究者从单一样本解析转向队列尺度的系统性比较。此外，多组学联合研究正在成为生命科学的重要发展方向。通过整合基因组、转录组及表观组等多维数据，研究者能够从系统层面理解复杂生物过程。然而，多源数据在结构和尺度上的差异使得生物信息学分析平台必须具备更强的数据整合能力。研究[7]指出，跨组学整合有助于提升生物学解释深度，但对计算方法和数据管理提出了更高要求。总体来看，科研场景中的集成模式体现为测序仪负责高效数据生成，生物信息学分析平台承担数据解析和整合任务，两者协同不仅提升了研究效率，更推动了生命科学由描述性研究向系统性研究的转变。

### 3.2 临床检测中的流程化集成模式

随着NGS技术逐步投入临床应用，如何在保证检测准确性的同时提升流程稳定性已成为测序应用的关键。与科研场景相比，临床检测通常对结果的可重复性、可追溯性及报告时效性要求更高，故需要更加规范化的技术体系。在此背景下，将测序仪运行、数据分析及结果解读整合为标准化流程正逐渐成为临床测序的重要发展方向。

当前，临床bulk测序通常遵循样本处理、文库构建、上机测序、生物信息学分析及报告生成的连续流程。研究[26]表明，经过系统验证的分析管线能有效降低人为操作误差，提升检测结果的一致性。美国分子病理学协会与美国病理学家学会联合发布的《下一代测序生物信息学流程的验证标准与指南》[26]强调，对生物信息学流程进行性能评估和持续监控是保障临床测序质量的重要环节。在肿瘤精准医疗领域，流程化集成尤为重要。大规模并行测序使多基因突变同步检测成为可能，而标准

化生物信息学分析平台则支持对变异进行快速识别和临床注释，从而为靶向治疗决策提供参考。例如，英国“十万人基因组计划”癌症项目对13 880个涵盖33种癌症类型的实体瘤样本开展全基因组测序，并将基因组分析结果与真实世界治疗及结局数据进行整合[29]，该项目揭示了不同癌种中潜在可干预的分子改变，展示了高通量测序、标准化分析和临床信息整合在患者分层及精准治疗研究中的应用价值。已有共识[8]指出，NGS正在逐步改变肿瘤分子检测模式，并推动诊疗策略向更加个体化的方向发展。此外，自动化趋势正在进一步提升临床测序的运行效率。例如，将实验室信息管理系统（laboratory information management system, LIMS）与生物信息学分析平台对接，可实现样本追踪、数据管理及报告归档的流程化管理，从而降低人为差错。实践表明，信息化管理对提高检测通量和缩短报告周期具有积极意义[26]。值得关注的是，随着测序在遗传病筛查和病原体检测中的应用，临床场景对生物信息学分析平台的稳定性和计算效率提出了更高要求，特别是在时间敏感的检测任务中，快速、可靠的数据处理能力已成为影响测序应用价值的重要因素。总体而言，临床检测中的集成模式体现为测序设备提供稳定的数据来源，标准化分析流程和信息管理系统共同保障结果输出的规范性。这种由实验到计算再到报告生成的流程化整合，不仅提升了检测效率，也为bulk测序技术在精准医学中的应用奠定了技术基础。

### 3.3 智能化与云端平台驱动的新型集成模式

随着NGS技术逐渐走向规模化应用，数据处理模式正在由本地计算向云端及分布式架构转变。云计算平台的弹性资源调度能力使研究者能够根据数据规模动态配置计算和存储资源，提升大规模测序项目的处理效率。研究[30]表明，基于云环境的基因组分析框架在支持跨机构协作和数据共享方面优势明显。该趋势正在推动测序数据管理模式由单一实验室模式向网络化协同模式转变。

与此同时，人工智能逐渐被引入测序数据分析流程。深度学习模型在变异检测、表观组学信号识别及疾病相关分子特征识别等方面展现出较高的应用潜力。研究[31]指出，基于神经网络的变异检测策略能够提高在复杂区域中的识别准确率，弥补传统统计模型的不足。随着计算能力的提升和训练数据的积累，智能算法在生物信息学分析平台中的应用将不断拓展。此外，实时分析能力的提升正在改变测序应用场景。部分测序平台已具备在数据生成过程中同步进行初步分析的能力，使研究者能够在测序尚未完成时获得关键结果。这种边测序边分析的模式在病原体快速检测、应急公共卫生等场景中具有重要意义[32]。例如，一项概念验证研究[32]将短时血培养、纳米孔测序与实时生物信息学分析相结合，在人工接种临床菌株的模拟脓毒症血液样本中实现了病原体及相关耐药基因的同步识别，从培养、DNA提取、文库构建到测序和数据分析完成的总周转时间为7～9 h。该研究体现了便携式测序设备与实时分析平台在感染性疾病快速检测中的应用潜力，但其临床性能仍需通过真实患者样本进一步验证。实时计算能力的提升，标志着测序系统与生物信息学分析平台之间的耦合程度在进一步加深。

值得关注的是，多模态数据融合正在成为生物信息学分析平台扩展的重要方向。随着bulk基因组、转录组和表观组测序数据以及蛋白质组、代谢组、临床表型和病理影像等多源数据的不断积累，未来生物信息学分析平台应具备整合异构数据的能力，以支持疾病机制解析、分子分型及生物系统建模。已有研究[7]强调，多源组学数据整合有助于系统理解健康与疾病的关系，为疾病机制解析和精准医学研究提供全面的信息支撑。总体而言，未来高通量测序仪与生物信息学分析平台的集成趋势将体现为云端化、智能化与多模态化。仪器端持续提升数据生成效率，计算端通过智能算法与分布式架构优化数据解析能力，两者协同推动精准医学及系统生物学研究的发展。

从应用场景看，科研场景、临床检测及公共卫生监测对高通量测序仪与生物信息学分析平台集成的要求并不相同。科研场景更强调数据规模、分析灵活性及多队列整合能力，适合通过开放式流程支持假设探索、分子分型及机制研究；临床检测更关注流程标准化、结果可追溯性、报告时效和变异解释规范，因此对分析管线验证、质控阈值及报告

生成的一致性要求更高；公共卫生及病原体监测更加重视实时性、可部署性及数据共享能力，要求测序系统与云端或分布式生物信息学分析平台高效衔接。由此可见，高通量测序仪与生物信息学分析平台的集成并非单一模式，应根据具体应用目标在数据产出效率、分析准确性、标准化程度及结果转化价值之间进行权衡。高通量测序仪与生物信息学分析平台的集成应用框架见图 3。

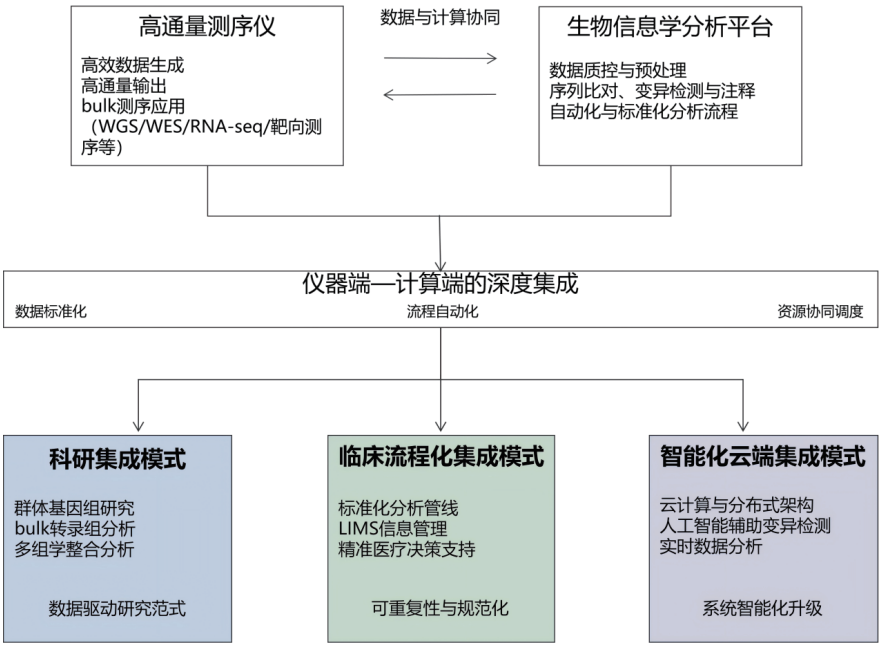



**图 3** 高通量测序仪与生物信息学分析平台的集成应用框架

注：WGS 为全基因组测序；WES 为全外显子组测序；RNA–seq 为转录组测序；LIMS 为实验室信息管理系统。

## 4 总结与展望

NGS 技术的持续进步正在深刻改变生命科学研究和临床诊疗模式。随着测序仪在通量、准确性及自动化等方面的不断提升，海量生物数据得以快速生成，生物信息学分析平台则通过算法优化和流程整合，将这些数据解析并转化为具有生物学意义的信息。两者之间形成从“数据产生”到“信息提取”的协同关系，已成为现代生物医学研究的重要技术基础。

从技术体系角度看，高通量测序仪与生物信息学分析平台的融合正在由单一工具协作向系统化集成演进。标准化分析流程的建立使测序数据处理更规范；自动化与信息化管理的引入在提升分析效率的同时，增强了结果的可重复性。在科研领域，这种集成模式推动研究由小规模探索向群体尺度和系统层面转变；在临床实践中，流程化整合为精准医学提供了稳定的技术支撑。

随着测序规模持续扩大，集成体系将面临诸多挑战。首先，数据量的快速增长对存储架构和计算能力提出了更高要求，如何在保证分析效率的同时控制计算成本，仍需重点关注；其次，不同测序平台在数据格式、误差特征等方面存在差异，一定程度上增加了跨平台数据整合的复杂性；此外，多组学研究的兴起使数据结构更加多样化，对算法的适配能力和平台的扩展性提出了新的要求。

值得关注的是，云计算与人工智能技术的发展正在为测序数据分析提供新的解决路径。弹性计算资源有助于提升大规模数据处理能力，智能算法在复杂模式识别方面展现出良好潜力。未来，随着实时分析能力的增强以及多模态数据融合技术的成熟，高通量测序仪与生物信息学分析平台之间的集成程度有望进一步加深，从而推动生命科学研究向智能化、系统化方向发展。

综上所述，NGS 技术与生物信息学分析平台的深度集成不仅提升了数据利用效率，也在一定程度上重塑了生物医学研究范式。随着技术的持续进

步，构建更加高效、稳定且具备扩展能力的集成体系将成为推动精准医学和系统生物学发展的重要方向。

**利益冲突：**作者声明没有利益冲突。

## [ 参 考 文 献 ]


[1] Collins F S, Morgan M, Patrinos A. The human genome project: lessons from large-scale biology [J]. Science, 2003, 300(5617): 286-290.

[2] Metzker M L. Sequencing technologies: the next generation [J]. Nat Rev Genet, 2010, 11: 31-46.

[3] Goodwin S, McPherson J D, McCombie W R. Coming of age: ten years of next-generation sequencing technologies [J]. Nat Rev Genet, 2016, 17: 333-351.

[4] Stephens Z D, Lee S Y, Faghri F, et al. Big data: astronomical or genomical? [J]. PLoS Biol, 2015, 13(7): e1002195.

[5] Marx V. The big challenges of big data [J]. Nature, 2013, 498: 255-260.

[6] All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program [J]. Nature, 2024, 627: 340-346.

[7] Slobodyanyuk M, Bahcheli A T, Klein Z P, et al. Directional integration and pathway enrichment analysis for multi-omics data [J]. Nat Commun, 2024, 15: 5690.

[8] Mosele M F, Westphalen C B, Stenzinger A, et al. Recommendations for the use of next-generation sequencing (NGS) for patients with advanced cancer in 2024: a report from the ESMO Precision Medicine Working Group [J]. Ann Oncol, 2024, 35(7): 588-606.

[9] Uhlen M, Quake S R. Sequential sequencing by synthesis and the next-generation sequencing revolution [J]. Trends Biotechnol, 2023, 41(12): 1565-1572.

[10] Li Q, Zhao X, Zhang W, et al. Reliable multiplex sequencing with rare index mis-assignment on DNB-based NGS platform [J]. BMC Genomics, 2019, 20: 215.

[11] Mak S S T, Gopalakrishnan S, Carøe C, et al. Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing [J]. GigaScience, 2017, 6(8): gix049.

[12] Korlach J, Bjornson K P, Chaudhuri B P, et al. Real-time DNA sequencing from single polymerase molecules [J]. Methods Enzymol, 2010, 472: 431-455.

[13] Clarke J, Wu H C, Jayasinghe L, et al. Continuous base identification for single-molecule nanopore DNA sequencing [J]. Nat Nanotechnol, 2009, 4: 265-270.

[14] Koren S, Schatz M C, Walenz B P, et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads [J]. Nat Biotechnol, 2012, 30: 693-700.

[15] Betschart R O, Riccio C, Aguilera-Garcia D, et al. Biostatistical aspects of whole genome sequencing studies: preprocessing and quality control [J]. Biom J, 2024, 66(5): e202300278.

[16] DePristo M A, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data [J]. Nat Genet, 2011, 43: 491-498.

[17] Wenger A M, Peluso P, Rowell W J, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome [J]. Nat Biotechnol, 2019, 37: 1155-1162.

[18] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. Bioinformatics, 2009, 25(14): 1754-1760.

[19] Langmead B, Salzberg S L. Fast gapped-read alignment with Bowtie 2 [J]. Nat Methods, 2012, 9: 357-359.

[20] Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment [J]. Bioinformatics, 2009, 25(15): 1966-1967.

[21] McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data [J]. Genome Res, 2010, 20(9): 1297-1303.

[22] Li H. Minimap2: pairwise alignment for nucleotide sequences [J]. Bioinformatics, 2018, 34(18): 3094-3100.

[23] Sedlazeck F J, Rescheneder P, Smolka M, et al. Accurate detection of complex structural variations using single-molecule sequencing [J]. Nat Methods, 2018, 15: 461-468.

[24] Jain M, Koren S, Miga K H, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads [J]. Nat Biotechnol, 2018, 36: 338-345.

[25] Di Tommaso P, Chatzou M, Floden E W, et al. Nextflow enables reproducible computational workflows [J]. Nat Biotechnol, 2017, 35: 316-319.

[26] Roy S, Coldren C, Karunamurthy A, et al. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: a joint recommendation of the Association for Molecular Pathology and the College of American Pathologists [J]. J Mol Diagn, 2018, 20(1): 4-27.

[27] The 1000 Genomes Project Consortium. A global reference for human genetic variation [J]. Nature, 2015, 526: 68-74.

[28] The UK Biobank Whole-Genome Sequencing Consortium. Whole-genome sequencing of 490,640 UK Biobank participants [J]. Nature, 2025, 645: 692-701.

[29] Sosinsky A, Ambrose J, Cross W, et al. Insights for precision oncology from the integration of genomic and clinical data of 13,880 tumors from the 100,000 Genomes Cancer Programme [J]. Nat Med, 2024, 30: 279-289.

[30] Deflaux N, Selvaraj M S, Condon H R, et al. Demonstrating paths for unlocking the value of cloud genomics through cross cohort analysis [J]. Nat Commun, 2023, 14: 5419.

[31] Poplin R, Chang P C, Alexander D, et al. A universal SNP and small-indel variant caller using deep neural networks [J]. Nat Biotechnol, 2018, 36: 983-987.

[32] Ali J, Johansen W, Ahmad R. Short turnaround time of seven to nine hours from sample collection until informed decision for sepsis treatment using nanopore sequencing [J]. Sci Rep, 2024, 14: 6534.