



海军军医大学学报

Academic Journal of Naval Medical University

ISSN 2097-1338, CN 31-2187/R

《海军军医大学学报》网络首发论文

题目：长读长测序技术在肿瘤领域中的应用：进展与挑战
作者：杨谨衔，陈淑桢，王红阳，文文
DOI：10.16781/j.CN31-2187/R.20220923
网络首发日期：2023-03-15
引用格式：杨谨衔，陈淑桢，王红阳，文文. 长读长测序技术在肿瘤领域中的应用：进展与挑战[J/OL]. 海军军医大学学报.
<https://doi.org/10.16781/j.CN31-2187/R.20220923>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

长读长测序技术在肿瘤领域中的应用：进展与挑战

杨谨銜¹, 陈淑桢^{2,3}, 王红阳^{2,3}, 文文^{1,2,3}✉

1. 海军军医大学第三附属医院, 上海 200438

2. 海军军医大学国家肝癌科学中心, 上海 200438

3. 海军军医大学东方肝胆外科医院国际合作生物信号转导研究室, 上海 200438

*通信作者 文文, Tel: 021-81875363, E-mail: wenwen_smmu@163.com

摘要 相对于目前普遍应用的二代测序 (NGS) 技术, 长读长测序 (LRS) 技术是一种能够连续读取数百万个碱基对的测序技术。本文回顾了 Pacific Biosciences (PacBio)、Oxford Nanopore Technologies (ONT) 和新型单分子 LRS 技术在肿瘤领域的研究应用。该技术能够更精确地检测肿瘤基因组变异, 包括结构变异、拷贝数变化、基因融合等。除基因组变异检测外, LRS 技术还为描绘转录组图谱、表观遗传修饰图谱提供了有效手段。该技术还可以有效弥补 NGS 的部分技术瓶颈, 深入揭示生物大分子在健康和疾病状态下的复杂性状, 为更好地理解肿瘤发生发展机制、制定治疗策略和开发药物提供新的理论基础。此外, 本文展望了 LRS 技术在临床应用中的潜在价值。

关键词 长读长测序; 单分子测序; 纳米孔测序; 肿瘤

DOI: 10.16781/j.cnki.31-2187/R.20220923

Application of long-read sequencing in oncology: progress and challenges

YANG Jin-xian¹, CHEN Shu-zhen^{2,3}, WANG Hong-yang^{2,3}, WEN Wen^{1,2,3}✉

1. The Third Affiliated Hospital of Naval Medical University, Shanghai 200438, China

2. National Center for Liver Cancer, Naval Medical University, Shanghai 200438, China

3. International Cooperation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai 200438, China

Abstract Compared to the commonly used second-generation sequencing (NGS), long-read sequencing (LRS) is capable of continuously reading millions of base pairs. This paper succinctly reviewed recent progress in the applications of long-read sequencing in oncology, such as Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT). These techniques facilitate the detection of genomic variations, including structural variations, copy number changes, and gene fusions. Moreover, long-read sequencing provides an effective tool for profiling transcriptome and epigenetic modification. LRS can also make up part of the technical bottleneck of the NGS and help to reveal the complexity of biological macromolecule in health and disease conditions, which will provide a new theoretical basis for better understanding the mechanisms of carcinogenesis and malignant progression, and subsequently leading to development of novel therapeutic strategies and drugs. Besides, This review outlines potential advantages of the application of LRS in clinical practice.

Key words long-read sequencing; single-molecule sequencing; nanopore sequencing; cancer

目前, 二代测序技术(next-generation sequencing, NGS)凭借其高通量、高准确性与低成本的优势在基因组测序市场占据主导地位。该技术将基因组打断为小片段, 在对每个片段进行测序后, 生成一小段 DNA 序列, 即读数 (reads)^[1]。但是, 这样的方法会导致基因组数据高度碎片化, 产生不完整甚至错误的组装。

近年来, 对读长的更高需求加速了长读长测序 (long-read sequencing, LRS) 技术的发展。LRS 技术与短读长测序技术相比, 无需将 DNA 片段化便可进行测序, 从而跨越整个重复序列并实现连续和完整的组装。随着测序通量和准确性的提升, LRS 技术可以测定上万至几兆碱基的连续序列, 技术

成熟度稳步提升。目前常使用的是以 PacBio SMRT 为代表的实时单分子测序技术和以 ONT 为代表的纳米孔测序技术，它们均能跨越基因组中高重复性的区域，以实现核酸分子更精确完整的检测，识别基因组中大片段的变异和更多的转录本。

1 LRS 技术原理

1.1 PacBio 测序技术

PacBio SMRT 测序技术采用了一种拓扑环状的 DNA 分子模板，称为 SMRTbell。该模板由两端带有单链发夹式接头的双链 DNA 插入物组成。组装好的 SMRTbell 与 DNA 聚合酶结合，并被装载到包含多达 800 万个零模波导孔（zero-mode waveguide, ZMW）的 SMRT CELL 芯片上^[2]。在测序反应过程中，聚合酶围绕 SMRTbell 进行反应，用荧光标记的脱氧核苷三磷酸合成新生链。在每次掺入后，激光将会激发荧光基团，并利用相机记录荧光信号（图 1A）。

2019 年，PacBio 开发了循环共识测序（circular consensus sequencing, CCS），该方法通过从单个模板分子的多次检测中得出共有序列，进而实现高碱基准确度^[3]，在长度约为 13 kb 的序列中碱基准确度超过 99%。但该测序过程需要约 10 μg DNA 样本，对检测样本量有限的微小肿瘤或早期肿瘤造成了一定的困难。同时，这种测序方式虽然准确性比较高，但是并未摆脱对光学系统的依赖；同时由于边合成边测序的基本技术原理，仪器的体积比较大，搬运和携带仍然比较困难。

1.2 纳米孔测序技术

2012 年，牛津纳米孔技术公司（Oxford Nanopore Technologies, ONT）公布了首个高通量测序平台 MinION。该测序平台先将测序接头与马达蛋白连接于双链 DNA 分子上，当双链 DNA 分子进入嵌在合成生物膜上的纳米孔后，马达蛋白即解开 DNA 双链，带负电荷的 DNA 在电场力与马达蛋白作用下以可控的速率通过纳米孔。DNA 分子在纳米孔中穿过时，会造成电流的扰动，通过记录电流序列依赖性变化，根据电子信号产生的差异识别不同的碱基，从而实现实时分析 DNA 链中的序列，该技术建库起始量低至 2 μg ^[4]（图 1B）。

MinION 和 PromethION 是 ONT 最常用的两种纳米孔测序仪，MinION 是一种纳米孔型便携式测序仪，最初商业化设备的读取长度约为 10 kb，单个流通池（Flowcell）的吐量约为 5 Gb^[5]。PromethION 增加了每个流通池的孔数，并支持运行多个流通池，最多可同时运行 48 个流通池，测序数据量可超过 100 Gb^[6]。基于纳米孔测序原理的测序仪特点是体积小，具有很好的便携性，适用于条件受限的特殊的应用场景。

近年来，国产纳米孔测序仪也逐步走向市场。齐碳科技有限公司推出了国内首个具有完全自主知识产权的纳米孔测序仪产品 QNome-3841，可实现生物芯片、集成电路等核心组件 100% 国产化。今是科技有限公司推出了国内首款中高通量纳米孔基因测序仪 Gseq500，单芯片测序单元数量达到 50 万个。此外，普译生物科技有限公司正在进行具有自主知识产权的纳米孔蛋白优化以及中高通量纳米孔基因测序仪的产品开发。

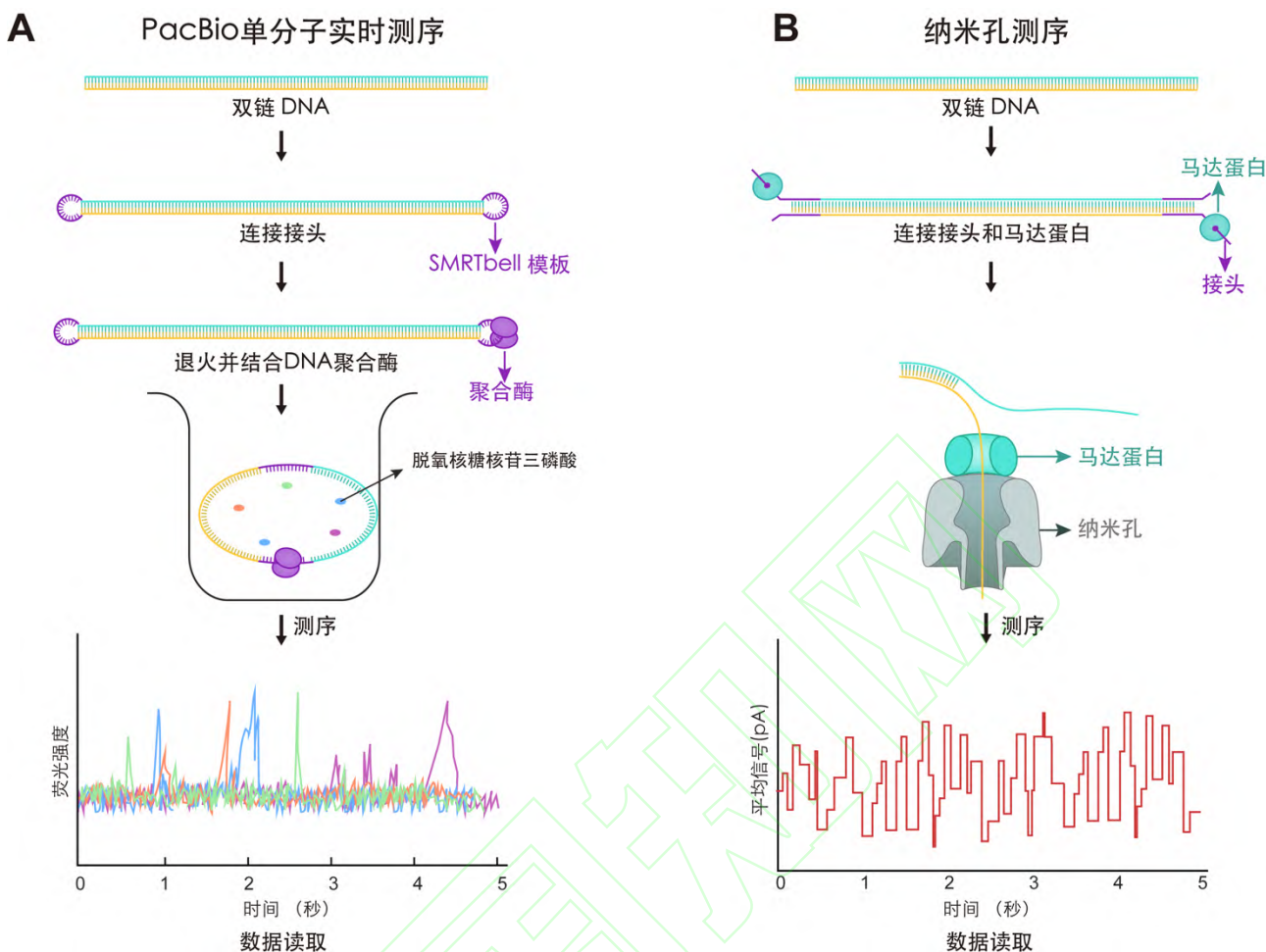


图1 PacBio与ONT测序技术工作原理

1.3 新型单分子 LRS 技术

上述两种测序技术为目前比较成熟的 LRS 技术，随着研究人员对测序原理的改进，新的单分子 LRS 技术应运而生，并且能够分析经过修饰的核酸分子。DiMeLo-seq (Directed Methylation with Long-read Sequencing, DiMeLo-seq) 结合了抗体靶向的蛋白-DNA 检测方法，将靶蛋白结合位点附近的 DNA 原位甲基化修饰后，使用 LRS 技术直接读取这些外源性甲基化标记^[7]。该方法在检测目标蛋白在单个 DNA 分子上的多种结合状态上具有独特优势，建库需要 3 μg DNA 样本，并且可以同时检测外源性抗体靶向的腺嘌呤甲基化和内源性的单分子 CpG 甲基化，从而实现在高度重复的基因区域内绘制蛋白质-DNA 相互作用图谱。

单管长片段测序 (single tube Long Fragment Read, stLFR) 是由华大智造科技股份有限公司自主研发的基于 DNBSEQ 平台的一种长片段读取技术，该方法利用 Tn5 转座酶将带有分子标签的杂交序列插入 DNA 中，以对每个长 DNA 分子进行编码，实现读取序列长度高达 10k-300k，建库起始量低至 1 ng^[8]。

TELL-seq 是一种使用 NGS 技术获取远程区域信息的方法^[9]，该方法使用转座酶来片段化 DNA 并添加条形码，进而帮助 NGS 技术重新组装 DNA 片段，让短读长数据产生超 LRS 效果，等价读长达 20 kb-200 kb。与 PacBio 和 ONT 测序方法比，碱基准确度更高，所需的 DNA 样本量更少，人类基因组 DNA 仅需 5 ng。但需要根据目标基因组进行复杂的生物信息学分析，以对获得的碎片信息进行精确注释。

通过 NGS 平台，使用测序技术对转座酶可及性染色质 (transposase-accessible chromatin) 进行单细胞测定 (即 scATAC-seq)，是一种成熟的对单个细胞内开放染色质区域进行检测的方法。然而，

由于基于 NGS 技术，该方法对检测基因组结构变异存在困难。北京大学汤富酬团队开发了一种基于 LRS 技术的转座酶可及性染色质单细胞测定方法 (scNanoATAC-seq)^[10]，可以实现同时检测单个细胞内的染色质可及性和结构变异。

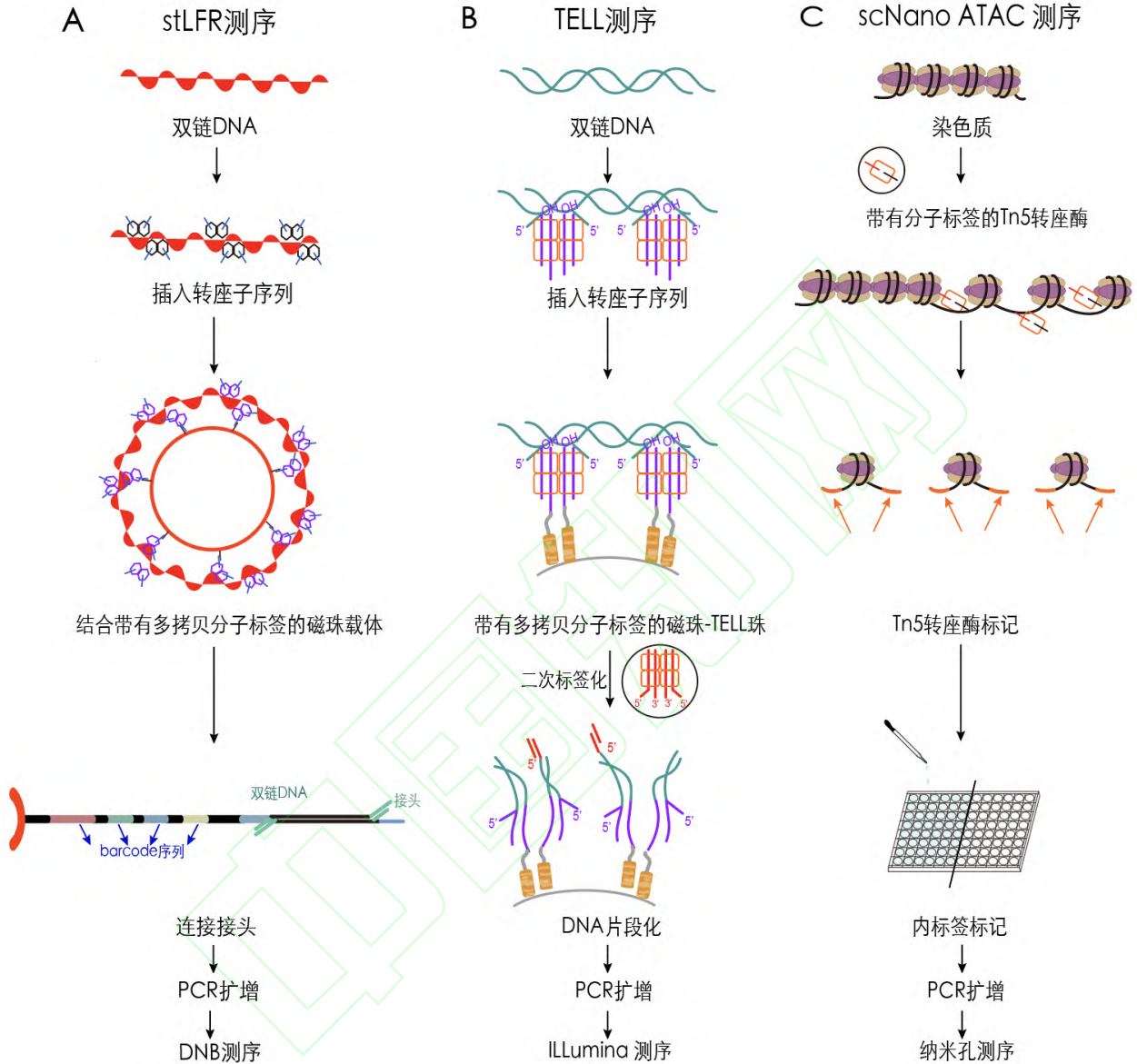


图2 3种新型单分子LRS技术工作原理

由于测序方式和 DNA 样品制备的差异，每一种 LRS 技术均可以产生不同类型的长序列数据，它们的长度和准确性都不同。在表 1 中，我们对长读长数据集特征进行总结整理，并与 Illumina NGS 数据进行比较，重点对比每种测序数据类型的相对长度和精度。

表1 长读长与短读长测序技术的比较^[11-14]

测序技术	单样本价格 (美元)	每百万读数 价格(美元)	运行时长(h)	最大输出 读数(千条)	平均读长(最长读长)	准确率(%)	错误来源
PacBio Sequel I (maximum 8 cells)	1 025	2,050	10-20	500	20 000 (50 000)	83-91/99.9 (CCS)	同聚物
PacBio Sequel II (maximum 16cells)	2 050	410	10-30	4 000	30 000 (130 000)	83-91/99.9 (CCS)	同聚物
ONT Flongle	82	400	0.1-12	200	30 000 (60 000)	96-99	同聚物
ONTMinION	615	615	0.1-48	1 000	50 000 (2.3 M)	96-99	同聚物
ONT GridION (maximum 5FCs)	615	615	48	1 000	50 000 (2.3 M)	96-99	同聚物
ONT PromethION (maximum 48FCs)	1435	256	72	6 000	30 000 (330 000)	96-99	同聚物
Illumina NovaSeq S4 (2 × 150PE; maximum 4 lanes)	5 126	2.6	44	2 000 000	250 (290)	99.9	低质量序列
Illumina MiSeq (2 × 300 PE; maximum 2 lanes)	2 050	103	56	20 000	550 (590)	99.9	低质量序列

FC: Flow cell; PE: Paired-end; CCS: Circular consensus sequencing.

2 LRS 技术在肿瘤研究中的应用

2.1 基因组变异检测

2.1.1 结构变异检测

结构变异是肿瘤的基因组不稳定性的重要标志，包括易位、大片段缺失、扩增和倒位。结构变异通常是通过驱动易位和扩增使癌基因（Oncogene）激活，驱动缺失和倒位使抑癌基因（tumor suppressor gene, TSG）失活。因此对肿瘤特异性突变的高灵敏检测，包括小的变异如单碱基替换及大的变异如结构变异，对于指导肿瘤患者早期检测、复发监测和靶向治疗等至关重要。

(1) 基于 LRS 检测癌细胞基因组的结构变异：NGS 的诞生为研究基因组微小变异提供了便利工具，但对于较大结构变异如结构变异，NGS 技术的短片段(<300 bp)检测仍有很大的局限性。重复区域（包括着丝粒、端粒和其他重复元件）包含人类基因组一半以上的序列（56%），单端测序对于不同位置重复出现的序列片段识别出相同的信息，会导致该序列比对时产生歧义^[15]，而 LRS（高达 20 kb）可以跨越这些重复区域进行片段识别。

LRS 针对结构变异检测策略通常分为 3 种：全面覆盖方法，混合覆盖方法和混合测序方法^[16]。全面覆盖方法指所有样本都以中到高覆盖率进行测序；混合覆盖方法指一部分样本以中到高覆盖率进行测序，其余样本使用低覆盖率进行测序；混合测序方法指一部分样本以中到高覆盖率进行测序，其余样本通过短读长测序进行测序。以上不同策略会影响对群体中常见或罕见变异的检测能力，具体总结为表 2。

表2 不同策略对结构变异检测能力的影响

测序策略	测序深度	优势/劣势
全面覆盖 测序	所有样本进行 LRS: 覆盖率>40 倍（从头组装检测结构变异）	结构变异检测全面，易于挖掘罕见变异
混合覆盖 测序	代表性材料以高覆盖率（覆盖率>30 倍）进行测序，其余样本以低覆盖率（覆盖率>5 倍）进行测序	整体实现较高的结构变异检测灵敏度，且性价比较高
混合测序	所有样本的 10%~20%进行 LRS（覆盖率>30 倍），并对剩余样本进行短读长测序（覆盖率>20 倍）	可获得已识别结构变异的等位基因频率，但会遗漏样本中的稀有结构变异

Hu^[17]和 Aganezov^[18]团队利用 ONT 和 PacBio 平台对乳腺癌样本进行了全基因组测序。研究结果表明，即使在相对较低的覆盖率（25X–30X）下，LRS 不仅可以检测到单核苷酸变异区域，也可以实现更准确和灵敏的结构变异检测，并且不同 LRS 方式产出的结果一致率达 90%~95%。

此外利用 LRS 技术检测结构变异在临床应用中也有巨大潜力。研究人员利用 MinION 技术评估了胰腺导管腺癌细胞系结构变异类型，包括单纯的缺失、易位、倒位以及易位和倒位的组合，该研究对 1:100 稀释度的 PCR 扩增产物进行检测，验证了该方法检测患者特异性结构变异的可行性，且每个样本测序深度仅需要 500 reads 即可快速得到检测结果^[19]。

(2) 基于 LRS 的基因组结构变异快速鉴定可用于癌症监测: 循环肿瘤 DNA(circulating tumor DNA, ctDNA)来源于肿瘤细胞, 已被证明与肿瘤负荷呈正线性相关, 且在癌症复发出现临床症状前几个月即可检出改变^[20]。液体活检可用于检测体液(如血液)中的肿瘤成分, 是一种对患者损伤小、可接受度高的肿瘤负荷动态监测方式^[21]。在液体活检中检测 ctDNA 中代表性结构变异, 可用于高灵敏度检测和量化 ctDNA^[22], 因此, 准确及时地检测血液中包括结构变异在内的肿瘤相关改变对于患者从早期预警到监测复发均具有重要作用。鉴于测序速度快、体积小和成本低等优点, LRS 可能成为早诊、复发或疗效监测过程中癌症相关结构变异检测的理想方式。

为了能够快速和高效地识别一组特异性体细胞结构变异, 用于 ctDNA 监测, Valle-Inclan 等^[23]利用纳米孔测序的长读长和快速测序能力, 开发出一种能够准确选择纳米孔测序数据体细胞 SVs 模型算法, 仅用 3 天时间即可检测卵巢癌和前列腺癌患者体细胞结构变异, 同时, 应用该方法 DNA 样本需求量可以进一步减少到 80 ng。此外, 研究报道了纳米孔测序技术在同一天内完成了对中枢神经系统肿瘤基因组结构变异、拷贝数及甲基化的检测, 并在短时间内进行了 IDH1、IDH2 和 H3F3A 的单核苷酸变异分析^[24]。Zhang 等人^[25]利用 SMART 测序技术检测 ctDNA, 追踪未经治疗的中高危/高危弥漫性大 B 细胞淋巴瘤(diffuse large b-cell lymphoma, DLBCL)的病情演变, 对 38 例中高危/高危 DLBCL 患者的 17 个肿瘤活检样本和 38 个血浆样本进行了基线评估。测序结果显示 ctDNA 与匹配肿瘤组织之间的突变谱具有良好的一致性。此外, 研究人员在 ctDNA 样本中检测到更多的突变位点, 并认为治疗前 TP53 或 B2M 突变均提示预后不良。通过长度长测序技术对血液样本的动态分析, 证实了其在实时评估疗效方面的应用潜力, 以及用于癌症的多模式和快速分子诊断的可行性。

2.1.2 拷贝数变化检测

拷贝数变异是人类基因组变异的一种常见形式, 已被证实与许多病理变化相关, 包括罕见的基因组疾病^[26]、神经系统疾病^[27]和癌症^[28]。在癌症中, 体细胞拷贝数变化(copy number alteration, CNA)促进了基因组的变异, 具有诊断、预后和治疗意义^[29]。CNA 信息可通过染色体分析、核型分析及 DNA-FISH^[30]等方法检测, 然而, 这些方法均存在灵敏度或准确度不高的局限性。NGS 从短读长测序数据中显示高准确度 CNA 信息来提高灵敏度和准确度^[31], 但此类方法成本较高, 应用场所局限于大型、资源丰富的临床中心。

Martignano 等^[32]利用低覆盖率纳米孔测序检测来自癌症患者血浆核酸分子的拷贝数变化, 仅用 200 万读数就完成了对 6 名肺癌患者和 4 名健康受试者的全基因组分子核型检测, 并发现了常见的与肺癌相关的拷贝数改变。这是纳米孔测序成功应用于血浆 DNA 拷贝数分析的案例, 展现出了 LRS 对患者高效、无创检测的潜力。

2.1.3 基因融合检测

基因融合是由于染色体倒位、串联重复、区段缺失或易位事件造成不同的、独立的基因或基因片段发生合并的过程, 在癌症中较为常见。例如, 慢性髓性白血病中可检出 ABL-BCR 基因融合, 导致费城染色体(Philadelphia chromosome, Ph)的形成^[33]; 肺腺癌患者的肿瘤组织中 EML4-ALK 融合基因编码的融合蛋白具有致癌作用^[34-35]。因此检测融合基因, 阐明其形成的机制和演化过程对于肿瘤研究具有重要价值。免疫组织化学和荧光原位杂交是临床检测融合基因使用的两种主要方法^[36]。以上方法用于分析单个基因, 且对于组织样本有一定的要求, 在临床实践中可行性低且耗时。因此, LRS 技术作为能够以最少的样本量测试多个基因的新技术, 可助力融合基因的快速发现。

Jeck William 等^[37]利用 Oxford Nanopore MinION 测序系统, 通过修改锚定多重 PCR 方法进行文库构建, 在测序 15 min 内获得>100 条 reads, 可实现 BCR-ABL1 融合转录本的检测。在 Romagnoli 等^[38]的研究中, LRS 技术能在 60 h 内鉴定出嗜酸性粒细胞增多症患者循环血 DNA 样本中 PDGFR α/β 、FGFR1 和 JAK2 重排, 该结果与诊断时独立进行的 FISH 分析完全一致。同时, 测序结果为识别这几名患者基因融合位点提供了理论依据, 这也是传统方法无法提供的信息。

2.2 全长 RNA 转录本检测: 转录本异构体和长链非编码 RNA 识别

一些转录亚型的异常表达可能会激活癌基因或使抑癌基因失活, 从而导致癌症发生^[39]。在 RNA

水平上研究癌细胞的分子特征，很大程度上依赖于参考转录组的精确度和完整性。然而，由于传统 Illumina 短读长 (-150 bp) 的限制，通常无法捕捉到转录本中的所有外显子，导致一些新的转录亚型无法识别^[40]。Fang 等^[41]对 3 例肝细胞癌 (hepatocellular carcinoma, HCC) 患者的肿瘤组织、门静脉癌栓和癌旁组织进行了纳米孔转录组 RNA 测序。研究发现，短读长的转录本组装可能会出现许多错误，而长读长在识别结构复杂的长片段新转录本方面具有明显的优势。此外，其他研究也使用 LRS 成功地发现了细胞中大量新的转录本^[42-44]。以上研究表明 LRS 是准确组装转录本的有效工具，并且相较于 NGS 技术能够识别更多未知的转录本。

长链非编码 RNA (Long non-coding RNA, lncRNA) 参与基因组修饰、转录激活、转录干扰、核内运输等过程^[45]，其在肿瘤发生发展过程中的调控作用已被大量研究证实。lncRNA 会因细胞类型、组织类型和物种类型的不同产生多种同工型。RNA-seq 可揭示 lncRNA 表达差异，但表达差异不足以代表其作用方式，因此 LRS 技术为探索 lncRNA 不同加工方式及生物学功能提供了有力工具^[46]。为了加速 lncRNA 注释，GENCODE 联盟开发了 RNA 捕获长测序 (RNA capture long seq, CLS)，它将靶向 RNA 捕获与第三代 LRS 相结合，所生成的全长转录本模型能够明确描述 lncRNA 的基因组特征，包括启动子和基因结构，以及蛋白质编码潜能^[47]。

2.3 表观遗传修饰识别

表观遗传学在驱动癌症及其异质性方面具有重要作用，基因表达的表观遗传调控主要依赖于 DNA 甲基化和组蛋白修饰^[48]。迄今为止，研究 DNA 甲基化通常先利用亚硫酸氢盐处理，将胞嘧啶转化为尿嘧啶进而将表观遗传信息转化为遗传信息，再进行大规模平行 DNA 测序。然而，基于亚硫酸氢盐处理的测序无法区分胞嘧啶、5-甲基胞嘧啶 (mC) 和 5-羟甲基胞嘧啶 (hmC)，这增加了文库制备的复杂性以及来自不完全化学转化等偏差的可能性^[49]。此外，常用的 NGS 技术存在读取长度短的限制，使得一些区域难以定位，无法进行等位基因特异性甲基化的研究。Rand 等^[50]和 Simpson 等^[51]首次使用纳米孔测序实现了全基因组 DNA 甲基化分析。Ewing 等人^[52]利用 ONT 测序技术实现了转座元件表观基因组的全面分析，同时评估转座因子 (transposable elements) 甲基化和检测体细胞转座因子动员。Davenport 等^[53]和 Zhang 等^[54]使用纳米孔测序技术在 HCC 全基因组范围内绘制 mC 信号以识别新的 TSG。该研究根据再生肝脏和原发性 HCC 的基因转录谱与甲基化数据的整合识别出 10 个潜在的 TSG，随后对其中一种候选物——葡萄糖激酶的作用进行了探究，验证其通过诱导细胞内乳酸积累来抑制 HCC 细胞的增殖，参与 HCC 的发生发展^[53]。此后，研究者们开发及改进了基于 LRS 技术对表观遗传景观进行建模和分析的工具^[55-56]，将现有 DNA 甲基化检测技术向前推进了一步。

3 总结与展望

LRS 技术助力人类癌细胞基因组及全转录组图谱绘制，使肿瘤患者游离 DNA 检测、大片段结构变异和甲基化等快速检测成为可能，为临床快速筛选、疾病诊断及预后判断提供了有力依据。LRS 技术在肿瘤研究中的应用正在不断扩展，特别是在缺少参考基因组信息的情况下，能够得出更为准确的拼接信息。例如，研究者将宫颈癌多组学数据与人乳头瘤病毒整合位点 LRS 数据相结合，识别和筛选出新的宫颈癌候选靶基因 (BNC1、RSBN1、USP36 和 TAOK3)，并在细胞迁移、增殖和集落形成等体外功能研究中验证了其潜在致癌作用^[57]。

LRS 技术尚在起步阶段，仍然存在一些尚待解决的问题。首要挑战则是对最低样本量的需求。LRS 技术相对 NGS 对检测 DNA 或 RNA 的质量和数量要求更低，但仍需要微克级 DNA 来制备用于全基因组测序的文库，特别是临床血液及体液样本提取的 DNA 量通常很有限，因此，LRS 应用于临床还需要进一步减少对样本量的需求。另一个挑战则是 ONT 和 PacBio 测序存在较高的原始错误率以及长扩增子嵌合体形成率，有待于通过升级优化检测平台和下游分析流程规避可能的错误。例如，ONT 发布的 PromethION 流通池 R10.3 版本测序准确度远高于之前版本，在单读时测序准确性可达到 99%，有助于改善当前长读测序的碱基准确度低的问题，进一步帮助改进下游分析。此外，针对 LRS 数据的下游分析如重复序列比对、基因组结构变异检测和甲基化分析等，依赖人工智能算法、软件和

数据库等生物信息学工具。目前已有的针对 NGS 数据开发的算法无法解决 LRS 数据高原始错误率的问题, 因此开发更适合处理长序列片段数据的算法, 有助于 LRS 技术准确度的提升。

LRS 技术在肿瘤研究中的应用思路和探索领域正在不断扩展, 特别是针对一些具有高突变率的肿瘤, 如血液肿瘤中的各类白血病、多发性骨髓瘤及恶性淋巴瘤等, 其可通过以下应用方向发挥重要作用: 检测肿瘤早期发生的基因突变, 对癌症进行早期预警; 根据分子标志物对疾病进行鉴别诊断及分子分型; 根据患者的基因变异选择合适的靶向药, 做到精准治疗; 利用基因变异导致患者对化疗药物耐受性不同, 进行化疗副作用预测; 利用 LRS 技术检测微小残留灶 (MRD) 进行疗效判断及复发监测, 从而进一步提高分子诊断在肿瘤疾病的早期诊断、精准治疗、判断疗效、复发预测等全程管理中的重要应用价值。

尽管该技术仍然存在诸多亟待解决的问题, LRS 仍是测序技术不断更新换代中的重要里程碑。相信随着技术的革新, 未来 LRS 技术将加速肿瘤基因组学及表观遗传组学研究, 窥探既往难以发现的分子生物学信息, 为新的肿瘤生物标志物识别及诊疗方式开发提供新的思路。

[本文编辑] 杨亚红

[参考文献]

- [1] SENOL CALI D, KIM J S, GHOSE S, ALKAN C, MUTLU O. Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions[J]. *Brief Bioinform*, 2019, 20: 1542-1559.
- [2] KORLACH J, BJORNSON K P, CHAUDHURI B P, CICERO R L, FLUSBERG B A, GRAY J J, et al. Real-time DNA sequencing from single polymerase molecules[J]. *Methods Enzymol*, 2010, 472: 431-455.
- [3] WENGER A M, PELUSO P, ROWELL W J, CHANG P C, HALL R J, CONCEPCION G T, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome[J]. *Nat Biotechnol*, 2019, 37: 1155-1162.
- [4] BOWDEN R, DAVIES R W, HEGER A, PAGNAMENTA A T, CESARE M D, OIKKONEN L E, et al. Sequencing of human genomes with nanopore technology[J]. *Nat Commun*, 2019, 10: 1869.
- [5] JAIN M, FIDDES I T, MIGA K H, OLSEN H E, PATEN B, AKESON M. Improved data analysis for the MinION nanopore sequencer[J]. *Nat Methods*, 2015, 12: 351-356.
- [6] JAIN M, KOREN S, MIGA K H, QUICK J, RAND A C, SASANI T A, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads[J]. *Nat Biotechnol*, 2018, 36: 338-345.
- [7] ALTEMOSE N, MASLAN A, SMITH O K, SUNDARARAJAN K, BROWN R R, MISHRA R, et al. DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome wide[J]. *Nat Methods*, 2022, 19: 711-723.
- [8] WANG O, CHIN R, CHENG X F, WU M K Y, MAO Q, TANG J B, et al. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and *de novo* assembly[J]. *Genome Res*, 2019, 29: 798-808.
- [9] CHEN Z T, PHAM L, WU T C, MO G Y, XIA Y, CHANG P L, et al. Erratum: Ultralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information[J]. *Genome Res*, 2021, 31: 934.
- [10] HU Y Q, JIANG Z H, CHEN K X, ZHOU Z X, ZHOU X, WANG Y, et al. scNanoATAC-seq: a long-read single-cell ATAC sequencing method to detect chromatin accessibility and genetic variants simultaneously within an individual cell[J]. *Cell Res*, 2023, 33: 83-86.
- [11] KARST S M, ZIELS R M, KIRKEGAARD R H, SØRENSEN E A, MCDONALD D, ZHU Q, et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing[J]. *Nat Methods*, 2021, 18: 165-169.
- [12] TEDERSOO L, ALBERTSEN M, ANSLAN S, CALLAHAN B. Perspectives and benefits of high-throughput long-read sequencing in microbial ecology[J]. *Appl Environ Microbiol*, 2021, 87: e0062621.
- [13] WANG Y, ZHAO Y, BOLLAS A, WANG Y, AU K F. Nanopore sequencing technology, bioinformatics and applications[J]. *Nat Biotechnol*, 2021, 39: 1348-1365.
- [14] FOOX J, TIGHE S W, NICOLET C M, ZOOK J M, BYRSKA-BISHOP M, CLARKE W E, et al. Performance assessment of DNA sequencing platforms in the ABRF Next-Generation Sequencing Study[J]. *Nat Biotechnol*, 2021, 39: 1129-1140.
- [15] SEDLAZECK F J, RESCHENEDER P, SMOLKA M, FANG H, NATTESTAD M, VON HAESELER A, et al. Accurate detection of complex structural variations using single-molecule sequencing[J]. *Nat Methods*, 2018, 15: 461-468.
- [16] DE COSTER W, WEISSENSTEINER M H, SEDLAZECK F J. Towards population-scale long-read sequencing[J]. *Nat Rev Genet*, 2021, 22: 572-587.
- [17] HU T B, LI J J, LONG M P, WU J B, ZHANG Z, XIE F, et al. Detection of structural variations and fusion genes in breast cancer samples using third-generation sequencing[J]. *Front Cell Dev Biol*, 2022, 10: 854640.
- [18] AGANEZOV S, GOODWIN S, SHERMAN R M, SEDLAZECK F J, ARUN G, BHATIA S, et al. Comprehensive analysis of structural variants in breast cancer genomes using single-molecule sequencing[J]. *Genome Res*, 2020, 30: 1258-1273.
- [19] NORRIS A L, WORKMAN R E, FAN Y F, ESHLEMAN J R, TIMP W. Nanopore sequencing detects structural variants in cancer[J]. *Cancer Biol Ther*, 2016, 17: 246-253.
- [20] SCHWARZENBACH H, HOON D S B, PANTEL K. Cell-free nucleic acids as biomarkers in cancer patients[J]. *Nat Rev Cancer*, 2011, 11: 426-437.
- [21] HEITZER E, HAQUE I S, ROBERTS C E S, SPEICHER M R. Current and future perspectives of liquid biopsies in genomics-driven oncology[J]. *Nat Rev Genet*, 2019, 20: 71-88.
- [22] OLSSON E, WINTER C, GEORGE A, CHEN Y L, HOWLIN J, TANG M H E, et al. Serial monitoring of circulating tumor DNA in patients with primary breast cancer for detection of occult metastatic disease[J]. *EMBO Mol Med*, 2015, 7: 1034-1047.
- [23] VALLE-INCLAN J E, STANGL C, DE JONG A C, VAN DESSEL L F, VAN ROOSMALEN M J, HELMIJER J C A, et al. Optimizing Nanopore sequencing-based detection of structural variants enables individualized circulating tumor DNA-based disease monitoring in

- cancer patients[J]. *Genome Med*, 2021, 13: 86.
- [24] EUSKIRCHEN P, BIELLE F, LABRECHE K, KLOOSTERMAN W P, ROSENBERG S, DANIAU M, et al. Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing[J]. *Acta Neuropathol*, 2017, 134: 691-703.
- [25] ZHANG S C, ZHANG T T, LIU H Q, ZHAO J, ZHOU H F, SU X X, et al. Tracking the evolution of untreated high-intermediate/high-risk diffuse large B-cell lymphoma by circulating tumour DNA[J]. *Br J Haematol*, 2022, 196: 617-628.
- [26] ANGULO M A, BUTLER M G, CATALETTO M E. Prader-Willi syndrome: a review of clinical, genetic, and endocrine findings[J]. *J Endocrinol Invest*, 2015, 38: 1249-1263.
- [27] STEFANSSON H, RUJESCU D, CICHON S, PIETILÄINEN O P H, INGASON A, STEINBERG S, et al. Large recurrent microdeletions associated with schizophrenia[J]. *Nature*, 2008, 455: 232-236.
- [28] SHLIEN A, MALKIN D. Copy number variations and cancer[J]. *Genome Med*, 2009, 1: 62.
- [29] DÖHNER H, WEI A H, APPELBAUM F R, CRADDOCK C, DINARDO C D, DOMBRET H, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN[J]. *Blood*, 2022, 140: 1345-1377.
- [30] GONZALEZ GARCIA J R, MEZA-ESPINOZA J P. Use of the international system for human cytogenetic nomenclature (ISCN)[J]. *Blood*, 2006, 108: 3952-3953.
- [31] ZEHIR A, BENAYED R, SHAH R H, SYED A, MIDDHA S, KIM H R, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients[J]. *Nat Med*, 2017, 23: 703-713.
- [32] MARTIGNANO F, MUNAGALA U, CRUCITTA S, MINGRINO A, SEMERARO R, DEL RE M, et al. Nanopore sequencing from liquid biopsy: analysis of copy number variations from cell-free DNA of lung cancer patients[J]. *Mol Cancer*, 2021, 20: 32.
- [33] MITELMAN F, JOHANSSON B, MERTENS F. The impact of translocations and gene fusions on cancer causation[J]. *Nat Rev Cancer*, 2007, 7: 233-245.
- [34] SODA M, CHOI Y L, ENOMOTO M, TAKADA S, YAMASHITA Y, ISHIKAWA S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer[J]. *Nature*, 2007, 448: 561-566.
- [35] SODA M, TAKADA S, TAKEUCHI K, CHOI Y L, ENOMOTO M, UENO T, et al. A mouse model for EML4-ALK-positive lung cancer[J]. *Proc Natl Acad Sci USA*, 2008, 105: 19893-19897.
- [36] PISAPIA P, PEPE F, SGARIGLIA R, NACCHIO M, RUSSO G, GRAGNANO G, et al. Methods for actionable gene fusion detection in lung cancer: now and in the future[J]. *Pharmacogenomics*, 2021, 22: 833-847.
- [37] JECK W R, LEE J, ROBINSON H, LE L P, IAFRATE A J, NARDI V. A nanopore sequencing-based assay for rapid detection of gene fusions[J]. *J Mol Diagn*, 2019, 21: 58-69.
- [38] ROMAGNOLI S, BARTALUCCI N, GESULLO F, BALLIU M, BONIFACIO S, FERNANDEZ A G L, et al. Nanopore sequencing for the screening of myeloid and lymphoid neoplasms with eosinophilia and rearrangement of PDGFR α , PDGFR β , FGFR1 or PCM1-JAK2[J]. *Biomark Res*, 2021, 9: 83.
- [39] CHEN H, GAO F, HE M, DING X F, WONG A M, SZE S C, et al. Long-read RNA sequencing identifies alternative splice variants in hepatocellular carcinoma and tumor-specific isoforms[J]. *Hepatology*, 2019, 70: 1011-1025.
- [40] STEIJGER T, ABRIL J F, ENGSTRÖM P G, KOKOCINSKI F, CONSORTIUM R G A S P, HUBBARD T J, et al. Assessment of transcript reconstruction methods for RNA-seq[J]. *Nat Methods*, 2013, 10: 1177-1184.
- [41] FANG Y C, CHEN G, CHEN F, HU E, DONG X Q, LI Z L, et al. Accurate transcriptome assembly by Nanopore RNA sequencing reveals novel functional transcripts in hepatocellular carcinoma[J]. *Cancer Sci*, 2021, 112: 3555-3568.
- [42] KUO R I, CHENG Y Y, ZHANG R X, BROWN J W S, SMITH J, ARCHIBALD A L, et al. Illuminating the dark side of the human transcriptome with long read transcript sequencing[J]. *BMC Genomics*, 2020, 21: 751.
- [43] BOTI M A, ADAMOPOULOS P G, TSIKANIKA P, SCORILAS A. Nanopore sequencing unveils diverse transcript variants of the epithelial cell-specific transcription factor eIF-3 in human malignancies[J]. *Genes*, 2021, 12: 839.
- [44] DE JONG L C, CREE S, LATTIMORE V, WIGGINS G A R, SPURDLE A B, KONFAB INVESTIGATORS, et al. Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events[J]. *Breast Cancer Res*, 2017, 19: 127.
- [45] CHEN L L. Towards higher-resolution and *in vivo* understanding of lncRNA biogenesis and function[J]. *Nat Methods*, 2022, 19: 1152-1155.
- [46] STATELLO L, GUO C J, CHEN L L, HUARTE M. Gene regulation by long non-coding RNAs and its biological functions[J]. *Nat Rev Mol Cell Biol*, 2021, 22: 96-118.
- [47] LAGARDE J, USZCZYNSKA-RATAJCZAK B, CARBONELL S, PÉREZ-LLUCH S, ABAD A, DAVIS C, et al. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing[J]. *Nat Genet*, 2017, 49: 1731-1740.
- [48] GUO M Z, PENG Y J, GAO A A, DU C, HERMAN J G. Epigenetic heterogeneity in cancer[J]. *Biomark Res*, 2019, 7: 23.
- [49] SCHATZ M C. Nanopore sequencing meets epigenetics[J]. *Nat Methods*, 2017, 14: 347-348.
- [50] RAND A C, JAIN M, EIZENGA J M, MUSSELMAN-BROWN A, OLSEN H E, AKESON M, et al. Mapping DNA methylation with high-throughput nanopore sequencing[J]. *Nat Methods*, 2017, 14: 411-413.
- [51] SIMPSON J T, WORKMAN R E, ZUZARTE P C, DAVID M, DURSI L J, TIMP W. Detecting DNA cytosine methylation using nanopore sequencing[J]. *Nat Methods*, 2017, 14: 407-410.
- [52] EWING A D, SMITS N, SANCHEZ-LUQUE F J, FAIVRE J, BRENNAN P M, RICHARDSON S R, et al. Nanopore sequencing enables comprehensive transposable element epigenomic profiling[J]. *Mol Cell*, 2020, 80: 915-928.e5.
- [53] DAVENPORT C F, SCHEITHAUER T, DUNST A, BAHR F S, DORDA M, WIEHLMANN L, et al. Genome-wide methylation mapping using nanopore sequencing technology identifies novel tumor suppressor genes in hepatocellular carcinoma[J]. *Int J Mol Sci*, 2021, 22: 3937.
- [54] ZHANG L L, RONG W Q, MA J, LI H X, TANG X K, XU S Y, et al. Comprehensive analysis of DNA 5-methylcytosine and N6-adenine methylation by nanopore sequencing in hepatocellular carcinoma[J]. *Front Cell Dev Biol*, 2022, 10: 827391.
- [55] ABANTE J, KAMBHAMPATI S, FEINBERG A P, GOUTSIAS J. Estimating DNA methylation potential energy landscapes from nanopore sequencing data[J]. *Sci Rep*, 2021, 11: 21619.
- [56] SAKAMOTO Y, ZAHA S, NAGASAWA S, MIYAKE S, KOJIMA Y, SUZUKI A, et al. Long-read whole-genome methylation patterning using enzymatic base conversion and nanopore sequencing[J]. *Nucleic Acids Res*, 2021, 49: e81.
- [57] IDEN M, TSAIH S W, HUANG Y W, LIU P Y, XIAO M Z, FLISTER M J, et al. Multi-omics mapping of human papillomavirus integration sites illuminates novel cervical cancer target genes[J]. *Br J Cancer*, 2021, 125: 1408-1419.