

DOI: 10.13376/j.cbbls/2025114

文章编号: 1004-0374(2025)09-1183-09

基于长读长测序的eccDNA检测方法与生信分析

张堂轩^{1,2}, 苗碧元^{1,2}, 曾庆松¹, 万绍贵^{1,3*}

(1 赣南医科大学基因组学与精准医学研究所, 赣州 341000; 2 赣南医科大学基础医学院, 赣州 341000; 3 赣南医科大学医学技术学院, 赣州 341000)

摘要: 染色体外环状 DNA (extrachromosomal circular DNA, eccDNA) 是一类独立于染色体存在的环状 DNA 分子, 广泛分布于真核细胞中, 与癌症发生、遗传异质性及治疗耐药性密切相关。当前检测 eccDNA 的主流方法依赖于高通量测序技术, 其中长读长测序凭借其超长读长优势, 可精准解析复杂结构 eccDNA 的完整序列, 克服短读长测序因串联重复序列导致的信息丢失问题。本文系统综述了基于长读长测序的 eccDNA 检测中的实验富集方法 (如 Circle-Seq、3SEP 等) 和生信分析工具, 并重点探讨了相关生物信息学工具的分析流程和在断点识别、序列组装及突变分析中的应用进展。尽管现有工具在提升检测灵敏度和准确性方面成效显著, 但仍面临假阳性率高、数据利用率低等挑战。未来需进一步优化算法、整合多组学数据, 以推动 eccDNA 在癌症标志物开发和精准医疗中的转化应用。

关键词: 染色体外环状 DNA; 长读长测序; 检测技术; 生信分析工具

中图分类号: Q811.4 文献标志码: A

eccDNA detecting technology and bioinformatical analysis using long-read sequencing

ZHANG Tang-Xuan^{1,2}, MIAO Bi-Yuan^{1,2}, ZENG Qing-Song¹, WAN Shao-Gui^{1,3*}

(1 Institute of Genomics and Precision Medicine, Gannan Medical University, Ganzhou 341000, China;

2 School of Basic Medicine, Gannan Medical University, Ganzhou 341000, China;

3 School of Medical and Information Engineering, Gannan Medical University, Ganzhou 341000, China)

Abstract: Extrachromosomal circular DNA (eccDNA) is a class of circular DNA molecules existing independently of chromosomes, widely distributed in eukaryotic cells and closely associated with cancer development, genetic heterogeneity, and therapeutic resistance. Current mainstream methods for eccDNA detection rely on high-throughput sequencing, among which long-read sequencing leverages its ultra-long read length advantages to accurately resolve the complete sequences of complex-structured eccDNA, thereby overcoming the information loss caused by tandem repeats in short-read sequencing. This review systematically summarizes experimental enrichment methods (e.g., Circle-Seq and 3SEP) and bioinformatics analysis tools in eccDNA detection utilizing long-read sequencing technologies, while highlighting advancements in bioinformatics tools for junction identification, sequence assembly, and mutation analysis. Despite progress in improving sensitivity and accuracy, challenges such as high false-positive rates and low data utilization persist. Future research should focus on algorithm optimization and multi-omics integration to advance the translational application of eccDNA in cancer biomarker discovery and precision medicine.

Key words: extrachromosomal circular DNA (eccDNA); long-read sequencing; detection technology; bioinformatical analysis tools

收稿日期: 2024-12-05; 修回日期: 2025-03-07

基金项目: 江西省自然科学基金重点项目(20244BAB28047)

*通信作者: E-mail: wansg@gmu.edu.cn; Tel/Fax: 0797-8169575

染色体外环状 DNA (extrachromosomal circular DNA, eccDNA) 是一类在染色体外独立存在的单链或双链环状 DNA 分子^[1], 它们广泛分布于人体各类组织细胞中^[2-5], 在血浆和尿液中也能被检测到^[6, 7]。其独特的环状结构能够抵御 RNA 酶和外切酶的消化, 使 eccDNA 在染色体结构外相较于线性 DNA 更稳定^[5], 此外, eccDNA 在细胞分裂过程中的非均匀分布模式进一步加剧了肿瘤内部的遗传异质性。根据大小和序列特征, eccDNA 可分为几种类型: 微小 DNA (microDNA, 约 200~400 bp)、小多分散 DNA (small polydispersed DNA, spcDNA, 100 bp~10 kb)、端粒环 (telomeric circles, t-circles) 以及环状 DNA (extrachromosomal DNA, ecDNA)。其中, microDNA 和 ecDNA 在近几年中最受关注。在真核细胞中, 主要来源于非重复性基因组序列的 microDNA 更为常见^[8]。而在肿瘤细胞中, 携带完整基因和调控区域的大分子 ecDNA 更为常见^[9]。它们能够从多方面影响癌症的发生和发展。例如, microDNA 能够在无需经典启动子序列的情况下进行转录, 产生的调控 RNA 通过 RNA 干扰途径抑制基因表达^[10]。ecDNA 上的肿瘤驱动基因更易扩增^[11], 并且相比于染色体上的癌基因, ecDNA 上的癌基因展现出更强的转录活性^[12]。临床样本研究直接表明, ecDNA 在恶性肿瘤中的出现频率高于良性肿瘤, 并且在正常组织中的出现频率较低, 显示出作为肿瘤标志物的潜力^[13, 14]。因此, microDNA 和 ecDNA 作为现阶段最受关注且研究最多的 eccDNA 类型, 对于理解癌症的分子机制和开发新的治疗策略具有重要意义。因为分类与鉴定方法的选择并无密切联系, 在本文中这两类均统称为 eccDNA。

随着高通量测序技术及生物信息学分析方法的不断进步, eccDNA 的检测手段也逐渐丰富。这些技术使得我们能够深入解析 eccDNA 所携带的遗传信息, 进而揭示其在遗传变异、肿瘤发生、免疫反应和耐药性中的关键作用, 为研究其生物学功能和临床应用提供了可能^[15-17]。高通量测序技术, 尤其是二代测序 (next generation sequencing, NGS) 技术, 因其高通量和高准确性, 在检测 eccDNA 方面发挥了重要作用。NGS 技术能够检测出跨越 eccDNA 断点的读段, 从而定位 eccDNA 在基因组上的位置。然而, NGS 在检测复杂结构 eccDNA 时存在一定的局限性, 尤其当 eccDNA 源自基因组内的重复区域或生成时发生结构变异。此外, 从 NGS 数据中鉴定 eccDNA 的生信分析工具种类繁多, 然而对于复

杂 eccDNA 的结构解析困难, 且往往只能提供 eccDNA 对应于基因组上的坐标, 无法完整展现 eccDNA 的实际序列内容。三代测序技术, 即长读长测序技术 (long-read sequencing), 以其更长的读长、直接测序和实时测序等优势, 越来越被广泛应用于 eccDNA 的检测。现有的从长读长测序数据鉴定 eccDNA 的生信分析工具, 包括 ciderseq2^[18]、ecc_finder^[19]、eccDNA_RCA_nanopore^[20, 21]、circular-calling^[22]、CReSIL^[23] 和 FLED^[24]。它们应用于不同的数据类型, 各有所长, 均能输出 eccDNA 对应于基因组上的坐标和 eccDNA 序列内容。本文归纳了基于长读长测序技术的 eccDNA 检测方法, 并探讨了相关生信分析工具的发展, 为 eccDNA 研究提供新视角。

1 长读长测序技术原理概述及其在 eccDNA 检测中的应用

长读长测序技术分为 Pacific Biosciences (PacBio) 的单分子实时测序和 Oxford Nanopore Technologies (ONT) 的纳米孔测序。PacBio 测序技术依托于 SMRT (单分子实时) 芯片, 在 DNA 聚合酶将新的脱氧核糖核苷酸整合到复制链的过程中, 产生微弱的光信号, 这些信号被精密的探测器捕获^[25]。通过分析这些光信号, 测序系统能够准确识别出合成的 DNA 中的脱氧核糖核苷酸类型。在 HiFi 测序模式下, DNA 聚合酶围绕样品 DNA 分子的环状结构进行多次循环工作, 通过对比同一分子的不同副本, 精确确定样品 DNA 的正确序列。与 PacBio 测序技术依赖于捕捉荧光信号不同, Nanopore 测序技术专注于检测电信号^[26]。该技术的核心在于一个纳米级别的孔洞, 即纳米孔 (nanopore), 孔内共价结合有分子接头。通过将纳米孔蛋白固定在电阻膜上, 并利用动力蛋白牵引核酸链穿过纳米孔, 可以实现对核酸的直接测序。核酸通过纳米孔时会引起电荷的变化, 进而导致电阻膜上电流的变化。由于纳米孔的直径非常细小, 仅允许单个脱氧核糖核苷酸通过, 而 A、T、C、G 四种碱基的带电性质各异, 因此它们通过纳米孔时对电流的干扰也各不相同。通过实时监测这些电流信号的变化并对其进行解码, 可以准确地确定碱基类型, 实现高效的测序过程。两种测序平台均具备无须扩增即可快速获得长读长数据的能力, 同时也能够进行直接的甲基化检测。特别是 Nanopore 测序技术, 以其超长读长而备受青睐, 因其无须像 PacBio 测序那样在建库过程中对 DNA 进

行片段化。在 eccDNA 的检测中, 长读长测序技术能够保留完整的 eccDNA 连续序列信息, 避免了短读长测序中通过片段重组所带来的信息丢失, 能够更高效地鉴定嵌合 eccDNA。然而, 值得注意的是, 尽管长读长测序技术在某些方面具有优势, 其测序准确度相较于 NGS 技术仍有提升空间。因此, 在利用长读长测序技术进行 eccDNA 检测时, 选择合适的生物信息分析工具至关重要。除此之外, 研究者们通常会随机挑选 eccDNA 样本进行聚合酶链式反应 (PCR) 和 Sanger 测序验证, 以确定检测到的 eccDNA 的真实性^[27, 28]。

2 基于长读长测序的eccDNA检测实验方法

eccDNA 在不同组织中的丰度存在显著差异^[16], 现如今也没有统一的量化其丰度的检测方法^[29]。在以往的研究中, eccDNA 富集的数量少和质量差一直是个难题。因此, eccDNA 检测的关键实验步骤是提取与纯化。目前, eccDNA 的提取纯化方法主要涉及全基因组测序的 gDNA 提取、eccDNA 富集 (eccDNA-seq) 及其他实验方法。传统的 eccDNA 提取方法, 如 Hirt 提取、CsCl-EtBr 密度梯度离心和 2D 凝胶电泳^[30, 31], 虽然在早期研究中发挥了重要作用, 但存在局限性。这些方法通常耗时较长, 且可能无法高效地分离 eccDNA, 特别是对于低丰度或小尺寸的 eccDNA^[20]。此外, 这些方法在处理复杂样本时, 检测的灵敏度和特异性较差。不同标本可能涉及不同的前期处理及保存期限, 如血液样本需要考虑其保存液及冷藏期限, 有研究表明, DNA 提取率与储存时间和条件有显著相关性, 相较于不同温度的长期储存样本, 反复冻融更加影响样本提取率^[32]。

为了提高 eccDNA 检测的灵敏度和特异性, 研究者们开发了多种 eccDNA 的提取与纯化方法, 并不断进行改良。现如今, 主要方法包括 Circle-Seq^[33]、3SEP^[21] 及 CIDER-Seq^[34] 技术。Circle-Seq 利用碱性溶液裂解细胞和离子交换柱的温和重力流从样本中粗提取环状 DNA, 结合质粒安全的 ATP 依赖性 DNA 内切酶 (plasmid-safe™ ATP-dependent DNase) 除去线性 DNA 和 phi29 DNA 聚合酶进行滚环扩增 eccDNA, 使用磁珠能够有效富集细胞中的 eccDNA, 并通过高通量测序进行验证。这种方法的创新之处在于其无需尺寸排除步骤, 适用于各种真核生物, 且能够以核苷酸分辨率检测大量独特的 eccDNA 类型^[33, 35]。3SEP 技术则是在 Circle-Seq 基础上进一步优化, 不仅将无缓冲氢氧化钠溶液改为 pH 值为 11.8 的改性碱性缓冲液裂解细胞, 除去线性 DNA 之前用 PacI 限制性核酸内切酶切割线粒体 DNA, 最关键的在于利用 solution A 溶液选择性地回收 DNA, 以此提高 eccDNA 的富集效率。该方案中改良的碱性裂解和 PS DNase 消化方法, 结合选择性回收, 显著提高了纯化效率和可重复性, 且能够在一天内完成纯化^[20]。CIDER-Seq 被开发用于环状病毒 DNA 和植物中 eccDNA 的提取, 基于随机引物的环状 DNA 扩增和酶促 DNA 修复步骤, 然后进行建库测序。大致方法与上述两种差别不大, 不同之处详见表 1。

在比较这些方法时, 我们可以看到它们在实验设计上的共同点 (图 1, 用 BioGDP.com 绘制): 通过物理或化学手段减少线性 DNA 的污染, 并利用特定的纯化步骤来富集 eccDNA。

已有实验用 qPCR 分析发现, solution A 溶液纯化和滚环扩增步骤都大大增加了环状 DNA 与线

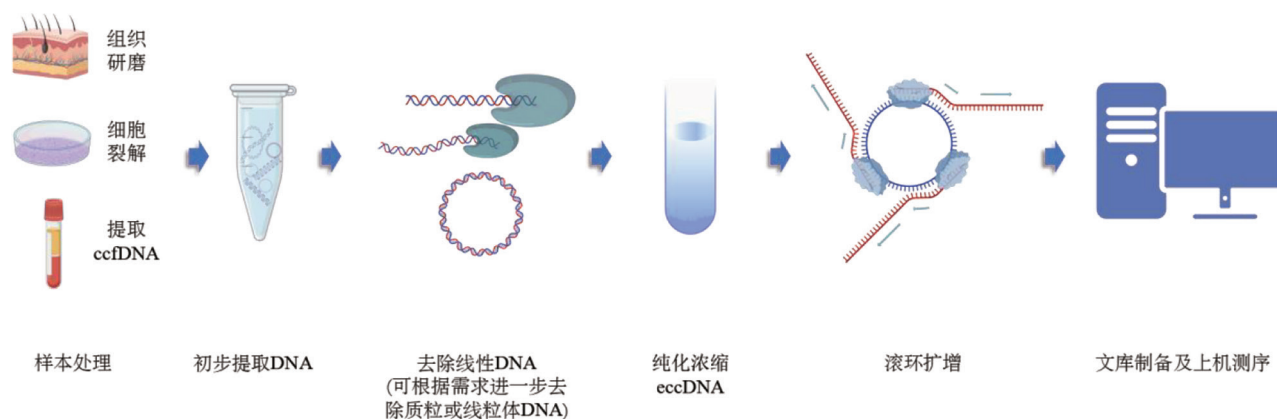


图1 eccDNA富集实验

表1 三种eccDNA富集实验

实验名称	适用样本	初步提取	去除线性DNA	纯化eccDNA	滚环扩增	纯化	进一步延伸	纯化	脱支处理	纯化	修复	纯化
Circle-seq	所有类型的真核生物、原核生物或离体DNA样本	柱膜富集 (Plasmid Mini AX, A&A Bio-technology)	外切酶：质粒安全DNA酶。外切酶V (已测试) 及外切酶VIII。(可选)内切酶： MssI、NotI及Swal	磁珠富集(Agencourt AMPure XP beads)	phi29 DNA聚合酶和随机引物滚环扩增eccDNA	磁珠富集	-	-	-	-	-	-
CIDER-seq	所有已知和未知病毒感染样本及真核植物	总DNA提取 (可根据需求进行片段分选)	-	-	phi29 DNA聚合酶和随机引物滚环扩增eccDNA	3 mol/L 钠和100% 冷冻乙醇化DNA	phi29 DNA聚合酶	3 mol/L 醋酸钠和100% 冷冻乙醇化DNA	SI核酸酶	3 mol/L 醋酸钠和100% 冷冻乙醇化DNA	DNA聚合酶I及T4 DNA聚合酶	磁珠富集 (KAPA Pure或AMPure)
3-Sep	细胞样本、组织样本或其他类型样品中纯化的eccDNA	柱膜富集 (QIAGEN Plasmid Plus Midi Kit)	ATP依赖的PS DNA酶-消化线性DNA、PacI-线性化mtDNA	酚-氯仿-异戊醇萃取，糖原沉淀，用solution A溶液结合Dynabeads链霉亲和素磁珠选择性富集	phi29 DNA聚合酶和随机引物滚环扩增eccDNA	SPRI select 磁珠	-	-	T7内切酶I	糖原沉淀	-	SPRI select磁珠

性 DNA 的 $\log_2$ 比率^[36]。eccDNA 富集方法检测到的 eccDNA 比例均明显高于 WGS 和 ATAC-seq, 除此之外, 包含滚环扩增步骤的方法比没有滚环扩增的方法实现了更高的 eccDNA 检测效率。目前, 推测滚环扩增可能偏好于优先扩增 10 kb 以下的 eccDNA, 鉴于 eccDNA 分子主要在 10 kb 以下, 或许进行 3SEP 和滚环扩增步骤能够更大程度地富集 eccDNA。这些方法的改进不仅提高了 eccDNA 研究的效率, 还为研究者保留了更准确的序列信息来探索 eccDNA 在生物学过程中的作用。然而, 这些方法也有其局限性。例如, 操作步骤较复杂, 实验时长长, 在大规模实验时, 需要降低实验操作的复杂性, 提高实验的可重复性。此外, 这些方法可能需要特定的试剂和设备, 这可能限制了它们在资源有限的实验室中的应用。综上所述, 尽管传统的 eccDNA 提取方法在某些情况下仍然有其价值, 但改良方法如 Circle-Seq 和 3SEP 提供了更高的灵敏度、特异性和效率。现如今, 各研究组在这些方法的基础上, 根据自身实验需求进行不同程度的调整优化以达到对应的实验目的, 例如调整酶切时间以期减少线粒体 DNA 和线性 DNA 的干扰^[27], 或者基于 Circle-seq 进一步开发出人工 eccDNA 合成方法 CAES^[37]。未来的研究可能会继续优化这些方法, 以进一步提高它们的适用性和成本效益。

3 基于长读长测序的eccDNA生信分析工具

基于 eccDNA 富集方法进行长读长测序, 现有六种成熟的 eccDNA 长读长测序数据的生信分析工具, 分别是: *cirdersseq2*^[18]、*ecc_finder*^[19]、*eccDNA_RCA_nanopore*^[21]、*circular-calling*^[22]、*CRoSIL*^[23] 和 *FLED*^[24]。以下按照其开发时间进行逐一介绍(图2)。

3.1 cirdersseq2

cirdersseq2 分析工具从结合 CIDER-seq 实验方法^[34]的 PacBio 测序数据中挖掘出 eccDNA 的基因组位点和序列信息, 主要应用于检测病毒环状基因组, 后续也被应用于检测植物中的 eccDNA^[38]。在使用 *cirdersseq2* 时需要对测序数据进行预处理: 使用 MUSCLE 将分箱的 ROI 与两个修饰的 EACMV DNA 序列 (DNA A 和 DNA B, 对于各自分箱中的序列) 进行比对。从路线的两端运行一个 10 nt 的滑动窗口。在比对读段和参考序列时, 滑动窗口第一次检测到同一性达 90% (即滑动窗口中 10 个核苷酸有 9 个相同) 的序列片段 (位于序列两端) 分别指定为序列的起点和终点。随后用 DeConcat 进一步分析这两个点之间的序列 (称为修剪的 ROI), 最后得到每个 sub-read 中推定的病毒环状 DNA 序列的起始和终止位置。大多数 reads 只需要循环 1~3 轮拼接即可解析其序列。其分析原理较为独特,

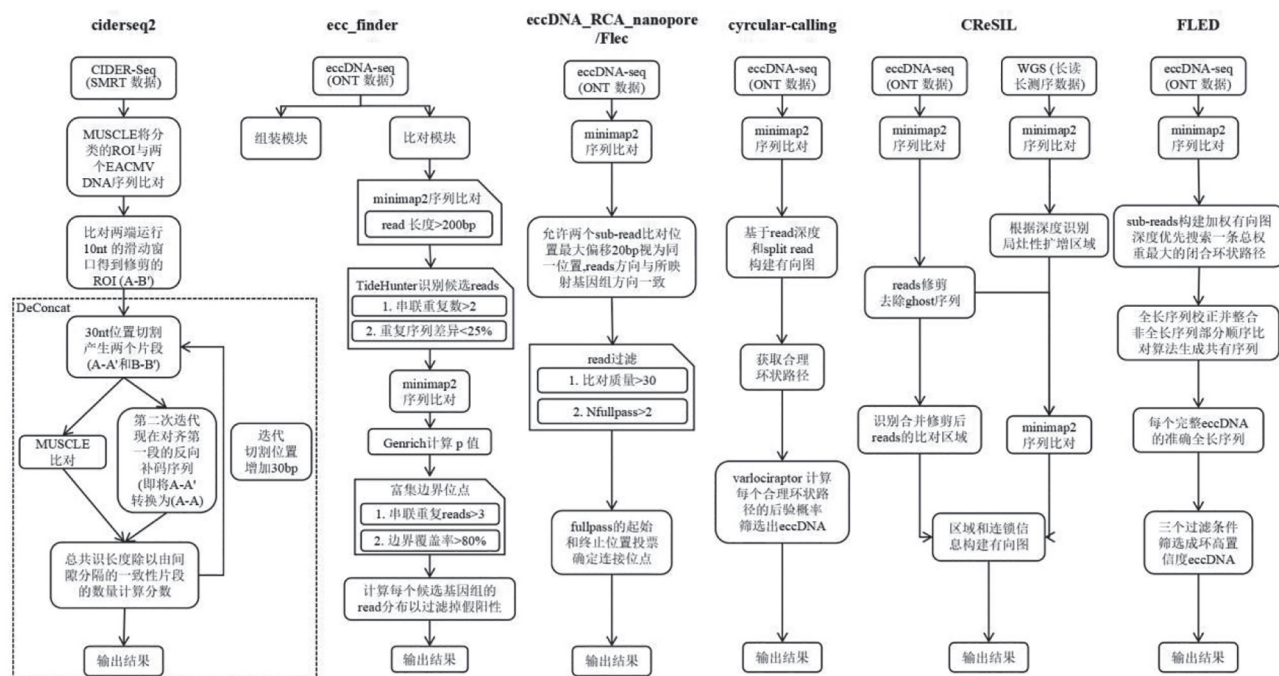


图2 六种分析方法原理流程比较图

能够在无参考基因组的情况下获取 eccDNA 的基因组位置和一致性序列。ciderseq2 目前应用范围较局限,只用于分析环状病毒和植物中的 eccDNA 序列^[39],尚未用其进行动物的 eccDNA 检测。

3.2 ecc_finder

ecc_finder 既可以分析 Illumina 短读长测序数据,又能够分析 ONT 长读长测序数据,并且提出了两种分析模式:由参考基因组引导的比对模式(mapping mode)或无参考的从头组装模式(assembly mode)^[19]。在比对模式中, ecc_finder 使用来自比对 reads 的串联重复序列识别候选基因座。首先, ecc_finder 使用 minimap2 进行序列比对,排除源自线性基因组的重复序列(例如卫星序列)。然后,利用 TideHunter 识别具有串联重复模式的候选 reads,并将这些 reads 划分为重复单元。再次使用 minimap2 将这些重复单元与参考基因组进行比对,筛选出两个以上 reads 覆盖的基因座。利用 Genrich 给出的具有对数正态分布的空模型计算基因组每个脱氧核糖核苷酸的 p 值,得到 eccDNA 的准确末端位置并过滤掉假阳性。最后生成每个产生 eccDNA 的基因座的坐标和相应的 eccDNA 序列的 bed 文件,随后被进一步归一化。在组装模块中, ecc_finder 使用 Tidehunter 组装 reads。通过对组装的重叠群进行聚类来构建一个代表集,使用 CD-hit 将重叠群与重叠群自动对齐,并过滤掉具有 80% 相似度的冗余重叠群,输出按支持 reads 数排序的所有重叠群的 FASTA 序列文件。最终输出 eccDNA 的基因组位置(不包括正负链情况)和一致性序列。相较于组装模式,开发者更倾向于用参考基因组的定位模块从 ONT 数据中鉴定 eccDNA。在实际应用中,已利用 ecc_finder 在长读长测序数据中发现 eccDNA 存在嵌合或截断插入转座子的现象^[40],但更多的是用 ecc_finder 对二代测序数据进行分析^[41],结合 Circle-Map 的分析结果进行两两交集,得到可信度更高的 eccDNA 再进行下一步分析^[42-44]。

3.3 eccDNA_RCA_nanopore (Flec)

eccDNA_RCA_nanopore 是一种单一用于分析经过滚环扩增的 eccDNA 的 ONT 数据的工具,通过在每个 read 中对齐且连续的 sub-read 来检测全长 eccDNA,也需要用 minimap2 进行比对预处理,得到 PAF 文件,然后运行 Flec 工具。

该工具使用 PAF 文件输入 eccDNA 片段组成(每个片段的染色体和基因组起始和终止位置)、连续片段覆盖率(pass 数)和来自每个 read 的一致性

序列。由于 Nanopore 测序的不准确性和易出现间隙,允许两个 sub-read 的映射基因组位置(起始和终止位置)的偏移最大 20 bp,被视为映射到同一位置。sub-reads 顺序、位置或链不一致的被丢弃。eccDNA 片段的确切末端是利用 sub-read 的起始和终止位置进行投票来确定。

该分析工具运行简单快速,其核心优势在于能够直接从测序 read 中获得 eccDNA 的完整序列,无须进行序列组装或剪接步骤。借助该工具,在胆道狭窄患者的胆汁中发现携带特殊增强子的 eccDNA 可能导致对应癌基因显著过表达,有作为预后不良生物标志物的潜能^[45]。除此之外,还发现癌组织和配对的癌血浆样本之间存在一些个体特异性的共享小 eccDNA^[46]。由于该分析工具要求至少两次跨越原始 eccDNA 连接位点的支持 reads,以确保 reads 中的重复模式得以准确识别。然而,由于 eccDNA 的大小和扩增酶活性的限制,这种严格的要求往往导致测序数据的利用率差。此外,在多次数据分析结果中,均发现 eccDNA_RCA_nanopore 的鉴定结果中存在 eccDNA 冗杂现象^[23, 36],这表明该方法尚需进一步的优化。

3.4 circular-calling

circular-calling 基于构建有向图进行组装从而从长读长测序数据中鉴定 eccDNA 环^[22]。长读长测序数据用 minimap2 进行参考基因组比对预处理后,基于 reads 深度信息和 split reads 构建了一个有向图,将片段连接位置的情况分为三种:覆盖(coverage)、缺失(deletion)和断裂(split)。随后筛选出合理路径:如果路径属于 coverage 和 deletion 或 coverage 和 split 连接,则认为该路径是合理的。对于每个合理的循环路径,通过将 deletion 和 split 转换为 VCF 格式中的断端链来生成一个候选事件。然后,用 varlociraptor 计算每个合理的循环路径的后验概率来筛选出 eccDNA。该工具要求用户具备一定的生物信息学背景,能够熟练使用 Snakemake 工具。

3.5 CReSIL

CReSIL (Construction-based Rolling-circle-amplification for eccDNA Sequence Identification and Location) 是基于 NanoCircle 和 CReSIL 两种分析工具的优化升级版^[23]。NanoCircle 利用 split reads 来确定 eccDNA 片段形成过程中基因组区域的连接断点坐标,但在处理复杂环状结构时存在组装错误的问题;CReSIL 利用与基因组区域对齐的序列 reads 直接进行从头组装^[3]。改良后的 CReSIL 能够

同时适用于分析 eccDNA 富集测序数据和全基因组测序 (WGS) 数据, 核心原理是利用比对片段构建有向图。具体步骤如下: 第一步是用 minimap2 进行比对, 对原始 reads 进行预处理修剪, 除去 ghost 序列。第二步聚合修剪 reads 的染色体对齐位置, 以识别代表 eccDNA 染色体起源的合并区域。此时, 获得三种类型的对齐 reads: (1) 仅对齐到一个区域的 reads (normal reads), (2) 在没有 CTC 的多个区域上对齐的 reads (breakpoint reads without CTC), 以及 (3) 在有 CTC 的多个区域上对齐的 reads (CTC reads)。第三步将记录区域、末端位置和链方向的信息构建有向图。第四步构建有向图, 利用区域和有向片段组装成环状结构。最后生成 eccDNA 文件。可选择性输出可视化 eccDNA 结果。该工具不仅能够定位 eccDNA 在基因组上的位置, 并获取相应的 eccDNA 序列, 而且还能够识别 eccDNA 上的突变。在使用高深度模拟数据进行评估时, CReSIL 相较于 ecc_finder、NanoCircle 和 eccDNA_RCA_nanopore, 获得了更高的 F1 值^[36]。F1 值是机器学习中用于评估分类模型性能的一个重要指标, 适用于不平衡数据集的分类任务, 其值较容易受到真值定义的影响。它是精确率 (Precision) 和召回率 (Recall) 的调和平均值, 通过一个单一的数字综合考虑这两个方面, 从而提供更全面的性能评价。通常情况下, F1 值越大, 说明该鉴定工具的综合可信度越高。然而, CReSIL 目前仍存在一些不足之处。例如, 其运行时间较长, 且鉴定得到的 eccDNA 数量相对较少。该工具目前常与 Circle-Seq 实验方法相结合, 被整合进成套 eccDNA 分析流程中, 用于 eccDNA 长读长测序数据的上游鉴定^[47]。

3.6 FLED

FLED 是一种结合滚环扩增和纳米孔长读长测序技术的生信工具, 专门用于鉴定经过滚环扩增的 eccDNA 长读长测序数据, 即能够分析经过 Circle-Seq 和 3SEP 富集的 eccDNA 长读长数据。首先, 利用 minimap2 进行序列比对, 然后基于 sub-reads 构建加权有向图, 并设置硬过滤器, 要求连续比对 sub-reads 的总长度超过 reads 的 70%。通过深度优先搜索 (DFS) 算法, 识别出总权重最大的封闭环状路径, 作为 eccDNA 的组成。在全长序列校正步骤中, FLED 整合了映射到同一基因组区域的高质量非全长读长, 并使用 SPOA 算法为每个组生成共有序列, 作为每个完整 eccDNA 的准确全长序列。候选 eccDNA 的支持 reads 编号包括至少具有 2 个断点的全长

reads、具有 1 个断点的全长 reads 和具有断点的部分 reads。第三步, FLED 对检测到的 eccDNA 应用三个过滤器, 计算 eccDNA 连接位点周围 50 bp 的测序深度, 并根据 Wilcoxon 检验和测序深度比率输出候选 eccDNA。最终, FLED 输出 eccDNA 在基因组中的准确位置、一致性序列和突变情况。该工具也利用模拟数据进行了准确性评估, 遗憾的是, 目前尚未有文献将其与 CReSIL 的 F1 值进行比较。

4 问题与展望

现如今已有多种成熟的分析方法用于检测 eccDNA, 但仍存在许多挑战亟待克服。短读长测序技术在解析基因组中的重复区域和结构变异时存在固有的局限性。现如今主要利用二代测序结合数据分析对 eccDNA 进行检测, 单独使用 Circle-Map 或 AmpliconArchitect (AA)^[48-50], 或者将 Circle-Map 和 ecc_finder 联合使用, 最后进行 eccDNA 筛选较为常见。这些早期开发的工具主要依赖有限的判定标准来识别潜在的 eccDNA, 这可能导致不同程度的假阳性结果或者遗漏, 并且这些工具通常只提供候选 eccDNA 的基因组位置信息。相比之下, 为长读长测序数据设计的分析工具能够基于测序读段直接输出 eccDNA 对应位置在基因组上的位置和完整的 eccDNA 序列。除此之外, eccDNA_RCA_nanopore、circular-calling、CReSIL 和 FLED 等工具不仅能够识别 eccDNA, 还能报告其突变信息, 这为深入理解 eccDNA 的功能和机制提供了更为丰富的数据。

由于处理 WGS 数据需要大量的计算资源, 当专注于研究 eccDNA 特征时, 研究者更倾向于进行 eccDNA 富集。除了 ciderseq2 和 ecc_finder (assembly mode), 其他工具分析的 eccDNA 测序数据均经过滚环扩增。这种扩增导致长读长测序的原始数据中, 单条 read 均包含原始 eccDNA 序列的串联重复。因此, 源自 eccDNA 的 reads 将在同一方向上显示两个或多个用于鉴定序列的 sub-read。ecc_finder (mapping mode)、eccDNA_RCA_nanopore、circular-calling、CReSIL 和 FLED 工具均遵守这一原则, 它们基于读段中包含的完整 eccDNA 序列及其两端的连接位点序列来构建 DNA 的环状结构。在所有长读长测序数据分析工具中, ecc_finder 和 CReSIL 的组装模块尝试利用这些 sub-read 片段进行组装, 但其效果仍需进一步验证。其他分析工具均利用了 minimap2 将测序数据与参考基因组进行序列比对。eccDNA_RCA_nanopore 能够从单个读段中提取完整的 eccDNA

序列, 无须进行序列组装, 但由于滚环扩增引入的错误序列和测序过程中的随机误差, 这种方法保留了 read 中偶然出现的随机错误情况, 最终导致鉴定结果的假阳性率较高。circular-calling、CReSIL 和 FLED 则基于比对 reads 构建有向图来重建 eccDNA 的内部结构, 它们在确定 eccDNA 连接位点和过滤假阳性方面各有千秋, 所有用于减少误报的参数都是可调整的, 以适应不同的数据集和研究需求。值得一提的是, CReSIL 和 FLED 特别保留了从跨越连接断点的非全长读段中识别出的 eccDNA 断点, 这提高了数据利用率, 但同时也增加了候选 eccDNA 位置不精确的风险, 因为两者纳入的片段质量要求不高, 存在引入错误的可能, 尤其是当测序深度增加时, eccDNA 的末端位置可能更不准确^[36]。

综上所述, 随着技术的不断发展, 我们对 eccDNA 的检测和解析能力得到了显著提升, 但仍旧存在着各种亟待解决的问题。倘若能够得到进一步的优化, eccDNA 的研究将会更进一步。这不仅加深了我们对 eccDNA 在多种生物学过程中的作用, 以及与癌症等疾病的关系的理解, 也为未来的临床应用奠定了基础, 推动 eccDNA 作为生物标志物的应用。未来的研究需要进一步整合和优化这些技术, 以实现 eccDNA 更全面、更精确的分析。

[参 考 文 献]

- [1] Dillon LW, Kumar P, Shibata Y, et al. Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep*, 2015, 11: 1749-59
- [2] Möller HD, Mohiyuddin M, Prada-Luengo I, et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat Commun*, 2018, 9: 1069
- [3] Henriksen RA, Jenjaroenpun P, Sjøstrøm IB, et al. Circular DNA in the human germline and its association with recombination. *Mol Cell*, 2022, 82: 209-17.e7
- [4] Zhao XK, Xing P, Song X, et al. Focal amplifications are associated with chromothripsis events and diverse prognoses in gastric cardia adenocarcinoma. *Nat Commun*, 2021, 12: 6489
- [5] Kumar P, Dillon LW, Shibata Y, et al. Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. *Mol Cancer Res*, 2017, 15: 1197-205
- [6] Sin STK, Jiang P, Deng J, et al. Identification and characterization of extrachromosomal circular DNA in maternal plasma. *Proc Natl Acad Sci USA*, 2020, 117: 1658-65
- [7] Lv W, Pan X, Han P, et al. Circle-Seq reveals genomic and disease-specific hallmarks in urinary cell-free extrachromosomal circular DNAs. *Clin Transl Med*, 2022, 12: 4
- [8] Shibata Y, Kumar P, Layer R, et al. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science*, 2012, 336: 82-6
- [9] Verhaak RGW, Bafna V, Mischel PS. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer*, 2019, 19: 283-8
- [10] Paulsen T, Shibata Y, Kumar P, et al. Small extrachromosomal circular DNAs, microDNA, produce short regulatory RNAs that suppress gene expression independent of canonical promoters. *Nucleic Acids Res*, 2019, 47: 4586-96
- [11] Turner KM, Deshpande V, Beyter D, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature*, 2017, 543: 122-5
- [12] Wu S, Turner KM, Nguyen N, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature*, 2019, 575: 699-703
- [13] Yan Y, Guo G, Huang J, et al. Current understanding of extrachromosomal circular DNA in cancer pathogenesis and therapeutic resistance. *J Hematol Oncol*, 2020, 13: 124
- [14] Chiu RWK, Dutta A, Henssen AG, et al. What is extrachromosomal circular DNA and what does it do? *Clin Chem*, 2020, 66: 754-9
- [15] Ling X, Han Y, Meng J, et al. Small extrachromosomal circular DNA (eccDNA): major functions in evolution and cancer. *Mol Cancer*, 2021, 20: 113
- [16] Paulsen T, Kumar P, Koseoglu MM, et al. Discoveries of extrachromosomal circles of DNA in normal and tumor cells. *Trends Genet*, 2018, 34: 270-8
- [17] Wang M, Chen X, Yu F, et al. Extrachromosomal circular DNAs: origin, formation and emerging function in cancer. *Int J Biol Sci*, 2021, 17: 1010-25
- [18] Mehta D, Hirsch-Hoffmann M, Were M, et al. A new full-length circular DNA sequencing method for viral-sized genomes reveals that RNAi transgenic plants provoke a shift in geminivirus populations in the field. *Nucleic Acids Res*, 2019, 47: e9
- [19] Zhang P, Peng H, Llauro C, et al. ecc_finder: a robust and accurate tool for detecting extrachromosomal circular DNA from sequencing data. *Front Plant Sci*, 2021, 12: 743742
- [20] Wang Y, Wang M, Zhang Y. Purification, full-length sequencing and genomic origin mapping of eccDNA. *Nat Protoc*, 2022, 14: 3540
- [21] Wang Y, Wang M, Djekidel MN, et al. eccDNAs are apoptotic products with high innate immunostimulatory activity. *Nature*, 2021, 599: 308-14
- [22] Tüns AI, Hartmann T, Magin S, et al. Detection and validation of circular DNA fragments using nanopore sequencing. *Front Genet*, 2022, 13: 867018
- [23] Wanchai V, Jenjaroenpun P, Leangapichart T, et al. CReSIL: accurate identification of extrachromosomal circular DNA from long-read sequences. *Brief Bioinform*,

- 2022, 23: bbac422
- [24] Li F, Ming W, Lu W, et al. FLED: a full-length eccDNA detector for long-reads sequencing data. *Brief Bioinform*, 2023, 24: bbad388
- [25] Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*, 2015, 13: 278-89
- [26] Wang Y, Zhao Y, Bollas A, et al. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol*, 2021, 39: 1348-65
- [27] Ran XK, Zhao XF, Wei ZW, et al. Circle-seq reveals that eccDNA may be a key blood biomarker for HBV-associated liver cancer. *Front Genet*, 2025, 15: 1454153
- [28] Lin Z, Dai F, Li B, et al. Integrating Circle-Seq with transcriptomics reveals genome-wide characterization of extrachromosomal circular DNA for dilated cardiomyopathy. *Biol Direct*, 2024, 19: 125
- [29] Yang L, Jia R, Ge T, et al. Extrachromosomal circular DNA: biogenesis, structure, functions and diseases. *Sig Transduct Target Ther*, 2022, 7: 342
- [30] Jones CY, Williams CL, Moreno SP, et al. Hyperextended telomeres promote formation of C-circle DNA in telomerase positive human cells. *J Biol Chem*, 2023, 299: 104665
- [31] Shoura MJ, Gabdank I, Hansen L, et al. Intricate and cell type-specific populations of endogenous circular DNA (eccDNA) in *Caenorhabditis elegans* and *Homo sapiens*. *G3 (Bethesda)*, 2017, 7: 3295-303
- [32] Schröder C, Steimer W. gDNA extraction yield and methylation status of blood samples are affected by long-term storage conditions. *PLoS One*, 2018, 13: e0192414
- [33] Møller HD. Circle-Seq: isolation and sequencing of chromosome-derived circular DNA elements in cells. *Methods Mol Biol*, 2020, 2119: 165-81
- [34] Mehta D, Cornet L, Hirsch-Hoffmann M, et al. Full-length sequencing of circular DNA viruses and extrachromosomal circular DNA using CIDER-Seq. *Nat Protoc*, 2020, 15: 1673-89
- [35] Møller HD, Parsons L, Jørgensen TS, et al. Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci USA*, 2015, 112: 2400
- [36] Gao X, Liu K, Luo S, et al. Comparative analysis of methodologies for detecting extrachromosomal circular DNA. *Nat Commun*, 2024, 15: 9208
- [37] Yu J, Zhang H, Han P, et al. Circle-seq based method for eccDNA synthesis and its application as a canonical promoter independent vector for robust microRNA overexpression. *Comput Struct Biotechnol J*, 2024, 21: 1115-21
- [38] Spier Camposano H, Molin WT, Saski CA. Sequence characterization of eccDNA content in glyphosate sensitive and resistant Palmer amaranth from geographically distant populations. *PLoS One*, 2022, 17: e0260906
- [39] Fu W, MacGregor DR, Comont D, et al. Sequence characterization of extrachromosomal circular DNA content in multiple blackgrass populations. *Genes*, 2023, 14: 1905
- [40] Zhang P, Mbodj A, Soundiramourthy A, et al. Extrachromosomal circular DNA and structural variants highlight genome instability in *Arabidopsis* epigenetic mutants. *Nat Commun*, 2023, 14: 5236
- [41] Kang J, Dai Y, Li J, et al. Investigating cellular heterogeneity at the single-cell level by the flexible and mobile extrachromosomal circular DNA. *Comput Struct Biotechnol J*, 2023, 21: 1115-21
- [42] Lin M, Chen Y, Xia S, et al. Integrative profiling of extrachromosomal circular DNA in placenta and maternal plasma provides insights into the biology of fetal growth restriction and reveals potential biomarkers. *Front Genet*, 2023, 14: 1128082
- [43] Ko I, Kranse OP, Senatori B, et al. A critical appraisal of DNA transfer from plants to parasitic cyst nematodes. *Mol Biol Evol*, 2024, 41: msae030
- [44] Zhuang J, Zhang Y, Zhou C, et al. Dynamics of extrachromosomal circular DNA in rice. *Nat Commun*, 2024, 15: 2413
- [45] Cheng Z, Luo X, Liu W, et al. Comprehensive landscape and oncogenic role of extrachromosomal circular DNA in malignant biliary strictures. *Cell Biosci*, 2025, 15: 16
- [46] Luo X, Zhang L, Cui J, et al. Small extrachromosomal circular DNAs as biomarkers for multi-cancer diagnosis and monitoring. *Clin Transl Med*, 2023, 13: e1393
- [47] Fang M, Fang J, Luo S, et al. eccDNA-pipe: an integrated pipeline for identification, analysis and visualization of extrachromosomal circular DNA from high-throughput sequencing data. *Brief Bioinform*, 2024, 25: bbac034
- [48] Peng Y, Li Y, Zhang W, et al. The characteristics of extrachromosomal circular DNA in patients with end-stage renal disease. *Eur J Med Res*, 2023, 28: 134
- [49] Yang Y, Song T, Liu S, et al. Circle-map profiling of extrachromosomal circular DNA as diagnostic biomarkers for lung cancer. *Precision Clin Med*, 2024, 7: pbae006
- [50] Lv W, Pan X, Han P, et al. Extrachromosomal circular DNA orchestrates genome heterogeneity in urothelial bladder carcinoma. *Theranostics*, 2024, 14: 5102-22