

DOI: 10.13376/j.cblls/2023092

文章编号: 1004-0374(2023)06-0824-09

基于高通量测序技术对染色体外环状DNA研究的最新进展

郭志涛^{1,2}, 武慧慧^{1,2}, 耿淑慧^{1,2}, 陈晨^{1,2}, 孟祥宁^{1,2}, 孙文靖^{1,2}, 朱静^{1,2*}

(1 哈尔滨医科大学医学遗传学研究室, 哈尔滨 150081; 2 中国遗传资源保护与
疾病防控教育部重点实验室(哈尔滨医科大学), 哈尔滨 150081)

摘要: 染色体外环状 DNA (extrachromosomal circular DNA, eccDNA) 是一种真核生物染色体外的闭合环状 DNA 结构, 长度和染色体起源具有较高异质性。eccDNA 这一名称目前主要指大小在数百 kb 以内的小分子染色体外环状 DNA, 包括 microDNA、小多分散环状 DNA (small polydispersed circular DNA, spcDNA) 以及其他未分类的小分子 eccDNA 等。高通量测序技术 (high-throughput sequencing, HTS) 是一种可以同时同时对百万条 DNA 分子进行序列测定的技术, 又名新一代测序技术 (next generation sequencing, NGS), 具有高通量、高灵敏度、高准确度等优势。近年来高通量测序结合生物信息学分析技术不仅在揭示 eccDNA 染色体起源、分子结构、发生机制和潜在功能以及循环系统中的 eccDNA 分子特征研究等方面发挥了重要作用, 而且推动了 eccDNA 在甲基化等表观遗传学方面的研究。生物信息学软件的发展和 eccDNA 分析算法的开发也对其研究提供了重要帮助。血浆以及尿液等液体活检常用液体样本中也鉴定出 eccDNA 的广泛存在, 并且表现出与组织、生理状态以及疾病相关的生物学特征。eccDNA 可能是重要的液体活检候选标志物。本综述旨在阐述利用高通量测序技术在研究 eccDNA 的结构、发生机制和功能上取得的最新成果, 并讨论 eccDNA 在液体活检等方面的潜在应用。

关键词: 染色体外环状 DNA; microDNA; 高通量测序技术; 液体活检

中图分类号: Q343.2

文献标志码: A

Research advances of extrachromosomal circular DNAs (eccDNAs) using next generation sequencing technology

GUO Zhi-Tao^{1,2}, WU Hui-Hui^{1,2}, GENG Shu-Hui^{1,2}, CHEN Chen^{1,2},
MENG Xiang-Ning^{1,2}, SUN Wen-Jing^{1,2}, ZHU Jing^{1,2*}

(1 Laboratory of Medical Genetics, Harbin Medical University, Harbin 150081, China; 2 Key Laboratory of Preservation of Human Genetic Resources and Disease Control in China (Harbin Medical University), Ministry of Education, Harbin 150081, China)

Abstract: Extrachromosomal circular DNA (eccDNA) is a closed circular DNA structure outside the chromosomes of eukaryotes. The sizes and chromosome origins of eccDNA are highly heterogeneous. At present, the name eccDNA mainly refers to small circular DNA sized less than thousands of kilo bases, which includes microDNA, small polydispersed circular DNA (spcDNA) and other unclassified eccDNA molecules. High-throughput sequencing (HTS) also known as next generation sequencing (NGS), is a technology that can simultaneously sequence millions of DNA molecules, with the advantages of high throughput, high sensitivity and high accuracy. In recent years, high-throughput sequencing combined with bioinformatics analysis methods has greatly promoted

收稿日期: 2022-10-17; 修回日期: 2022-12-17

基金项目: 2019年度黑龙江省自然科学基金-优秀青年基金(YQ2019H002); 哈尔滨医科大学少帅揭榜项目(HMUMIF-21007); 2018年黑龙江省留学回国人员择优资助项目(启动类)

*通信作者: E-mail: zhuj@hrbmu.edu.cn; Tel: 0451-86632775

research of the origin, molecular structure, genetic mechanism and potential function of eccDNA, the research on the molecular characteristics of eccDNA in the circulating system, as well as the research on epigenetics such as methylation of eccDNA. The development of bioinformatic software and eccDNA analysis algorithm also provides important help for eccDNA research. EccDNAs have also been widely identified in plasma, urine and other body fluid samples, which are commonly used for liquid biopsy, and show biological characteristics related to tissue, physiological status and diseases. EccDNA may become important biomarkers for liquid biopsy. This review aims to discuss the latest research progress of the structure, mechanism and function of eccDNAs by using high-throughput sequencing technologies and the potential clinical application of eccDNAs in liquid biopsy.

Key words: extrachromosomal circular DNAs (eccDNAs); microDNA; high-throughput sequencing (HTS); liquid biopsy

染色体外环状 DNA (extrachromosomal circular DNAs, eccDNAs) 是指位于染色体外的单链或双链闭合环状 DNA 结构。广义上来说, eccDNA 包括了所有染色体外环状 DNA 分子。而实际上, 在目前很多研究报道中, eccDNAs 这一名称指的是一大类小分子染色体外环状 DNA, 包括 microDNA^[1]、小多分散环状 DNA (small polydispersed circular DNA, spcDNA)、端粒环 (T-circle) 和其他未分类小环状 DNA 分子^[2]。而一些大分子环状 DNA, 比如双微体 (double minute chromosomes, DMs) (近年有研究者使用 extrachromosomal DNA, ecDNA 这一名称), 这种与肿瘤中基因扩增密切相关的大分子环状 DNA, 一般用自己的专用名称。此外, 本文围绕的是真核生物中的起源于自身基因组且非细胞器的环状 DNA, 所以并没有讨论真核生物的线粒体 DNA、原核生物的环状 DNA 以及对于宿主基因组来说是外源性的质粒或病毒 DNA 等环状 DNA。

高通量测序技术 (high-throughput sequencing, HTS) 与传统 Sanger 测序技术相比, 它不仅通量高、灵敏度高, 而且能同时对几十万到几百万条 DNA 分子进行序列测定, 也被命名为新一代测序技术 (next generation sequencing, NGS)。近年来, 随着新一代测序技术的不断成熟, 通过二代测序、三代测序等测序技术实现全基因组测序 (whole genome sequencing, WGS) 等 DNA 水平的测序, 以及对 mRNA、非编码 RNA 等转录本进行全转录组测序 (whole transcriptome sequencing), 人类得以更全面、更精细地观察和分析遗传物质的结构和功能。生物信息学分析方法、软件和数据库的发展以及高性能计算平台的搭建也对高通量测序技术起到重要的支撑作用。本文主要阐述新一代测序技术及其衍生技术在 eccDNAs 分子结构、特征和功能研究中所取得的最新进展, 也为这些技术在其他基因组学和表观基因

组学研究中的应用提供思路。

1 eccDNA的研究进程

早在 1965 年, Hotta 等^[3]就在高等植物和动物细胞中发现染色体外环状 DNA 的存在。随后, 在 HeLa 细胞中, 研究人员也发现了大量大小不等的闭合环状 DNA^[4]。1972 年, Smith 等^[5]研究了环状 DNA, 发现它们的大小在 0.2 μm 到 2.0 μm 之间, 并将其命名为 spcDNA。早期主要是利用电子显微镜、电泳和杂交技术等方法对 eccDNA 的结构和特征进行研究, 在植物、酵母、小鼠和人等样本中证明了 eccDNA 的广泛存在。

随着高通量测序技术的不断进步, eccDNA 的相关研究也取得了重要的突破。2012 年, Dutta 研究团队^[1]利用二代测序技术首次在小鼠组织以及小鼠和人细胞系中发现一类长度相对较小的染色体外环状 DNA, 并且并非来自染色体重复序列, 基于其长度和序列特征将其命名为 microDNA。随后, 此类 eccDNA 的起源、机制和功能研究逐步展开, 而高通量测序技术在其中发挥非常重要的作用^[6-11]。近年来, 研究人员也将 eccDNA 的研究拓展到血液和其他体液样本中, 在外周血 (血浆、血清)、尿液等样本中鉴定了 eccDNA 的存在, 并研究了其在肿瘤、慢性肾病和产前诊断等相关的液体活检中的潜在应用^[12-16]。

科研方法的确立和技术的进步对研究不同的 eccDNA 非常重要。早期研究报道的 spcDNA 使用的技术多为电泳和杂交技术, 并提示 spcDNA 的分子特征表现为主要起源于重复序列, 发生机制上则可能包括同源重组、非同源重组和逆转录转座等。近年来, 随着 eccDNA 富集方法的不断完善以及各种高通量测序技术的不断进步, 目前研究 eccDNA 多以高通量测序技术为主, 并且针对 eccDNA 不同

的研究目的,建立了特异的高通量研究方法。不同的研究方法是我们能否攻克 eccDNA 相关问题的关键,因此是我们讨论 eccDNA 的分子特征的重要基础。

2 新一代测序技术助力eccDNA结构与序列特征的揭示

目前在 eccDNA 分子结构的研究中使用较多的高通量测序技术是二代测序技术,主要是采用 Illumina 测序平台的双端测序方法 (paired-end sequencing)。microDNA 以及与其富集和分析方法相类似的一些未分类 eccDNA 是运用此方法的重要代表。对于 microDNA 分子结构特征的研究一般包括对其分子大小、数量、基因组起源、GC 含量分析等方面。

首先发现和定义 microDNA 的是 Dutta 的研究团队。他们对 microDNA 在长度方面的描述为:大多在 1 000 bp 以下,在 200 bp 和 400 bp 有两个主要峰。同时,也观察到其他长度分布峰的存在,峰值周期为 150 bp 左右^[1,6]。此后,该团队比较正常组织和肺癌组织中的 microDNA,发现肿瘤样本中鉴定的 microDNA 的序列长度大多比匹配的正常组织中的 microDNA 更长,相应地,肿瘤手术切除前循环系统中的 eccDNA 的长度要长于手术切除后^[15]。此外, Mehanna 等^[17]用化疗药物氨甲喋呤 (methotrexate, MTX) 或者门冬酰胺酶 (L-asparaginase, ASP) 处理淋巴瘤细胞系 (lymphoblastoid cell lines, LCLs),发现药物处理前 microDNA 在 200 bp 和 400 bp 处存在峰值,平均大小在 379 bp;而经过 MTX 或 ASP 处理后, microDNA 的量更多,片段长度变长,并且这种长度改变趋势在敏感细胞系中的结果比天然耐药细胞系中更明显。同时,他们也观察到 microDNA 大小分布出现了间隔约为 190 bp 的周期性峰值,该周期比较符合细胞凋亡产生的 DNA 片段化 (180~200 bp 的整倍数),提示细胞凋亡可能是部分 microDNA 的产生机制。在另一则研究中, Sin 等^[13]证明孕妇血浆 eccDNA 的长度分布的峰值在 202 bp 和 338 bp 处,以 338 bp 峰为主,并且两个主峰上都有 10 bp 的周期性小峰;而血浆中游离的线性 DNA 长度分布的主峰约为 166 bp,因此游离的 eccDNA 长于游离的线性 DNA。此外,该团队也证明孕妇血浆中胎儿来源的游离 eccDNA 的存在,并且胎儿来源的 eccDNA 长度要短于母体来源的 eccDNA。Pang 等^[18]对 4 例痛风性关节炎

和健康个体的血浆 eccDNA 进行分析,提示健康个体血浆 eccDNA 的大小比痛风性关节炎患者更长 (针对 2 kb 以下的 eccDNA 进行的分析)。Lv 等^[16]近期研究发现晚期慢性肾病患者的尿液 eccDNA 比健康人对照在数量上更多,大小上更长。综上所述, eccDNA 分子大小是 eccDNA 分子特征研究的重要方面,表现出与组织来源、生理或病理状态以及药物作用的相关性,有潜在的诊断和液体活检等方面的应用价值。

关于 eccDNA 的染色体起源也是 eccDNA 分析的重要方面。端粒环 (T-circle) 包含端粒重复序列,因为其自身序列和功能的特点,不在此讨论; spcDNA 的研究较为早期,采用的实验方法多为电泳和杂交,证明其主要来源于重复序列。在此我们主要关注和讨论 microDNA 以及与其有类似特征的未分类 eccDNA 的染色体起源。根据高通量测序结果的分析, microDNA 可以起源于整个基因组,但其起源并非随机,而是具有一定倾向性,更多起源于基因区域,特别是外显子区、5' 非翻译区 (5'untranslated regions, 5'UTR) 或 3' 非翻译区 (3'untranslated regions, 3'UTR) 以及 CpG 岛 (CpG island) 等^[1,6,13,16]。比起整个基因组或者其上下游区域 DNA, eccDNA 有着更高的 GC 含量^[1,6,14];并且比较同组中的不同长度的 eccDNA 发现,更小分子的 eccDNA (<500 bp) 比大分子的 eccDNA (>500 bp) 具有更高的 GC 含量^[14]。此外,标志着高转录活性、高开放程度或者有调节功能的特征性位置,比如核酸内切酶敏感位点 (DNase hypersensitivity sites, DHS)、H3K27Ac 等标志也在 eccDNA 上富集^[14,17]。这些结果说明, eccDNA 的起源并非随机,而是更易于起源于基因区、活性功能区和调控区, eccDNA 可能发挥着调节基因表达等功能。

除了二代测序,单分子荧光测序 (代表性技术如美国螺旋生物的 SMS 技术和美国太平洋生物的 SMRT 技术) 和纳米孔测序等三代测序技术也被应用到 eccDNA 的结构和特征研究。这种测序技术在 DNA 测序时不经过 PCR 扩增过程,对每一条 DNA 分子进行单独测序,具有精确度高、读长长、测序速度快等优势。Wang 等^[11]对分离纯化后的 eccDNA 进行滚环复制 (rolling cycle amplification, RCA),对其产物进行处理后进行纳米孔测序,并对 eccDNA 的分子结构和产生机制以及免疫刺激功能进行了深入研究。从 eccDNA 长度方面来说,他们观察到了鼠胚胎干细胞中 eccDNAs 长度中位数约

为 1 kb, 并且也呈现出平均间隔为 188 bp 的多个峰值分布的特点^[11]。与 Mehanna 等^[17]的观点一致, 他们也认为 eccDNA 可能与细胞凋亡密切相关, 并进行了 eccDNA 发生机制的研究^[11]。

3 eccDNA 的发生机制

不同类别的 eccDNA 可能有不同的产生机制, 而同一种 eccDNA 的产生也可能有多种机制共同参与。早期对于 spcDNA 的研究认为, spcDNA 主要来源于基因组中的重复序列如 Alu 家族序列, 同源重组 (homologous recombination, HR) 可能是 spcDNA 的主要介导机制。但同时非同源末端连接 (non-homologous end joining, NHEJ) 也被证明参与小鼠主卫星 DNA (major satellite DNA) 序列来源的 spcDNA 的产生^[19]。在此, 我们依然是围绕基于高通量测序技术的 eccDNA 研究, 重点介绍 microDNA 等 eccDNA 的发生机制。

目前关于 microDNA 的形成途径尚未完全阐明, 而对高通量测序结果的分析则提供了揭示其分子机制的线索。总体来说, microDNA 的产生包括染色体外 DNA 片段的产生和 DNA 片段的异常修复/环化两大步骤 (图 1)。对于 DNA 片段的产生, 目前认为有可能是 DNA 双链断裂 (double strand breaks, DSBs) 损伤^[1, 6, 10, 14]、复制压力 (replication stress)^[6]、过度转录 (hypertranscription)^[20]、基因组重排时的染色体碎裂^[21] 以及早期细胞凋亡驱动的 DNA 碎片化^[11, 17] 等。其中 DSB 有助于 eccDNA 生成的观点在 2018 年得到了 Moller 等^[22] 利用 CRISPER-C 进行

的实验验证, 他们发现细胞中 DSB 产生的碎片 DNA 以末端连接的方式形成 eccDNA, 其大小在几百 bp 到几百 kb 不等。该实验也为后续人工合成 eccDNA 提供了一种思路。对于 DNA 片段的异常修复/环化过程, 主要利用的是 NHEJ、微同源介导末端连接 (microhomology mediates end joining, MMEJ) 等重要修复通路^[1, 14, 16]。有趣的是, 近期 Paulsen 等^[10] 发现 microDNA 产生的数量依赖于 DSB 和末端降解, 其主要通过 MMEJ 途径修复; 而不进行末端降解的经典非同源末端连接 (canonical non-homologous end joining, C-NHEJ) 通路通过修复 DSB 来抑制 microDNA 的产生。该项研究提示 C-NHEJ 途径可能是抑制疾病相关 microDNA 产生的靶通路。

除了 DNA 断裂和修复机制, 其他机制也参与了 microDNA 的产生: 比如 Dillon 等^[6] 分析 microDNA 起始位点两侧有 2~15 bp 的短直接重复序列, 可能是复制滑移 (replication slippage) 的位点, 可以被错配修复 (mismatch repair, MMR) 途径切除。MMR 途径的重要分子 MSH3 的突变显著减少了 microDNA 的丰度, 并改变了 microDNA 的基因组位点的分布。因此, 该项研究认为, DNA 损伤和修复、复制滑移 -MMR 途径等多个机制均可介导 microDNA 的产生。另外, 根据对 eccDNA 分子大小特征和凋亡诱导后的观察, Wang 等^[11] 认为凋亡相关 DNA 片段化以及后续的 DNA 连接酶 3 (DNA ligase 3) 介导的 DNA 连接过程是 eccDNA 产生的重要机制。综上所述, microDNA 等 eccDNA 的发生机制复杂, 在不同背景下, DNA 断裂和修复、DNA 复制以及细

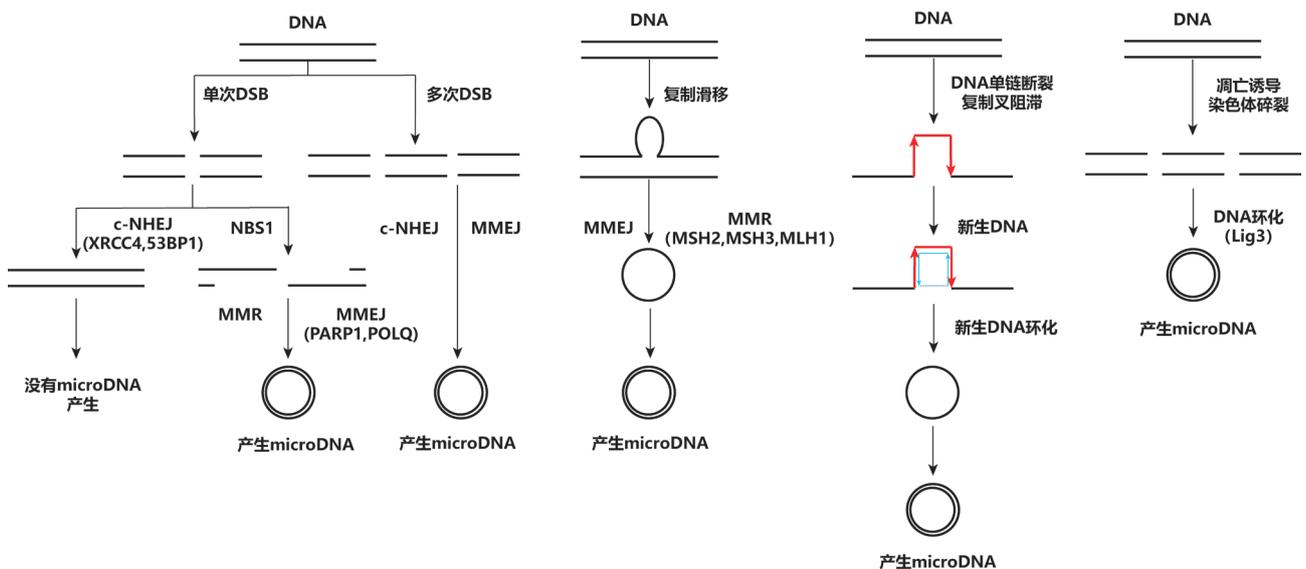


图1 eccDNA 产生机制

胞凋亡 DNA 片段化和连接等多种机制均可能参与 eccDNA 的产生。

4 eccDNA的功能

研究 eccDNA 的功能主要从其分子特征的分析、不同生理病理条件下的丰度、是否有 mRNA 转录本、是否表达小的调节性 RNA 等几个方面考量。首先, microDNA 虽然可以起源于整条染色体,但是它的起源具有非随机性和倾向性,其易起源于基因区、高 GC 含量区、CpG 岛区、UTR 区等,提示其可能具有一定的功能,而不仅仅是核酸降解的副产物。其次,肿瘤细胞中的 spcDNA 具有与调控基因重排和肿瘤基因组不稳定有关的逆转座子成分和端粒序列^[23]。更有研究表明, eccDNA 的长度、丰度与细胞或组织特异性、基因组稳定性、病理状态、是否化疗敏感等情况相关^[15-17, 24-26],提示其发挥一定的生物学功能。

虽然 microDNA 大小多在 1 kb 以内,但也有比较大的 eccDNA 分子(几个 kb 或者几十 kb)有可能表达完整基因或者部分截短基因,从而发挥功能、影响表型。eccDNA 的异质性很高,鉴定 eccDNA 表达的转录本并不容易。得益于高测序深度的转录组测序技术, Moller 等^[8]通过对人肌肉组织样本的转录组测序证明一些 mRNA 序列可以完全匹配 microDNA 的连接序列。他们从 8 个肌肉样本中鉴定出了 25 种不同的 microDNA 转录本,说明 microDNA 能够通过表达 mRNA 来影响细胞表型,并表达 mRNA 产物,从而可能影响细胞表型。

另外, Paulsen 等^[9]通过构建人工 microDNA 分子研究其功能特性,发现 microDNA 能够表达功能性小调节 RNA 如 microRNA 和 si-like RNA,靶向 microDNA 来源亲本基因,从而抑制基因的表达。同时发现 microDNA 的转录不依赖标准启动子序列,内源性 microDNA 能与 RNA 聚合酶亚基 POLR2H 和 POLR3F 结合,表明 microDNA 可能被 RNA 聚合酶 III 转录,也可能被 RNA 聚合酶 I 和 II 转录。总之, microDNA 可能通过产生已知的和新的调节 RNA 来调节基因表达。

值得注意的是, Wang 等^[11]的研究还证明了 eccDNA 能够刺激免疫细胞做出免疫应答。他们用与 eccDNA 大小相同的基因组线性 DNA 与提取的 eccDNA 同时转染巨噬细胞,检测其免疫因子,发现 eccDNA 能够有效增加免疫因子释放。进一步研究表明,这种刺激与 cGAS-STING 通路有关,之后

他们将 eccDNA 线性化,发现其刺激作用消失,表明此刺激作用不是依赖于 eccDNA 序列,而是因为 eccDNA 的环状特性和胞质 DNA 传感器 STING 通路。此结果给肿瘤的免疫治疗提供了新的思路。

5 eccDNA在液体活检中的潜在应用

多项研究证明了 eccDNA 在循环系统中的存在及其特征。Zhu 等^[4]鉴定了人外周血血浆中的游离 eccDNA,发现它们更倾向于起源于染色体上的外显子区、DNA 酶超敏区、组蛋白 H3K27Ac 标志物富集区等位置。Kumar 等^[15]在人和小鼠的血浆和血清中证明了 microDNA 的存在。此外,肿瘤和正常细胞都可以释放 microDNA 到循环系统中,而肿瘤来源的 microDNA 比正常细胞来源的片段长度更长。2022 年, Wu 等^[27]研究肺腺癌患者血浆 eccDNA 后发现,部分 eccDNA 可作为早期诊断肺腺癌的潜在生物标志。Pang 等^[18]对痛风性关节炎患者及其对照人群的血浆 eccDNA 进行了富集和测序分析,发现了可能与痛风性关节炎特异性相关的 eccDNA 分子,提示血浆 eccDNA 在该病中可能的生物学功能以及作为标志物进行临床应用。

尿液也是液体活检的样本之一。Lv 等^[16]研究了人体尿液中小分子 eccDNA 的相关特征。他们发现尿液中 eccDNA 的长度主要在 1 000 bp 以内,GC 含量高,易起源于基因区、短散布核元件(short interspersed nuclear elements, SINE)等区域,晚期慢性肾病患者尿液中有更多的 eccDNA,并且富集了更多 miRNA 基因,提示 eccDNA 是尿液活检的潜在标志物。

在产前诊断领域,2020 年, Sin 等^[13]在研究母体外周血胎儿 eccDNA 时发现,胎儿来源的 eccDNA 分子通常比母体来源的短。2021 年,他们接着研究了母体和胎儿 eccDNA 的甲基化水平,通过新的建库和测序流程对其进行序列和甲基化分析,发现胎儿 eccDNA 的甲基化水平低于母体的 eccDNA;较小的 eccDNA 的甲基化水平低于较大的 eccDNA^[12]。eccDNA 的甲基化水平与其片段大小呈正相关。而母胎起源的线性 DNA 和 eccDNA 之间的甲基化水平相当。这一研究开拓了 eccDNA 表观遗传学领域的研究,为疾病相关异常甲基化的液体活检方面的应用提供了重要思路。

液体活检是通过分析循环系统中的细胞、核酸和蛋白质等组分来实现癌症等疾病的早期筛查、分子分型、预后、用药指导以及复发监测等临床应用。

血浆、血清、尿液等循环系统中存在大量 eccDNA, 并且与其细胞来源有特异相关性。经过更多的研究和开发, 循环系统中游离的 eccDNA 可能成为液体活检的潜在分子标志物^[28]。

6 eccDNA研究策略总结

6.1 eccDNA的主要研究策略

对于小分子的 eccDNAs, 如 microDNAs, 因为其异质性非常高, 而且并不像双微体那样呈现高扩增, 所以很难从普通全基因组 DNA 测序中获得其相关信息。而目前采取的富集和分析方法主要步骤是: (1) 总 DNA 的提取。提取出来的 DNA 中包括线性 DNA 和环状 DNA。(2) DNA 的消化。利用核酸外切酶 VII、plasmid-safe ATP-dependent DNase 等消化去除线性 DNA, 利用蛋白酶和 RNA 酶消化去除蛋白质和 RNA。(3) 经过滚环扩增 (rolling circle amplification, RCA) 或者多重置换扩增 (multiple displacement amplification, MDA) 将小分子环状 DNA 优先富集扩增出来。(4) 通过物理或化学方法将 DNA 制成适合的小片段, 进行 DNA 建库和测序^[1, 6, 14](图 2)。也有报道不经过扩增, 只是进行非常充分的线性 DNA 的消化, 并对获得的 DNA 利用 Tn5 酶等方法进行微量建库、测序和分析^[13, 26, 29], 这样可能具有更好的实验重复性^[26]。而 Wang 等^[11]则对比了 RCA 扩增和不经过 RCA 扩增直接进行转座酶建库测序的实验结果, 并排除了 RCA 扩增过程产生的偏差。2020 年, Kumar 等^[30]依据 ATAC-seq (Assay

for Transposase-Accessible Chromatin with high throughput sequencing) 设计了识别 eccDNA 的新策略 (图 3), 他们不进行消化在裂解细胞后直接进行 ATAC-seq 建库, 利用 eccDNA 特殊的连接位点应用生信手段提取 eccDNA 信息。

6.2 主要eccDNA生物信息提取工具

随着对 eccDNA 研究的不断深入, 科研工作者开发出了多个用于 eccDNA 信息提取的工具。目前已发表的用于 eccDNA 分子信息提取的工具具有 circle-Map^[31]、ECCsplorer^[32]、ecc_finder^[33] 等。其中 circle-seq、ECCsplorer 是基于 Illumina 短读测序的工具, CIDER-seq^[34] 运用于长读测序的数据, ecc_finder 则主要对 Illumina 长读以及纳米孔测序数据进行分析 (表 1)。最常用的分析工具是 circle-seq 工具, 其基于 eccDNA 的 map reads 中包含 split reads 和 Discordant reads 两个特殊 reads 而进行 eccDNA 识别, 它能够将包含这两种特殊结构的 reads 进行提取, 并对其进行剪切拼接, 从而得到完整的 eccDNA 信息。不过, 实验发现当 split reads 数过低时, 假阳性较高, 因此可以选择 split reads 数较高的数据作为 eccDNA 进行分析^[35]。

6.3 eccDNA甲基化研究策略

随着二代测序衍生技术的飞速发展, 为了研究 eccDNA 的甲基化, 2021 年 Sin 团队介绍了一种新的适合研究甲基化的建库方法^[12]。酶切去除线性 DNA 后, 利用 Tn5 转座酶打开环状 DNA 的环状结构并同时 DNA 片段两端加上接头序列, 并用酶

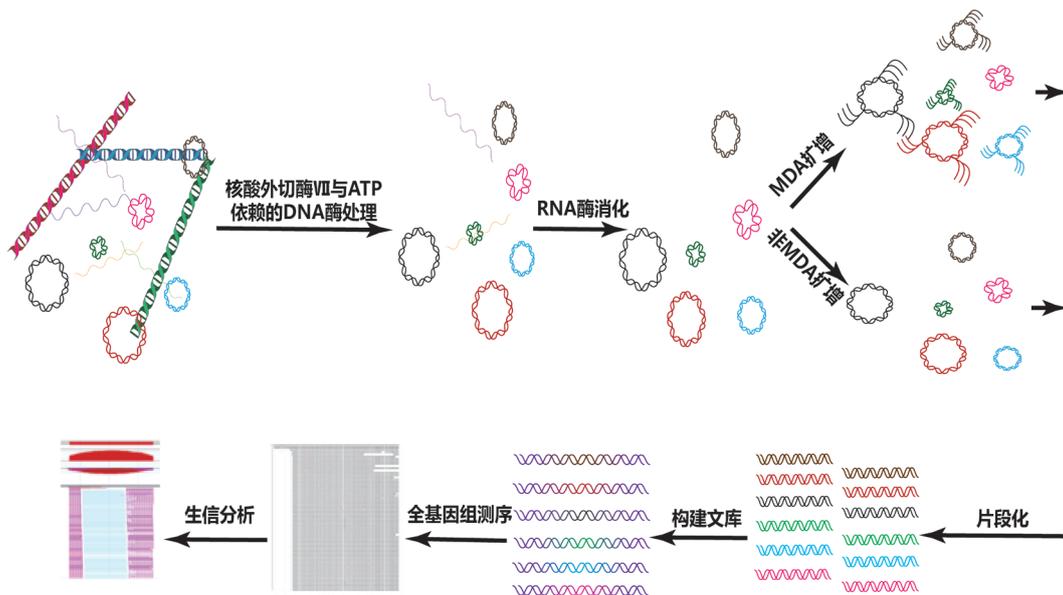


图2 小分子eccDNA的富集和测序研究策略

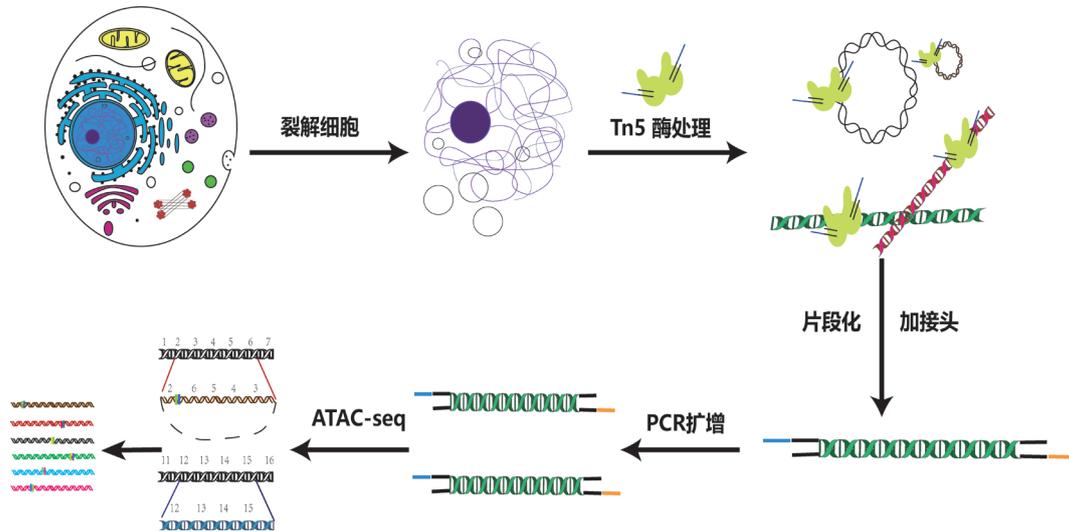


图3 ATAC-seq识别eccDNA

表1 eccDNA分子提取的生物信息学算法

算法名称	下载地址	参考文献
circle-Map	https://github.com/iprada/Circle-Map	[36]
ECCsplorer	https://github.com/crimBubble/ECCsplorer	[32]
ecc_finder	https://github.com/njaupan/ecc_finder	[33]
CIDER-seq	https://github.com/devang-mehta/ciderseq2	[34]

转化法将未甲基化的 C 转化为 U, PCR 复制将 U 转化为 T, 进行建库及测序。单次建库测序能够同时检测样品中环状 DNA 及其甲基化位点的信息 (图 4)。

7 问题和展望

尽管现在 eccDNAs 的分离和提取方法、高通量测序和分析等都取得了很大的进步, 我们对 eccDNA 有了一定了解, 但这是远远不够的。在其结构与功能方面还有许多亟待解决的问题。如在结构和发生机制上: (1) 驱动 eccDNAs 形成和维持的潜在机制是哪些? (2) eccDNA 能否以及在什么位点重新整合到染色体中? 在功能上: (1) eccDNA 的异质性非常高, 如何从中找到疾病特异性相关的分子, 并深入研究其功能? (2) eccDNA 究竟能否调控相关基因的表达及其机制是什么? (3) 外周血中的 eccDNA、外泌体 eccDNA 是否能进入细胞并发挥生物学功能?

eccDNA 表现出与组织、个体、生理或病理状态等的相关性, 是潜在的临床诊断标志物。eccDNA 的环状结构使其能对抗核酸酶的降解, 比线性 DNA 更稳定, 并且在外周血循环系统中普遍存在,

可能是液体活检的候选标志物。并且 eccDNA 表现出起源的非随机, 富集活性调控序列可能通过产生 miRNA 等小分子发挥调控作用, 也可以刺激免疫系统产生特异反应。综上, 我们有以下展望: (1) 进一步应用最新的三代测序技术, 助力于揭示 eccDNAs 更精确的结构和发生机制; (2) 表观遗传组学技术在 eccDNA 的表观学研究上进行更多的应用, 揭示 eccDNA 的基因调控功能和生物学功能; (3) 在更多临床样本中检测 eccDNA 与疾病和产前状态的相关性, 开发特异性的 eccDNA 分子或 eccDNA 种类检测方法, 应用于疾病的诊断和产前诊断, 尤其是无创性液体活检领域。技术的发展常常是揭开科学神秘面纱的关键。新一代测序技术为 eccDNA 的结构、功能和应用的研究开启了新篇章, 相信不久将取得更多有价值的研究成果。

[参 考 文 献]

- [1] Shibata Y, Kumar P, Layer R, et al. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science*, 2012, 336: 82-6
- [2] Wang T, Zhang H, Zhou Y, et al. Extrachromosomal circular DNA: a new potential role in cancer progression. *J Transl Med*, 2021, 19: 257

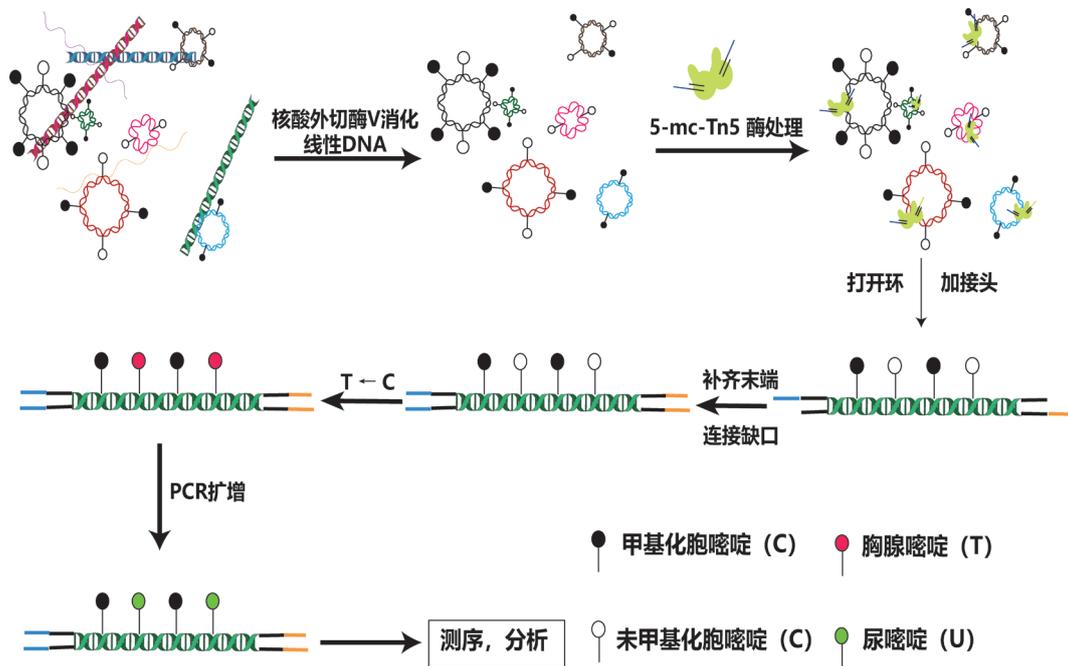


图4 eccDNA甲基化研究策略

- [3] Hotta Y, Bassel A. Molecular size and circularity of DNA in cells of mammals and higher plants. *Proc Natl Acad Sci U S A*, 1965, 53: 356-62
- [4] Radloff R, Bauer W, Vinograd J. A dye-buoyant ant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. *Proc Natl Acad Sci U S A*, 1967, 57: 1514-21
- [5] Smith C, Vinograd J. Small polydisperse circular DNA of HeLa cells. *J Mol Biol*, 1972, 69: 163-78
- [6] 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
- [7] Reon BJ, Dutta A. Biological processes discovered by high-throughput sequencing. *Am J Pathol*, 2016, 186: 722-32
- [8] Moller HD, Mohiyuddin M, Prada LI, et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat Commun*, 2018, 9: 1069
- [9] 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
- [10] Paulsen T, Malapati P, Shibata Y, et al. MicroDNA levels are dependent on MMEJ, repressed by c-NHEJ pathway, and stimulated by DNA damage. *Nucleic Acids Res*, 2021, 49: 11787-99
- [11] Wang Y, Wang M, Djekidel MN, et al. EccDNAs are apoptotic products with high innate immunostimulatory activity. *Nature*, 2021, 599: 308-14
- [12] Sin STK, Ji L, Deng J, et al. Characteristics of fetal extrachromosomal circular DNA in maternal plasma: methylation status and clearance. *Clin Chem*, 2021, 67: 788-96
- [13] Sin STK, Jiang P, Deng J, et al. Identification and characterization of extrachromosomal circular DNA in maternal plasma. *Proc Natl Acad Sci U S A*, 2020, 117: 1658-65
- [14] Zhu J, Zhang F, Du M, et al. Molecular characterization of cell-free eccDNAs in human plasma. *Sci Rep*, 2017, 7: 10968
- [15] 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
- [16] 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: e817
- [17] Mehanna P, Gagne V, Lajoie M, et al. Characterization of the microDNA through the response to chemotherapeutics in lymphoblastoid cell lines. *PLoS One*, 2017, 12: e0184365
- [18] Pang J, Pan X, Lin L, et al. Characterization of plasma extrachromosomal circular DNA in gouty arthritis. *Front Genet*, 2022, 13: 859513
- [19] Cohen Z, Bacharach E, Lavi S. Mouse major satellite DNA is prone to eccDNA formation via DNA ligase IV-dependent pathway. *Oncogene*, 2006, 25: 4515-24
- [20] Hull RM, King M, Pizza G, et al. Transcription-induced formation of extrachromosomal DNA during yeast ageing. *PLoS Biol*, 2019, 17: e3000471
- [21] Yerlici VT, Lu MW, Hoge CR, et al. Programmed genome rearrangements in *Oxytricha* produce transcriptionally active extrachromosomal circular DNA. *Nucleic Acids*

- Res, 2019, 47: 9741-60
- [22] Moller HD, Lin L, Xiang X, et al. CRISPR-C: circularization of genes and chromosome by CRISPR in human cells. *Nucleic Acids Res*, 2018, 46: e131
- [23] Schmidt H, Taubert H, Lange H, et al. Small polydispersed circular DNA contains strains of mobile genetic elements and occurs more frequently in permanent cell lines of malignant tumors than in normal lymphocytes. *Oncol Rep*, 2009, 22: 393-400
- [24] Cohen S, Regev A, Lavi S. Small polydispersed circular DNA (spcDNA) in human cells association with genomic instability. *Oncogene*, 1997, 14: 977-85
- [25] Cohen S, Lavi S. Induction of circles of heterogeneous sizes in carcinogen-treated cells two-dimensional gel analysis of circular DNA molecules. *Mol Cell Biol*, 1996, 16: 2002-14
- [26] 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
- [27] Wu X, Li P, Yimiti M. Identification and characterization of extrachromosomal circular DNA in plasma of lung adenocarcinoma patients. *Int J Gen Med*, 2022, 15: 4781-91
- [28] Zhu J, Chen S, Zhang F, et al. Cell-free eccDNAs: a new type of nucleic acid component for liquid biopsy? *Mol Diagn Ther*, 2018, 22: 515-22
- [29] Su Z, Saha S, Paulsen T, et al. ATAC-Seq-based identification of extrachromosomal circular DNA in mammalian cells and its validation using inverse PCR and FISH. *Bio Protoc*, 2021, 11: e4003
- [30] Kumar P, Kiran S, Saha S, et al. ATAC-seq identifies thousands of extrachromosomal circular DNA in cancer and cell lines. *Sci Adv*, 2020, 6: eaba2489
- [31] Moller HD, Parsons L, Jorgensen TS, et al. Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci U S A*, 2015, 112: E3114-22
- [32] Mann L, Seibt KM, Weber B, et al. ECCsplorer: a pipeline to detect extrachromosomal circular DNA (eccDNA) from next-generation sequencing data. *BMC Bioinformatics*, 2022, 23: 40
- [33] 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
- [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] Prada-Luengo I, Krogh A, Maretty L, et al. Sensitive detection of circular DNAs at single-nucleotide resolution using guided realignment of partially aligned reads. *BMC Bioinformatics*, 2019, 20: 663
- [36] Moller HD. Circle-seq: isolation and sequencing of chromosome-derived circular DNA elements in cells. *Methods Mol Biol*, 2020, 2119: 165-81