DOI:10.12113/202104013

环形 RNA 全长组装与定量工具的研究进展

谢宏文[1],吴　静[1,2],宋晓峰[1]*
(1.南京航空航天大学 自动化学院,南京 211106;2.南京医科大学 生物医学工程与信息学院,南京 211166)

摘　要:环形 RNA 是一种广泛存在于真核细胞的内源性 RNA,由前体 RNA 反向剪接而成,不具有 5'末端帽子和 3'末端 poly(A)尾巴,呈封闭环状结构。环形 RNA 通过 miRNA 海绵结合等方式参与基因表达调控等许多重要的生物学过程。环形 RNA 可以通过可变剪接产生不同的环形 RNA 转录本,因此获取环形 RNA 转录本内部全长序列信息以及对环形 RNA 内部可变剪接产物进行精确定量是揭示环形 RNA 调控功能的前提。生物信息学工具能够高效便捷的处理高通量测序数据,被普遍用来鉴别和分析环形 RNA。本文介绍了环形 RNA 的产生机制以及功能特性,对环形 RNA 检测、全长序列组装以及定量相关计算工具进行综述。

关键词:环形 RNA;生物信息学计算工具;可变剪接;高通量测序
中图分类号:R318.04　　**文献标志码**:A　　**文章编号**:1672-5565(2022)03-155-08

Review on full-length assembly and quantification tools of circular RNA

XIE Hongwen[1], WU Jing[1,2], SONG Xiaofeng[1]*
(1. *College of Automation Engineering*, *Nanjing University of Aeronautics and Astronautics*, *Nanjing* 211106, *China*;
2. *School of Biomedical Engineering and Informatics*, *Nanjing Medical University*, *Nanjing* 211166, *China*)

Abstract: Circular RNA(circRNA) is a type of endogenous RNA that widely exists in eukaryotic cells. It is formed by back-splicing of pre-mRNA without 5' cap and 3' poly(A) tail, showing a closed circular structure. CircRNA are involved in many important biological processes such as gene expression regulation through miRNA sponge binding. CircRNA can produce different circular transcripts through alternative splicing. Therefore, obtaining the full-length sequence information of circRNA transcripts and accurately quantifying the internal alternative splicing products of circRNA are the premise to reveal the regulatory function of circRNA. Bioinformatics tools can handle high-throughput sequencing data efficiently and conveniently, and are widely used to identify and analyze circRNA. This paper introduces the generation mechanism and functional characteristics of circRNA, and reviews the computational tools for circRNA detection, full-length sequence assembly and quantification.

Keywords: CircRNA; Bioinformatics computational tools; Alternative splicing; High-throughput sequencing

环形 RNA 是一种特殊的内源性非编码 RNA,其由前体 RNA(Pre-mRNA)经过反向剪接形成,呈封闭环状结构,不具有 5'末端帽子和 3'末端 poly(A)尾巴。研究发现环形 RNA 参与许多重要的生物学过程,如作为 miRNA 海绵体,参与基因调控,编码蛋白等。环形 RNA 的反向剪接位点是鉴别和定量环形 RNA 的关键,而环形 RNA 可以通过可变剪接产生不同的环形转录本,这些转录本的全长序列信息以及精确定量对于进一步研究环形 RNA 功能具有重要作用。生物信息学方法因其能够高效便捷的处理高通量 RNA-seq 数据,被普遍用来鉴别和分析环形 RNA。本文介绍了真核生物环形 RNA 的产生及功能,对环形 RNA 检测以及全长组装和定量方面的研究工具进行了综述。

收稿日期:2021-04-19;**修回日期**:2021-05-25.
基金项目:国家自然科学基金项目(No.61973155,61901225).
作者简介:谢宏文,男,硕士研究生,研究方向:生物信息学.E-mail: hongwen@nuaa.edu.cn.
*通信作者:宋晓峰,男,教授,研究方向:生物信息学.E-mail: xfsong@nuaa.edu.cn.

1 环形 RNA 的产生及其功能

1.1 环形 RNA 的产生

Sanger 等在 20 世纪 70 年代发现某些高等植物中存在可致病的单链环状类病毒,这是人类首次发现环形 RNA[1]。但在过去的几十年中,人们把环形 RNA 看作剪接副产物,环形 RNA 的研究进展十分缓慢。近年来,随着第二代高通量测序技术的出现及环形 RNA 分子纯化方法的运用和发展,人们可以重新认识和研究环形 RNA。

环形 RNA 不具有 5’末端帽子和 3’末端 poly(A)尾巴,是以反向剪接的方式形成的一种共价环状结构。环形 RNA 主要来源于蛋白编码基因,大多由其中的外显子衍生而来并积累在细胞质中,少部分为包含外显子和内含子序列的环形 RNA,以及仅来自蛋白编码基因内含子区域的环形 RNA[2]。环形 RNA 的产生机制可分为内含子环化和外显子环化。内含子环化发生在外显子剪接过程中,该过程会产生内含子套索结构的中间产物,这些套索在剪接完成后仍然存在,形成内含子环形 RNA(见图 1a),如 ci-ankrd52 由内含子套索结构形成,可以与 RNA 合成酶Ⅱ结合促进其转录作用[3];外显子环化又可分为三种,即内含子配对驱动环化、套索驱动环化和 RNA 结合蛋白驱动环化[4]。内含子配对驱动环化将环化外显子的侧翼内含子互补配对,使环化外显子的剪接供体和剪接受体位置更接近,形成环状结构,然后将环状结构中的内含子切除后形成环形 RNA(见图 1b),如秀丽隐杆线虫侧翼内含子上反向互补重复序列可以使转录本形成发夹结构,促进外显子环化[5];套索驱动环化发生在外显子跳跃过程中,该过程可以使下游 5’剪接供体与上游 3’剪接受体结合,形成包含外显子的套索结构,再通过套索内的剪接切除内含子,形成环形 RNA(见图 1c),如 Barrett 对酵母 mrps16 基因进行质粒表达,通过删除剪接位点证明了包含外显子的套索结构对于酵母细胞中环形 RNA 的生成是必要的[6]; RNA 结合蛋白可以结合到环化外显子两侧的内含子的结合位点上,拉近两端的内含子促进环化,形成环形 RNA(见图 1d),如可变剪接调控因子 quaking 可以与环化外显子侧翼内含子上的 quaking 结合位点结合,促进两侧内含子相互靠近,连接成环[7]。

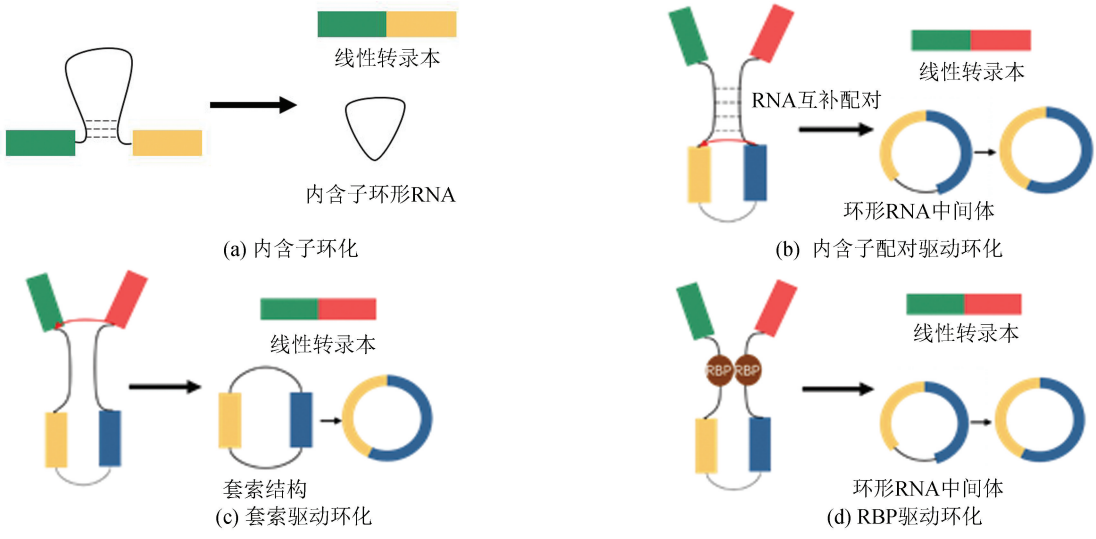

图 1 环形 RNA 环化模型

Fig.1 CircRNA cyclization model

1.2 环形 RNA 的功能

由于环形 RNA 独特的剪接方式及其环状结构,曾经被认为是剪接副产物,不具有生物功能。近年来的研究表明环形 RNA 广泛存在、结构稳定并且具有组织特异性,在生物体生长发育过程中发挥重要作用[8]。以下将从环形 RNA 与 miRNA 相互作用、调控基因表达、与 RNA 结合蛋白相互作用、翻译蛋白质四个方面介绍环形 RNA 的功能。

1.2.1 环形 RNA 与 miRNA 相互作用

近期研究表明,环形 RNA 可以作为 miRNA 分子海绵,通过竞争性的结合 miRNA,降低 miRNA 对其靶分子的抑制作用,进而调控基因表达水平(见图 2 a)。Hansen 等人研究发现 CDR1as (Antisense to the cerebellar degeneration-related protein 1 transcript)含有 miR-7 的超过 70 个保守结合位点,可以调节 miR-7 靶基因的表达[9]。Memczak 等人在

CDR1as 敲除的人类细胞中观察到 miR-7 靶点的下调,认为 CDR1as 是作为 miRNA 海绵来减弱 miRNA 介导的反应[10]。You 等人经过研究认为大多数 circRNA 并不充当 miRNA 海绵,并不比其同源 mRNA 具有更多的 miRNA 结合位点[11]。

1.2.2　环形 RNA 与 RNA 结合蛋白相互作用

研究表明,RNA 结合蛋白(RBPs)如 Argonaute、RNA 聚合酶 II 和 MBL 可与环形 RNA 结合。一些环形 RNA 可以储存、分类或定位 RBP,并且可能像它们调节 miRNA 活性一样,通过作为竞争元素来调节 RBP 的功能(见图 2a)[12]。环形 RNA 独特的三级结构在 RNA 或 RBP 复合物的组装中起重要作用,其可以作为脚手架(Scaffolding)为蛋白质与 RNA、蛋白质与 DNA、蛋白质与蛋白质之间的相互作用提供平台[13]。

1.2.3　环形 RNA 调控基因表达

环形 RNA 可以认为是一种选择性剪接产物,因此环形 RNA 可能在选择性剪接水平上起到调节基因表达的作用。环形 RNA 可与 U1 snRNP 相互作用,增强其亲本基因的表达(见图 2 b)[14]。Burd 等人通过研究推测 cANRIL(Circular ANRIL)具有转录调控功能,cANRIL 的形成降低了 ANRIL 转录本的表达,从而对编码基因 INK4/ARF 的转录进行调控[15]。Yang 等人发现环形 RNA circ-ITCH 可以作为海绵体结合 miR-17 和 miR-224,上调靶基因 p21 和 PTEN 的表达[16]。

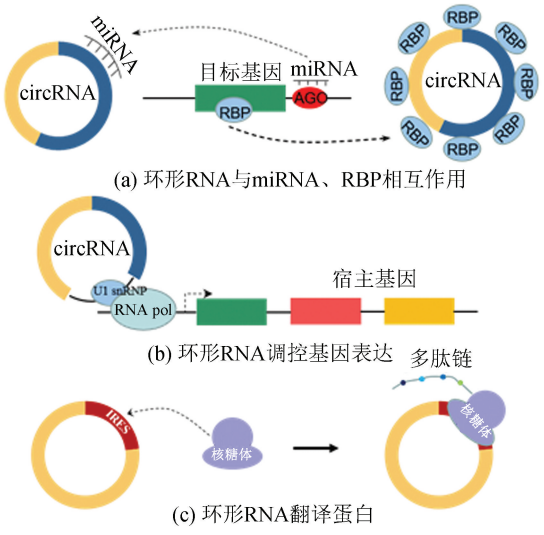

图 2　环形 RNA 功能示意图

Fig.2　Functional diagram of circRNA

1.2.4　环形 RNA 翻译蛋白质

环形 RNA 缺乏帽依赖性翻译的必需元件,例如 5′帽子和 poly(A)尾,而有些环形 RNA 也可以翻译产生蛋白(见图 2 c)。环形 RNA 的非帽依赖性翻译可以通过内部核糖体进入位点(IRES)或在 5′非翻译区(UTR)中加入 m6A 修饰后发生[17]。Legnini 等人证明 circ-ZNF609 包含开放阅读框(ORF),可以通过非帽依赖性翻译编码蛋白控制肌细胞的增殖[18]。Yang 等人通过 Northern blotting 和质谱检测等技术发现 circ-FBXW7 能够编码蛋白 FBXW7-185 aa,可以控制癌细胞周期和增殖[19]。

2　环形 RNA 检测工具

生物信息学方法因其能够高效便捷的处理高通量 RNA-seq 数据,被普遍用来鉴别和分析环形 RNA。为了研究和探索环形 RNA 的普遍性质和多样功能,研究者们开发了各种计算工具从 RNA 测序数据中检测环形 RNA。这些工具根据检测方法可以为两类:基于注释检测和不基于注释检测。

2.1　基于注释的环形 RNA 检测工具

基于注释的环形 RNA 检测工具根据鉴定策略可分为两类,第一类是基于“伪参考序列”的检测策略,第二类是基于“比对片段”的检测策略。

“伪参考序列”检测策略需要根据基因注释信息来构建环形 RNA 的伪序列,通过伪序列检测环形 RNA。基于这种策略的检测工具有 KNIFE 和 PTESFinder 等[20-21]。其中 KNIFE 应用较为广泛,该工具首先从基因组注释信息中构建所有可能的无序“外显子-外显子”连接序列,在 Bowtie2 的帮助下,将读取的数据分别映射到基因组、rRNA 序列、线性“外显子-外显子”连接序列和无序“外显子-外显子”连接序列[22]。真实的 BSJ 读段必须覆盖指定的核苷酸数量且不能比对到基因组和 rRNA 序列以及规范的线性“外显子-外显子”连接序列。KNIFE 增加了从头分析模块检测来自未注释剪接位点的环形 RNA,但 Zeng 等人认为这种从头检测方法是基于统计学进行推断,不能提供准确的断点[23]。

“比对片段”的检测策略的大致思路便是将测序读段与参考基因组比对,跨越 BSJ(Back-spliced junction)的读段被分裂成片段并以相反的顺序与参考基因组对齐,根据这些读段去识别反向剪接位点。基于该策略的工具有 CIRCexplorer、UROBORUS 和 DCC 等[24-26]。其中 CIRCexplorer 使用较为广泛,该工具首先使用 TopHat 将读段比对到参考基因组,然后提取未映射的读段与 TopHat Fusion 的结果进行比对,若读段映射到染色体上的顺序相反且映射位置与已知基因注释中的外显子剪接边界一致,便得到环形 RNA 的反向剪接位点。

相较于不基于注释的检测方法,基于注释的检测方法能够更可靠的检测反向剪接位点,但无法应

用于缺乏基因组注释的物种上。

2.2 不基于注释的环形 RNA 检测工具

不基于注释的环形 RNA 检测工具有 circRNA finder、find_circ 和 CIRI 等[10,27-28]。其中 CIRI 使用较为广泛,该工具使用 BWA-MEM 将读段映射到参考基因组[29]。与上述工具需要从未映射上的读段中提取锚序列来检测反向剪接不同,BWA-MEM 可以自动确定环形 RNA 派生读段的断点。CIRI 对 BWA-MEM 比对完成的结果进行两次扫描计算,过滤掉假阳性的反向剪接位点,精度和灵敏度优于其他检测工具,且运算时间耗费不大,实现了更好的平衡性能。相较于基于注释的检测方法,不基于注释的检测方法应用范围更广,能够检测新的候选环状 RNA,但需要提高检测和过滤过程中的灵敏度和准确性。

3 环形 RNA 全长组装与定量工具

研究表明环形 RNA 在同一个反向剪接位点内部可通过可变剪接形成多个不同的转录本(见图 3)[19]。若要深入研究环形 RNA 的功能特性,必须获取环形 RNA 转录本内部全长序列信息以及对不同环形 RNA 内部可变剪接产物进行精确定量。研究者们开发了多种计算工具用于环形 RNA 内部结构的探索、全长序列的组装及表达量的分析,目前已知的工具(见表 1)。

表 1 环形 RNA 下游分析计算工具

Table 1 Downstream analysis and calculation tools of circRNA

工具名称	功能特点	语言	网址
CIRI-AS	识别环形 RNA 内部剪接结构	Perl	https://sourceforge.net/projects/ciri/files/CIRI-AS/
FUCHS	基于长读段测序,识别环形 RNA 内部可变剪接信息	Python, R	https://github.com/dieterich-lab/FUCHS/tree/master/GCB_testset
CircSplice	识别并比较多组样本环形 RNA 可变剪接差异	Perl	https://github.com/GeneFeng/CircSplice
CIRCexplorer2	基于线性转录本组转工具 Cufflinks 对环形 RNA 全长进行组装	Python	https://github.com/YangLab/CIRCexplorer2
CIRI-full	利用长读段测序组装环形 RNA 全长序列	Java	https://sourceforge.net/projects/ciri-full/
CircAST	环形 RNA 全长序列组装和定量	Python	https://github.com/xiaofengsong/CircAST
IsoCirc	基于纳米孔测序技术识别环形 RNA 全长序列	Python, R	https://github.com/Xinglab/isoCirc
CIRI-long	基于纳米孔测序技术识别环形 RNA 全长序列,可以对矫正纳米孔测序错误	Python	https://github.com/Kevinzjy/CIRI-long
Sailfish-cir	基于线性定量模型 Sailfish 对环形转录本进行定量分析	Python	https://github.com/zerodel/Sailfish-cir
CIRCexplorer3	环形 RNA 与线性 RNA 的定量比较	Python	https://github.com/YangLab/CLEAR
CIRIquant	矫正 RNaseR 数据,对环形 RNA 转录本进行定量分析	Python	https://sourceforge.net/projects/ciri/files/CIRIquant/

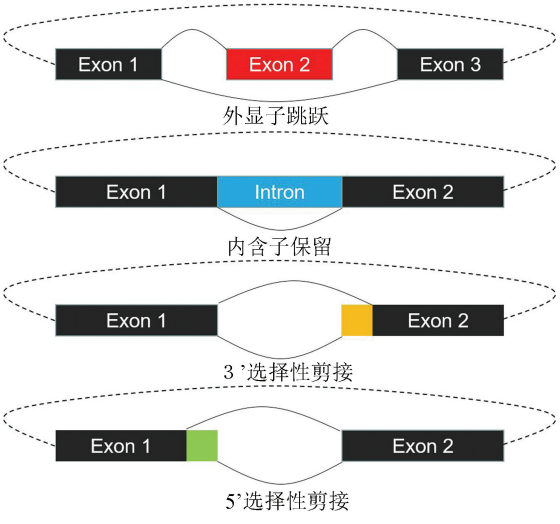

图 3 环形 RNA 内部可变剪接示意图

Fig.3 Schematic diagram of alternative splicing within circRNA

3.1 环形 RNA 内部结构探索及全长序列组装工具

为了更好的揭示环形 RNA 的调控机制,需要深入研究环形 RNA 内部复杂的选择性剪接结构,构建环形 RNA 全长序列。目前识别环形 RNA 内部可变剪接结构的工具主要有 CIRI-AS、FUCHS 和 CircSplice[30-32];针对于环形 RNA 全长序列组装的工具主要有 CIRCexplorer2、CIRI-full、CircAST、isoCirc 和 CIRI-long[33-37]。

Gao 等人开发了 CIRI-AS 工具研究环形 RNA 内部的剪接结构[30]。该工具首先利用 CIRI 找到环形 RNA 反向剪接位点以及跨越反向剪接位点的读段,然后利用双端测序信息,分析跨越反向剪接读段的另一端读段是否支持环形 RNA 内部的前向剪接位点,根据支持前向剪接位点读段信息构建前向剪接图,列出可能发生可变剪接的外显子和所有可能的路径,以检测环形 RNA 内部的可变剪接事件。通过

计算与实验验证相结合的手段,揭示了外显子跳跃,5' 或 3' 可变剪接位点以及内含子保留这四种可变剪接事件在环形 RNA 中普遍存在。CIRI-AS 证实了环形 RNA 内部存在可变剪接事件,但并没有给出环形转录本内部的组成。

Metge 等人开发了工具 FUCHS 对环状 RNA 内可变剪切等信息进行分析和解读[31]。FUCHS 基于长读段测序(大于 150 bp),在运行前需要用户手动输入比对和环形 RNA 位点检测信息。FUCHS 从输入的比对信息中提取出嵌合比对的读段信息,识别出同一反向剪接位点内的前向可变剪接事件。类似于 CIRI-AS,FUCHS 利用双端测序信息来剪接验证环形 RNA 内部的可变剪接,筛选掉只有一段跨越反向剪接位点的双端测序读段实现对假阳性环形 RNA 的过滤。

Feng 等人开发了 CircSplice 工具用于识别环形 RNA 内部的可变剪接事件[32]。该工具通过 STAR 进行比对,通过 GT-AG 和 CT-AC 两种剪切位点过滤并识别环形 RNA 反向剪接位点,保留两端都落在反向剪接位点之间的读段,根据一端比对到反向剪接位点,另一端比对到反向剪接位点内部的读段对外显子跳跃、内含子保留、5' 选择性剪接,3' 选择性剪接四种可变剪接类型进行判断,计算支持可变剪接事件的读段数量,最后根据基因组坐标融合不同的可变剪接事件,并比较不同样本间的可变剪接差异。相较于上文提到的 CIRI-AS,该方法支持不同样本环形 RNA 可变剪接差异的比较。

CIRCexplorer2 是环形 RNA 检测软件 CIRCexplorer 升级版[33]。CIRCexplorer2 将 Tophat 未比对上但映射到 Tophat-fusion 的读段重新比对到已知和从头组装的注释上,可以检测来自注释或新外显子边界的反向剪接位点;增加了检测环形 RNA 中的可变剪接事件功能用于确定环形 RNA 全长,该工具通过比较分析 ploy(A)+和 poly(A)-的数据集识别环形 RNA 内部可变剪接事件,重构环形 RNA 转录本序列。虽然 CIRCexplorer2 能够通过这种方法给出环形 RNA 的全长转录本,其使用的工具却是线性转录本的组装工具 Cufflinks,会得到错误的环形 RNA 序列,给计算带来偏差。

Zheng 等人开发了工具 CIRI-full,提出了一种环形 RNA 转录本组装的新方法[34]。由于二代测序数据读长较短,限制了研究者们只能定位环形 RNA 的反向剪接位点,难以获得环形 RNA 内部的全长结构信息。Zheng 等人于是增加测序读长到 250 bp,在较长的读长下,若某一双端测序读段来自与环形 RNA,两端读段会存在反向重叠区(Reverse overlap,RO),该特征不仅可以识别环形 RNA 的反向剪接位点,而且可以判断该序列是否覆盖整个环形 RNA,从而获取环形 RNA 的全长序列。

2020 年,Wu 等人开发了组装环形 RNA 全长序列的工具 CircAST[35]。该工具首先利用 Tophat 的比对结果找到跨越外显子发生剪接的读段信息,后根据用户通过环形 RNA 识别工具(如 UROBORUS、CIRCexplorer2、CIRI2 等)检测到的反向剪接位点信息以及基因组注释文件中的外显子边界信息,对每一个基因位点构建反向剪接事件的有向无环图(Directed acyclic graph ,DAG),图中的起点和终点对应环形 RNA 中的两个发生反向剪接的外显子。相较于传统的线性转录本拼接算法,CircAST 若检测到某一基因位点存在超过一个反向剪接事件,则构建多个剪接图,因此可以进行准确拼接和组装。CircAST 利用扩展最小路径覆盖算法(Extended minimum path cover,EMPC)结合测序读段信息计算出最优拼接方式,实现环形 RNA 的全长序列的拼接组装。

Xin 等人提出了识别环形 RNA 全长序列的工具 isoCirc[36]。由于短读段测序不能通过实验确定环形 RNA 全长序列的组成,该工具将数据进行线性 RNA 消化,并进行滚环扩增,提高低丰度环形 RNA 的比例,后利用纳米孔长读段测序技术得到包含环形 RNA 全长序列的长度段测序数据。isoCirc 基于上述长读段测序数据,识别出其中重复共有的片段,并将两个相同的重复共有片段连接起来。将连接片段比对到参考基因组以识别环形 RNA 反向剪接位点和环形 RNA 内部前向剪接位点信息。通过多重策略(如比对质量,BSJ/FSJ 精确度等)对比对片段进行过滤,之后便得到环形 RNA 反向剪接位点及全长序列信息。Xin 等人使用 isoCirc 对 12 种人类组织及人类 HEK293 细胞系进行测试,一共检测到 107 147 个环形 RNA 全长序列,其中包含 40 628 个长度大于 500 nt 的环形 RNA 亚型。isoCirc 工具利用纳米孔测序技术,弥补了短读段测序的缺陷,提供了一种研究环形 RNA 全长序列的新策略。

Zhang 等人同样基于纳米孔测序技术,开发了 CIRI-long 工具[37]。在环形 RNA 建库过程中,CIRI-long 与 isoCirc 稍有不同,在对数据进行线性 RNA 消化和环形 RNA 滚环扩增之后,针对于长度更长 cDNA 序列加入了片段长度筛选流程。经过纳米孔测序后得到长读段测序数据,利用 k-mer 匹配方法得到长度段中的重复片段,并使用偏序比对修正测序错误,然后将重复片段比对到参考基因组,结合剪接信号信息,识别反向剪接位点和环形 RNA 内部正

向剪接信息,实现了环形 RNA 的识别及全长序列的重建。Zhang 等人评估了该方法在模拟数据上的效果,并与 Illumina 测序和实时定量 RT-PCR 进行了比较,验证了该方法的准确性。

3.2 环形 RNA 定量工具

为了深入研究环形 RNA 的调控机制以及不同环形 RNA 转录本之间的表达差异,需要对环形转录本进行精确定量。目前对于环形 RNA 进行定量的工具主要有 sailfish-cir、CIRCexplorer3、CIRIquant、CIRI-full 和 CircAST[38-40]。

Li 等人开发了工具 sailfish-cir 对环形 RNA 的表达量进行计算[38]。先前的研究主要根据检测到的反向剪接读段数量来量化环形 RNA 的表达,这种检测方法由于反向剪接读段较少,精度较低。因此 Li 等人将检测到的环形 RNA 在反向剪接位点处切开,将 3'端的序列追加到 5'端,构成伪线性转录本,并加入到真实线性转录本集合中构成混合转录本集合。后利用对线性转录本定量的工具 sailfish 对混合转录本进行定量分析。sailfish-cir 基于期望最大化模型,通过迭代将读段分配到不同转录本上,并估计这些转录本的表达量。与基于反向剪接读段计数的直接定量法相比,sailfish-cir 与 qRT-PCR 定量结果有更强的相关性。

Ma 等人对 CIRCexplorer 分析流程进行了升级,开发了环形 RNA 与线性 RNA 定量比较的工具 CIRCexplorer3-CLEAR[39]。该方法提出了一种定量转录本表达量的参数 FPB(Fragments per billion mapped bases),该参数相较于 FPKM 能够不受测序读段长度的影响。该工具首先利用 HISAT2 将测序数据比对到参考基因组,利用 StringTie 对线性转录本进行组装,计算跨越前向剪接位点的测序片段得到线性转录本的表达 FPBlinear;然后利用 TopHat-Fusion 处理未比对上的测序片段,通过计算跨越反向剪接位点片段得到环形转录本的表达量 FPBcirc,通过环形转录本表达量与对应线性转录本表达量的比值 CIRCscore(FPBcirc/FPBliner)衡量环形 RNA 的相对表达量。相较于其他基于跨越反向剪接位点读段进行定量的方法,该工具提出了新的定量参数 FPB,可以利用 CIRCscore 对环形 RNA 进行相对定量,具有更广泛的适应性。

Zhang 等人开发了鉴定和定量环形 RNA 的算法 CIRIquant[40]。CIRIquant 首先利用 CIRI 等工具从测序数据中识别环形 RNA 的反向剪接位点,然后基于环形 RNA 反向剪接位点构造环形转录本伪参考序列,重新映射读段到伪参考序列。CIRIquant 运用这种方法可以更准确的识别环形 RNA 的反向剪接序列,根据反向剪接读段和前向剪接读段的比例对环形 RNA 进行定量。考虑到环形 RNA 建库时 RNase R 处理存在偏差,CIRIquant 结合未经 RNase R 处理过的数据利用高斯混合模型(Gaussian mixture model,GMM)对 RNase R 处理数据进行校正,使定量更为准确。

除了这些工具外,上文提到的工具 CIRI-full 和 CircAST 不仅可以组装环形转录本全长序列,还可以对环形转录本进行表达量估计。CIRI-full 在构建环形 RNA 的全长后,采用蒙特卡洛方法模拟不同环形 RNA 剪接产物读段在全长序列上的分布,通过梯度下降法求得最优的表达量组合,实现了对不同环形转录本相对丰度的预测评估,并在模拟和真实数据上验证了该方法的准确性。但该方法受到测序长度的限制,需要更高的测序成本和更先进的测序技术来获得较长读长的测序数据;CircAST 在实现环形 RNA 全长序列组装之后,通过期望最大化(Expectation maximization,EM)算法来估计每个组装的环形 RNA 转录本的丰度。其在似然函数的构造等方面充分注意到环形 RNA 的结构特点,在绝对定量和相对定量中均表现出了良好的性能。

4 总结与展望

随着深度测序技术和分子纯化方法的发展,人们对于环形 RNA 的认识正在逐渐发展。在对环形 RNA 的研究中,计算方法因其在高通量 RNA-seq 数据分析中的便利性以及在分析环形 RNA 表达方面的优势而起着重要的作用。由于环形 RNA 自身的结构特点和其特有的剪接事件,我们不能简单地用线性 RNA 的相关工具来解决环形 RNA 的相关计算,需要开发更多针对于环形 RNA 的计算工具。为了更好的揭示环形 RNA 的特殊功能,需要对环形 RNA 内部全长序列进行准确重建。由于二代测序的读段较短,基于二代测序数据重建环形 RNA 全长序列难度较大;基于纳米孔长读段测序技术构建环形 RNA 全长的工具将会是未来研究开发的方向。在环形 RNA 定量方面,现有的工具或基于线性 RNA 开发而来,或对于输入数据存在特定要求。而在全转录组测序数据上由于环形 RNA 与线性 RNA 重叠部分的干扰,精准定量环形 RNA 转录本仍然是一个挑战。开发基于全转录组测序数据环形 RNA 特定的计算工具是目前环形 RNA 研究中迫切需要解决的问题,这对于进一步研究环形 RNA 的生物学功能至关重要。随着计算算法、测序技术、基因组学和生物信息学的发展,相信会有更多环形 RNA 相关计算工

具的出现,促进对于环形 RNA 机制和功能的深入研究。

参考文献(References)

[1]SANGER H L, KLOTZ G, RIESNER D, et al. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures [J]. Proceedings of the National Academy of Sciences of the United States of America, 1976, 73(11): 3852-3856. DOI: 10.1073/pnas.73.11.3852.

[2]GUO J U, AGARWAL V, GUO H, et al. Expanded identification and characterization of mammalian circular RNAs[J]. Genome Biology, 2014, 15(7): 409. DOI: 10.1186/s13059-014-0409-z.

[3]JECK W R, SORRENTINO J A, WANG K, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats[J]. RNA, 2013, 19(2): 141-157. DOI: 10.1261/rna.035667.112.

[4]ZHANG Y, ZHANG X O, CHEN T, et al. Circular intronic long noncoding RNAs [J]. Molecular Cell, 2013, 51(6): 792-806.DOI: 10.1016/j.molcel.2013.08.017.

[5]IVANOV A, MEMCZAK S, WYLER E, et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals [J]. Cell Reports, 2015, 10(2): 170-177. DOI: 10.1016/j.celrep.2014.12.019.

[6]BARRETT S P, WANG P L, SALZMAN J. Circular RNA biogenesis can proceed through an exon-containing lariat precursor[J]. Elife, 2015, 4: e07540. DOI: 10.7554/eLife.07540.

[7]CONN S J, PILLMAN K A, TOUBIA J, et al. The RNA binding protein quaking regulates formation of circRNAs[J]. Cell, 2015, 160(6): 1125-1134. DOI: 10.1016/j.cell.2015.02.014.

[8]LI X, YANG L , CHEN L L. The biogenesis, functions, and challenges of circular RNAs [J]. Molecular Cell, 2018, 71(3): 428-442. DOI: 10.1016/j.molcel.2018.06.034.

[9]HANSEN T B, JENSEN T I, CLAUSEN B H, et al. Natural RNA circles function as efficient microRNA sponges [J]. Nature, 2013, 495 (7441): 384 - 388. DOI: 10. 1038/nature11993.

[10]MEMCZAK S, JENS M, ELEFSINIOTI A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency[J]. Nature, 2013, 495(7441): 333-338. DOI: 10.1038/nature11928.

[11]YOU X, VLATKOVIC I, BABIC A, et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity [J]. Natural Neuroscience, 2015, 18(4): 603-610. DOI: 10.1038/nn.3975.

[12]HENTZE M W, PREISS T. Circular RNAs: Splicing´s enigma variations[J]. EMBO Journal, 2013, 32(7): 923-925. DOI: 10.1038/emboj.2013.53.

[13]VAN ROSSUM D, VERHEIJEN B M, PASTERKAMP R J. Circular RNAs: Novel regulators of neuronal development [J]. Frontiers in Molecular Neuroscience, 2016, 9: 74. DOI: 10.3389/fnmol.2016.00074.

[14]LI Z, HUANG C, BAO C, et al. Exon-intron circular RNAs regulate transcription in the nucleus [J]. Natural Struction Molecular Biology, 2015, 22(3): 256-264. DOI: 10.1038/nsmb.2959.

[15]BURD C E, JECK W R, LIU Y, et al. Expression of linear and novel circular forms of an INK4/ARF-associated noncoding RNA correlates with atherosclerosis risk[J]. PLoS Genetics, 2010, 6(12): e1001233. DOI: 10.1371/journal.pgen.1001233.

[16]YANG C, YUAN W, YANG X, et al. Circular RNA circ-ITCH inhibits bladder cancer progression by sponging miR-17/miR-224 and regulating p21, PTEN expression [J]. Molecular Cancer, 2018, 17(1): 19. DOI: 10.1186/s12943-018-0771-7.

[17]DIALLO L H, TATIN F, DAVID F, et al. How are circRNAs translated by non-canonical initiation mechanisms? [J]. Biochimie, 2019, 164: 45-52. DOI: 10.1016/j.biochi.2019.06.015.

[18]LEGNINI I, DI TIMOTEO G, ROSSI F, et al. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis[J]. Molecular Cell, 2017, 66(1): 22-37 e29. DOI: 10.1016/j.molcel.2017.02.017.

[19]YANG Y, GAO X, ZHANG M, et al. Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis [J]. Journal of the National Cancer Institute, 2018, 110(3): 304-315. DOI: 10.1093/jnci/djx166.

[20]SZABO L, MOREY R, PALPANT N J, et al. Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development[J]. Genome Biology, 2015, 16: 126. DOI: 10.1186/s13059-015-0690-5.

[21]IZUOGU O G, ALHASAN A A, ALAFGHANI H M, et al. PTESFinder: a computational method to identify post-transcriptional exon shuffling (PTES) events[J]. BMC Bioinformatics, 2016, 17: 31. DOI: 10.1186/s12859-016-0881-4.

[22]LANGDON W B. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks[J]. BioData Mining, 2015, 8(1): 1. DOI: 10.1186/s13040-014-0034-0.

[23]ZENG X, LIN W, GUO M, et al. A comprehensive overview and evaluation of circular RNA detection tools [J]. PLoS Comput ational Biology, 2017, 13(6): e1005420. DOI: 10.1371/journal.pcbi.1005420.

[24]ZHANG X O, WANG H B, ZHANG Y, et al. Complementary sequence-mediated exon circularization [J]. Cell,

2014,159(1):134-147. DOI:10.1016/j.cell.2014.09.001.
[25]SONG X, ZHANG N, HAN P, et al. Circular RNA profile in gliomas revealed by identification tool UROBORUS[J]. Nucleic Acids Research,2016,44(9):e87. DOI:10.1093/nar/gkw075.
[26]CHENG J, METGE F, DIETERICH C. Specific identification and quantification of circular RNAs from sequencing data[J]. Bioinformatics,2016,32(7):1094-1096. DOI:10.1093/bioinformatics/btv656.
[27]WESTHOLM J O, MIURA P, OLSON S, et al. Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation[J]. Cell Reports,2014,9(5):1966-1980. DOI:10.1016/j.celrep.2014.10.062.
[28]GAO Y, WANG J, ZHAO F. CIRI: An efficient and unbiased algorithm for de novo circular RNA identification[J]. Genome Biology,2015,16:4. DOI:10.1186/s13059-014-0571-3.
[29]LI H, DURBIN R. Fast and accurate long-read alignment with Burrows-Wheeler transform[J]. Bioinformatics,2010,26(5):589-595. DOI:10.1093/bioinformatics/btp698.
[30]GAO Y, WANG J, ZHENG Y, et al. Comprehensive identification of internal structure and alternative splicing events in circular RNAs[J]. Natural Communication,2016,7:12060. DOI:10.1038/ncomms12060.
[31]METGE F, CZAJA-HASSE L F, REINHARDT R, et al. FUCHS-towards full circular RNA characterization using RNAseq[J]. PeerJ,2017,5:e2934. DOI:10.7717/peerj.2934.
[32]FENG J, CHEN K, DONG X, et al. Genome-wide identification of cancer-specific alternative splicing in circRNA[J]. Molecular Cancer,2019,18(1):35. DOI:10.1186/s12943-019-0996-0.
[33]ZHANG X O, DONG R, ZHANG Y, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs[J]. Genome Research,2016,26(9):1277-1287. DOI:10.1101/gr.202895.115.
[34]ZHENG Y, JI P, CHEN S, et al. Reconstruction of full-length circular RNAs enables isoform-level quantification[J]. Genome Medicine,2019,11(1):2. DOI:10.1186/s13073-019-0614-1.
[35]WU J, LI Y, WANG C, et al. CircAST: Full-length assembly and quantification of alternatively spliced isoforms in circular RNAs[J]. Genomics Proteomics Bioinformatics,2019,17(5):522-534. DOI:10.1016/j.gpb.2019.03.004.
[36]XIN R, GAO Y, GAO Y, et al. isoCirc catalogs full-length circular RNA isoforms in human transcriptomes[J]. Natural Communication,2021,12(1):266. DOI:10.1038/s41467-020-20459-8.
[37]ZHANG J, HOU L, ZUO Z, et al. Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long[J]. Natural Biotechnology,2021,39:836-845. DOI:10.1038/s41587-021-00842-6.
[38]LI M, XIE X, ZHOU J, et al. Quantifying circular RNA expression from RNA-seq data using model-based framework[J]. Bioinformatics,2017,33(14):2131-2139. DOI:10.1093/bioinformatics/btx129.
[39]MA X K, WANG M R, LIU C X, et al. CIRCexplorer3: A clear pipeline for direct comparison of circular and linear RNA expression[J]. Genomics Proteomics Bioinformatics,2019,17(5):511-521. DOI:10.1016/j.gpb.2019.11.004.
[40]ZHANG J, CHEN S, YANG J, et al. Accurate quantification of circular RNAs identifies extensive circular isoform switching events[J]. Natural Communication, 2020,11(1): 90. DOI:10.1038/s41467-019-13840-9.