



Poly(A)尾检测技术研究进展

贾铮^{1,2†}, 刘静雯^{1,2†}, 陆发隆^{1,2*}

1. 中国科学院遗传与发育生物学研究所, 整合生理调控实验室, 北京 100101

2. 中国科学院大学, 北京 100049

† 同等贡献

* 联系人, E-mail: flul@genetics.ac.cn

收稿日期: 2025-05-03; 接受日期: 2025-05-26; 网络版发表日期: 2025-06-10

国家自然科学基金(批准号: 32170606)和中国科学院稳定支持基础研究领域青年团队计划(批准号: YSBR-012)资助

摘要 绝大部分真核生物的mRNA会在其3'末端发生非模板来源的多聚腺苷酸化, 生成称为poly(A)尾的多聚腺苷酸链。Poly(A)尾不仅帮助mRNA从细胞核转运到细胞质, 促进mRNA翻译为蛋白质, 其长度和碱基组成亦与密切影响着mRNA的稳定性密切相关。然而, 由于poly(A)尾的长同聚物特性, 其序列难以被准确测定。长期以来, poly(A)尾的功能与其调控机制研究受限于其检测技术。近年来, 随着高通量测序技术的飞速发展, 各类poly(A)尾的检测技术有了长足的进步, 推动了mRNA poly(A)尾的组成与调控研究。本文系统总结了当前poly(A)尾检测技术的研究进展, 涵盖传统分子生物学方法、基于二代测序和基于三代单分子测序的测序方法, 并阐述了各方法的原理、样本制备与捕获策略, 以及在尾长测定、尾内非A残基识别与检测灵敏度等方面的性能特点与局限。Poly(A)尾检测技术的发展不仅揭示了poly(A)尾长度介导的RNA转录后调控功能, 也为探究非模板来源的poly(A)尾序列携带的poly(A)尾表观遗传调控信息奠定了坚实的技术基础。

关键词 poly(A)尾, 翻译, RNA稳定性, poly(A)尾非A残基, poly(A)尾表观遗传调控, RNA表观遗传学

mRNA转录后加工在基因表达调控与生命系统稳态维持中扮演着重要角色。在真核生物中, mRNA的3'末端通过多聚腺苷酸化生成poly(A)尾结构。这一过程不仅是mRNA成熟的标志性事件, 更通过调控mRNA的核质转运、翻译及稳定性, 深刻影响基因表达^[1]。随着全转录组poly(A)尾研究的深入, poly(A)尾长度动态与其内部的非A残基的动态与功能逐渐被揭示。其中, 精确检测poly(A)尾是揭示其在mRNA代谢调控中功能与机制的关键突破口之一。近年来, 随着核酸分子检测技术的进步, poly(A)尾检测方法已从针对单一基因的传统检测技术(如PAT assay)发展到全转录组高通量

测序解决方案。本文系统梳理了poly(A)尾检测不同技术平台的技术演进, 着重探讨高通量测序时代下的poly(A)尾动态研究, 旨在为RNA poly(A)尾研究提供方法学参考。

1 Poly(A)尾的代谢和功能

mRNA是基因转录生成的, 是“中心法则”中关键的遗传信息传递载体。几乎所有真核生物的mRNA前体在其3'端均会经历3'末端切割和多聚腺苷酸化的过程, 生成由数十至数百个腺苷酸残基构成的多聚腺苷

引用格式: 贾铮, 刘静雯, 陆发隆. Poly(A)尾检测技术研究进展. 中国科学: 生命科学, 2025, 55: 1134–1148

Jia Z, Liu J W, Lu F L. Technologies for measuring RNA poly(A) tails (in Chinese). Sci Sin Vitae, 2025, 55: 1134–1148, doi: 10.1360/SSV-2025-0143

酸链,称为poly(A)尾.作为一种普遍存在的mRNA转录后修饰,poly(A)尾在mRNA的稳定性、核质转运及翻译中起着至关重要的作用^[1].

1.1 Poly(A)尾的代谢

Poly(A)尾的加工依赖于多种蛋白复合体和RNA顺式调控元件的协同作用.这些蛋白质在真核生物中高度保守,通过不同的功能模块相互作用,构成了一个复杂的加工机器^[2].参与3'末端加工的关键蛋白包括:切割因子 I (cleavage factor Im, CFIm)、切割因子 II (cleavage factor IIm, CFIIIm)、切割与多聚腺苷酸化特异性因子(cleavage and polyadenylation specificity factor, CPSF)、切割刺激因子(cleavage stimulation factor, CstF),以及poly(A)聚合酶(poly(A) polymerase, PAP)^[3].多聚腺苷酸化信号(polyadenylation signal, PAS)通常是一段保守的AAUAAA或者AUUAAA序列,位于剪切位点的上游,其上下游分别分布有多种顺式调控元件^[4].RNA剪切发生时,CPSF特异性识别并结合PAS以及其下游的序列,CFIm和CstF也分别识别PAS上游和下游的序列^[4].最终,CFIm, CPSF和CstF共同形成复合体,在CPSF的作用下于PAS下游约10~30 nt处切割RNA前体.剪切完成后,CPSF与FIP1(factor interacting with PAP1)结合,招募PAP到剪切后的mRNA 3'端,启动腺苷残基的添加^[5].细胞核poly(A)结合蛋白(poly(A) binding protein nuclear, PABPN)继而结合到短的poly(A)尾上,进一步促进poly(A)尾的延长形成完整的poly(A)尾^[2].

mRNA在完成细胞核内的多聚腺苷酸化后,poly(A)尾帮助其转运到细胞质.在细胞质内,poly(A)尾可经历进一步的加工,包括尾巴去腺苷酸化和重新延伸.Poly(A)尾的去腺苷酸化主要通过PAN2-PAN3复合体和CCR4-NOT复合体来完成^[1,6],而细胞中也存在抑制去腺苷酸化的调控机制^[7,8].重新延伸则主要依赖于末端核苷酸转移酶(terminal nucleotidyltransferases, TENTs)^[9].例如,TENT4A和TENT4B可催化poly(A)尾的重新延伸,并且其催化生成的poly(A)尾中会掺入少量非A残基(以G为主),形成含有非A残基的poly(A)尾^[10].G残基的掺入能够抑制CCR4-NOT复合体的去腺苷酸化作用,从而有助于提高转录本的稳定性^[10,11].TUT4(terminal uridylyl transferase 4)和TUT7(terminal uridylyl transferase 7)这两种末端尿苷酸转移酶,可在

短于25个核苷酸,或完全去腺苷酸化的poly(A)尾的3'端添加U残基^[12].带有末端U残基修饰的mRNA可进一步通过DIS3L2和外切体的作用,从3'端开始降解;或在LSM1-7复合物和XRN1核酸外切酶的作用下,从5'端开始降解^[12].此外,TENT2与TENT5能够在特定的场景下在细胞质中对特定的mRNA进行多聚腺苷酸化,延长其poly(A)尾,以激活翻译或增强稳定性^[9,13~15].近期的一项研究显示,TENT5A介导的多聚腺苷酸化可能能够作用于外源导入的mRNA,增强mRNA疫苗的效果^[16].有意思的是,尽管功能与机制仍然未知,在有丝分裂细胞周期中,poly(A)尾末端的G和U残基的修饰呈现明显的互斥的随细胞周期的丰度波动^[17].

1.2 Poly(A)尾的功能

Poly(A)尾不仅是mRNA成熟和转运的必要结构组分,更通过多种机制参与基因表达的精细调控,其主要功能可简要归纳为mRNA稳定性调控、翻译调控,以及发育特异性调控等三个方面.

通常认为poly(A)尾是mRNA在细胞中稳定存在的重要因素,其对mRNA的保护作用主要体现在:一方面,通过保护3'末端,抑制外切酶对mRNA的降解^[6];另一方面,poly(A)尾长度的缩短往往先于mRNA降解,是触发去腺苷酸化依赖性降解通路的关键步骤^[1].在细胞命运转换的过程中,主动poly(A)去尾可有效清除前一个状态的调控网络的相关mRNA^[18].例如,在炎症刺激下,促炎因子TNF α (tumor necrosis factor α)和Cox-2(cyclooxygenase-2)的mRNA去腺苷酸化可迅速降低其稳定性和翻译效率,从而调控炎症反应的时空特异性^[19].需要注意的是,稳态poly(A)尾的长度与mRNA的稳定性之间不存在明显的正相关性^[20].

一般认为,poly(A)尾与5'端帽结构通过环状蛋白复合体相互作用,共同形成mRNA“闭环”结构,促进了eIF4F复合体的招募与翻译起始子复合体的组装,从而起始mRNA的翻译^[1].在神经元和生殖细胞等具备高度极化结构的细胞中,poly(A)尾的长度与局部mRNA翻译密切相关.以海马突触后致密区为例,CPEB1在 α CaMKII mRNA 3' UTR识别CPE序列后,招募CPE依赖的poly(A)聚合酶进行再加尾,增强局部翻译,从而推动突触可塑性和记忆巩固^[21,22].同样,在果蝇、斑马鱼、非洲爪蟾及小鼠等多种模式生物的卵母细胞和早期胚胎中,poly(A)尾长度与翻译效率呈显著正相关,

进而影响卵母细胞和胚胎的发育^[23~29]。在果蝇(*Drosophila melanogaster*)的成熟卵母细胞和受精卵中, mRNA的poly(A)尾越长, 其翻译效率越高^[23]。对斑马鱼(*Danio rerio*)和非洲爪蟾(*Xenopus laevis*)的早期胚胎进行的研究亦显示, poly(A)尾的长度与蛋白翻译效率成正相关^[24]。在小鼠GV卵母细胞中也发现了poly(A)尾的长度与mRNA翻译效率之间的正相关关系^[25,26]。

近期的全转录组poly(A)尾分析发现, 在发育过程中poly(A)尾不仅在长度上有动态变化, 还广泛掺入非A残基(U, C, G), 这些修饰在生殖细胞和胚胎发育过程中可能扮演了关键角色^[25,27,30~35]。例如, 斑马鱼与非洲爪蟾早期胚胎中, poly(A)尾3'末端富含U残基, 可以促进母源mRNA的降解^[34,36]。在人类1~4细胞期胚胎中, 超过60%的mRNA poly(A)尾内部含有U残基^[27], 这种母源mRNA中非A残基随发育的动态变化在小鼠、大鼠及猪中也同样存在, 提示其在哺乳动物生殖发育中的重要性^[37]。此外, poly(A)尾中的碱基是否存在碱基化学修饰值得关注, 目前有一项在布氏锥虫中的研究发现, 其编码变异表面糖蛋白(variant surface glycoprotein, VSG)的转录本的poly(A)尾中存在m⁶A修饰, 可帮助提高该转录本的稳定性^[38]。

更多关于poly(A)尾长度与非A残基调控机制的深入综述, 请参阅本团队近期关于poly(A)尾介导的表观遗传调控的综述^[39]。

2 Poly(A)尾检测技术

Poly(A)尾作为一种特殊的核酸序列, 其检测技术归属于核酸序列检测技术的范畴, 旨在获取特定基因转录本或全转录组中poly(A)尾的长度及其碱基组成的信息, 而其长同聚物的特性一方面为基于传统分子生物学方法的检测提供了条件, 另一方面又导致其基于测序技术的检测难于常规的核酸序列检测^[30]。随着主流核酸分子检测技术的发展, poly(A)尾检测技术亦不断演进(图1)。早期主要基于电泳、层析等传统分子生物学技术; 随后, Illumina等NGS(next-generation sequencing)二代测序平台兴起, 催生了一批基于二代高通量测序的poly(A)尾检测方法; 近年来, 以PacBio(Pacific BioSciences)公司推出的PacBio HiFi和ONT(Oxford Nanopore Technologies)公司推出的Nanopore

等为代表的三代单分子测序平台又带来了poly(A)尾准确测定的新技术(图1)。尽管各方法原理与流程存在差异, poly(A)尾检测通常可分为四个基本步骤: 含完整poly(A)尾的转录本或转录本片段的捕获、测定平台特异的样本或文库准备、poly(A)尾测定与数据解读。其中, 各类poly(A)尾巴测定方法在特定应用场景中的适用性主要由其对含完整poly(A)尾的转录本或转录本片段的捕获策略与分子检测技术平台的发展水平共同决定。

对于完整poly(A)尾的捕获是样本文库构建的重要步骤, 主要有3'端接头连接(3'-end adapter ligation)^[23,24,30,35,40~50]、末端延伸(end extension)^[25,51~56]、G/I延伸(G/I extension)^[33,57,58]和模板置换(template switch)^[34]等四种常用的基本策略, 这四种基本捕获策略在各个时期的poly(A)尾检测技术当中被广泛采用, 分别用于应对不同场景下的poly(A)尾检测。其中, 3'端接头连接策略将一段特定序列的接头片段连接到感兴趣的转录本的3'端(图2A), 末端延伸(或称为模板指导的末端延伸)策略利用一段有特定序列且能够与poly(A)尾互补配对的核酸片段与poly(A)尾配对后作为模板指导在原poly(A)尾后进行延伸以添加接头序列(图2B), G/I延伸策略在相应转录本的3'端以非模板依赖的形式直接合成、延伸一段由G残基与I残基的片段(图2C), 而模板置换策略则利用II型内含子反转录酶和与转录本最3'端仅一个碱基互补的接头引物进行反转录(图2D)。

对于基本捕获策略进行改动, 可以提高其特定方面的性能, 例如3'端夹板接头连接(splint ligation)^[23,40~42,49]对3'接头片段进行了改进, 通过添加一个能够同时与原3'接头片段、poly(A)尾互补配对的“夹板”片段, 提升转录本与接头片段连接的效率, 进而提高其灵敏度, 并可靶向特定的poly(A)⁺ RNA。相应地, 我们将3'端夹板接头连接以外的3'端接头连接方法称为3'端接头直接连接。此外, 对于一些特化的场景, 有一些特殊的捕获方式, 例如环化^[59]等, 这里不展开介绍。

同时, 根据poly(A)尾测定所依赖的核心技术平台, 可将poly(A)尾检测技术的发展大致分为三个阶段: 基于传统分子生物学方法的检测技术、基于二代测序的测序技术和基于三代单分子测序的测序技术。此外, 在基于三代单分子测序的测序技术当中, 基于Nanopore

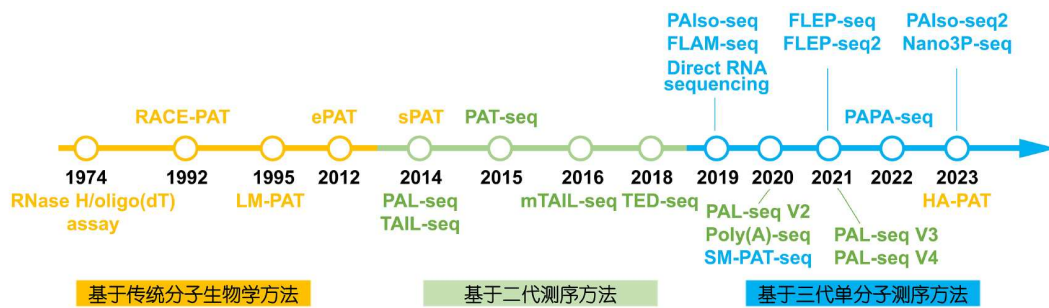


图 1 Poly(A)检测技术发展示意图。横线自左向右为时间轴, 圆圈标注各技术出现的年份。黄色标示基于传统分子生物学方法的检测技术, 绿色标示基于二代测序方法的检测技术, 蓝色标示基于三代单分子测序方法的检测技术

Figure 1 Schematic of poly(A) detection technology development. The horizontal line from left to right represents the timeline, with circle marks indicating the year each technology appeared. Yellow denotes methods based on traditional molecular biology techniques; green denotes methods based on second-generation sequencing; blue denotes methods based on third-generation single-molecule sequencing

平台的测序技术的性能也取决于其采用的数据分析方法^[60], 且其对于poly(A)尾长^[61,62]、内部非A残基的相关信息提取都需要设计专门的计算工具^[63]。

总体而言, 对转录本捕获和文库构建的优化可提升技术的灵敏度或降低偏倚, 而检测结果的精度与准确性则主要取决于其检测技术平台。接下来, 本文将结合上述四个基本步骤与各个发展阶段的特点、进展及其局限性, 对当前主流的poly(A)尾检测技术进行讨论。

2.1 基于传统分子生物学方法的poly(A)尾检测技术

Poly(A)尾由均一的腺苷酸残基组成, 因此, 其可与poly(U)或oligo(dT)序列特异性地配对。早期的检测方法主要依托电泳或色谱技术, 通过分析是否含poly(A)尾, 以及含有较长或较短的poly(A)尾的转录本在凝胶中的迁移或在oligo(dT)/poly(U)柱层析中的相对结合能力, 对特定基因的转录本或总的转录本进行定性或定量测定, 无需复杂算法即可直观评估尾长变化。

2.1.1 基于电泳的poly(A)尾检测技术

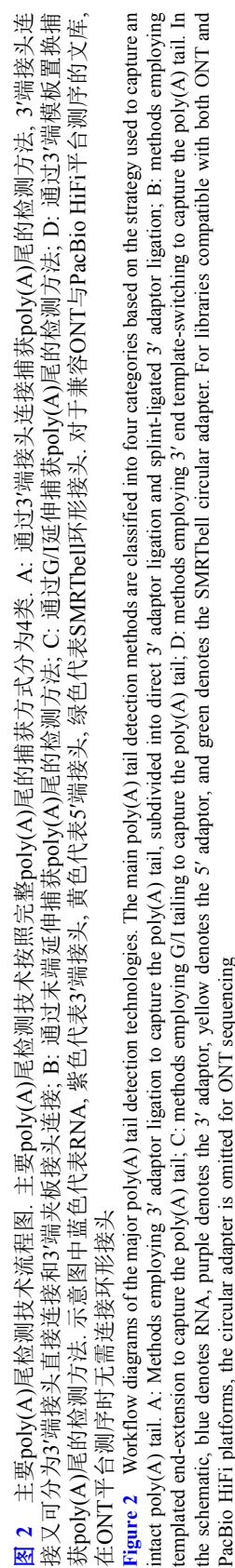
基于电泳的方法利用poly(A)尾与oligo(dT)的特异性杂交, 通过RNase H降解或PCR扩增等手段, 生成尾长各异的RNA或cDNA片段, 随后利用凝胶电泳分析降解或扩增的产物的分子量, 从而推测原转录本poly(A)尾的长度。

RNase H/oligo(dT) assay是最早应用于研究poly(A)尾长度的技术^[64]。该方法利用RNase H特异性

降解RNA:DNA杂交片段中RNA序列的特性, 将oligo(dT)片段与RNA进行杂交并加入RNase H, 特异性降解与oligo(dT)互补的poly(A)尾序列, 然后通过Northern杂交或凝胶电泳检测其与带有poly(A)尾的原转录本迁移率的差别, 从而判断poly(A)尾的长度^[64]。该方法直观且特异, 但由于Northern杂交技术复杂, 难度较大, 一次只能检测一个基因, 且受限与凝胶电泳的分辨率, 难以检测2 kb以上的mRNA的poly(A)尾。因此, 该方法通常仅局限于检测个别相对高丰度基因转录本的poly(A)尾长度, 且通量极为有限。

随着PCR技术的发展, 出现了一系列基于RT-PCR(反转录PCR)的poly(A)尾检测技术, 其代表为RACE-PAT(rapid amplification of cDNA ends poly(A) test)^[51,52]和LM-PAT(ligase-mediated poly(A) test)^[65]等。RACE-PAT通过带有接头序列的oligo(dT)引物进行反转录, 反转录随机起始于poly(A)尾内部, 产生含有不同长度poly(A)尾的cDNA片段, 经用接头序列和基因特异序列的PCR扩增, 有较长poly(A)尾的转录本的扩增产物在电泳凝胶上呈现为一个较长的弥散状态, 反之则呈现为一个较短的弥散状态或一个单一条带^[51,52]。LM-PAT则在T4 DNA连接酶作用下, 将与poly(A)尾互补的5'磷酸化的oligo(dT)片段连接成一个整体, 进行反转录, 然后再在反转录产物的最5'端连接上带有接头序列的锚定oligo(dT)序列, 经PCR扩增, 可以获得包含完整poly(A)尾序列的扩增片段, 用于poly(A)尾长度分析^[65]。与RACE-PAT相比, LM-PAT具有更高的灵敏度^[65,66]。

后续对基于PCR的poly(A)尾检测技术的改进主



要集中于含有完整poly(A)尾的转录本的捕获策略, 以提高其准确性与灵敏度. ePAT(extension poly(A) test)方法采用末端延伸策略捕获转录本, 利用3'端为oligo(dT)的包含5'接头序列的寡聚核苷酸为模板, 利用Klenow聚合酶, 在poly(A)尾的3'端以上述寡聚核苷酸为模板合成一段DNA序列作为后续PCR的接头^[53]; sPAT(splint-mediated poly(A) test)方法则采用3'端夹板接头连接策略, 以poly(A)尾序列为引导在mRNA的3'端连接一段RNA接头^[40]; Hire-PAT(high-resolution poly(A) test)方法则采用G/I延伸策略, 在mRNA的3'端酶促合成G尾或G/I尾, 以作为 C_mT_n 的反转录引物的结合位点, 进行接下来的RT-PCR^[57]. 这些完整poly(A)尾捕获策略已经比较成熟, 并被广泛用于后续出现的高通量测序相关poly(A)尾检测方法的文库构建中. 然后, 这些PAT方法的PCR产物可通过凝胶电泳进行分析, 用于估算poly(A)尾的长度. 此外, 这些PAT方法可以在PCR引物上添加荧光基团, 可以对产物进行毛细管电泳, 获得更高的分辨率. 当然, 现有高灵敏度核酸毛细管电泳技术可直接分析未进行荧光标记的PCR产物, 同样可获得高分辨率的poly(A)尾长度检测结果^[25,57]. 图1中黄色部分为典型的非高通量检测方法.

总体来看, 基于PCR的方法比之前的方法更便捷, 需要的样本材料也更少, 且oligo(dT)片段结合至转录本内部的情况对其影响也更小. 但PCR扩增会导致不同程度的PCR偏差(PCR bias), 是基于PCR的poly(A)尾测定中常见的噪声来源^[66].

这些基于电泳的poly(A)尾检测方法使得对于poly(A)尾检测成为可能, 具有便捷、易推广、易复现、成本低等优势, 目前仍广泛用于单个基因转录本的poly(A)尾检测和新方法的验证等, 近年来也出现了HA-PAT等新方法^[67]. 但这类技术对于基因特异的PCR扩增和电泳技术的依赖限制了其检测通量.

2.1.2 基于色谱层析的poly(A)尾检测技术

由于poly(A)尾可以与poly(U)或oligo(dT)序列特异性结合, 因此可以利用poly(U)或oligo(dT)介质, 通过与介质结合能力的差异将含有较长的和较短的poly(A)尾的转录本进行有效的色谱层析分离. 基于色谱层析分离的方法最初用于检测单个基因转录本的poly(A)尾长度^[68,69]. 在此基础上进一步结合微阵列技术, 可以用于定性检测全转录组mRNA poly(A)尾的长

与短^[70,71]. 这类技术在对含poly(A)尾的转录本的捕获方面有不同的技术路线, 一种常用策略是基于oligo(dT)和盐浓度梯度洗脱. RNA首先与生物素化的poly(dT)序列结合并固定到磁珠介质上, 并以盐浓度梯度洗脱, 从而分别获得短poly(A)尾和长poly(A)尾两个组分^[70]. 另一种路线则是通过poly(U)层析和热洗脱得到两种不同长度poly(A)尾组分^[71]. 两种路线可以混合使用, 各取所长. 分离后的RNA样本通过微阵列进行杂交与信号检测, 即可获得各基因poly(A)尾长短定性分布情况^[70,71].

上述色谱层析结合微阵列的poly(A)尾检测技术被成功应用于神经元中CPEB介导的poly(A)尾聚腺苷酸化的检测^[72]、受poly(A)聚合酶GLD4调控的转录本的鉴定^[73], 以及酵母体系中poly(A)尾长变化、Pab1结合与核糖体占据情况的相关性的分析^[74]等研究工作中. 该类方法存在两方面缺陷, 一是对poly(A)尾只能进行定性的长短分类, 无法精准测量poly(A)尾长度及其动态变化, 另一方面, 该技术带有微阵列转录组检测的所有缺陷, 而随着高通量测序技术的兴起, 微阵列技术在转录组检测中的绝大部分应用已被高通量测序取代. 因此, 除色谱层析介导的不同poly(A)长分离技术由于其技术独特性仍在特定应用场景下有其价值外, 这类技术目前已被弃用.

2.2 基于二代测序方法的poly(A)尾检测技术

过去20多年间, 基因组学测序技术快速发展, 一系列以Illumina SBS(sequencing by synthesis)为代表的被称为NGS的新型测序技术纷纷出现, 为核酸序列的测定提供了超高通量、低成本的方案, 这类技术也被称为二代测序技术^[75]. 二代测序的超高通量短读长的特性催生了转录组测序技术, 也称为RNA-seq. 该技术通过对测到的cDNA片段计数来进行全转录组RNA的定量, 是转录组分析的革命性技术进步^[75].

然而, 由于同聚物在Illumina平台测序时容易产生相位错误, 并且少量未切割的荧光分子积累会导致持续的信号干扰, 以Illumina平台为代表的大多数二代测序平台的标准碱基识别算法难以直接测定长同聚物的序列^[24,30]. 因此, poly(A)尾的序列在Illumina平台上的常规转录组测序中通常在文库制备、测序或数据分析阶段被丢弃. 例如, Smart-seq2在构建测序文库过程中反转录时使用锚定3'UTR的Oligo(dT)引物(5'-AAG-

CAGTGGTATCAACGCAGAGTACT₃₀VN-3', 其中“N”表示A, T, C或G, “V”表示A, C或G), 通过末端的“N”和“V”碱基将反转录引物精确定位于3'UTR末端, 从而在最终cDNA文库中去除了大部分的poly(A)尾序列, 避免了同聚合物序列的干扰^[76]。

为了能在二代测序平台对全转录组poly(A)尾进行检测, 研究者开发了以PAL-seq(poly(A)-tail length profiling by sequencing)^[24]和TAIL-seq^[30]为代表的基于二代平台的全转录组poly(A)尾检测技术(表1)。这些技术的poly(A)尾测定方式跳出了二代测序常规的碱基读取程序的范畴, PAL-seq修改了已经停产的Illumina GA II测序仪的测序的化学反应, 在预计为poly(A)位置的互补链合成时掺入非末端终止的dTTP和生物素修饰的dUTP分子, 使其在合成poly(A)尾的互补链时掺入, 然后通过测定其与荧光标记的链霉亲和素孵育所产生的荧光的强度推断相应T/U序列的长度, 即poly(A)尾的长度^[24]。TAIL-seq则修改了测序仪的碱基读取算法, 使用测序仪获取的原始荧光数据文件, 并利用Tailseeker算法来估算应该为poly(A)尾序列处的poly(A)尾长度^[30]。这两个技术通过不同的poly(A)尾长度估计方法避开了二代测序常规碱基读取程序无法识别长同聚物的缺陷, 实现了对poly(A)尾长度的相对准确的测量。之后在TAIL-seq的基础上进一步衍生出了mTAIL-seq^[23], PAL-seq v2^[41], PAL-seq v3和PAL-seq v4技术^[42]。

在含有完整poly(A)尾的转录本的捕获方面, 大部分基于二代测序平台的poly(A)尾检测技术采用了与基于PCR电泳的检测技术类似的策略。TAIL-seq方法采用3'接头直接连接的策略^[24], PAL-seq则采用了3'端夹板接头连接的策略^[30]。在TAIL-seq的后续改进版本的mTAIL-seq中, 最重要的改进也是采用包含与poly(A)尾末端匹配的oligo(dT)的发卡夹板连接策略取代了原有的3'接头直接连接^[23]。PAL-seq采用的是已经停产的Illumina GA II平台^[24], 其应用与推广高度受限。因此, PAL-seq的后续迭代PAL-seq v2^[41], PAL-seq v3和PAL-seq v4^[42]技术使用了类似TAIL-seq的策略来进行poly(A)尾测定。PAL-seq v2采用与PAL-seq类似的3'端夹板接头连接策略捕获转录本, 不过对夹板接头进行了改进, 使其能够捕捉并检测末端为U残基的poly(A)尾^[41]。在此基础上, PAL-seq v3通过在3'接头中添加条码(barcode)序列^[42], 使得来自多个样本的文库能够被

混合测序, 同时在建库过程中省去PCR扩增, 直接测定反转录得到的cDNA的序列, 从而去除PCR扩增带来的偏差; 但这种方法的灵敏度较低(作者认为可能是因为添加了条码的接头序列的连接效率欠佳), 因此原作者团队^[42]将添加了条码的接头序列更换回一般的接头序列, 并将来自各个样本的转录本分别测序, 得到了灵敏度更高的PAL-seq v4。

3'接头直接连接策略的优势在于不受RNA 3'端序列影响, 可捕获各种末端序列的poly(A)尾以及无poly(A)尾的RNA, 为了去除高丰度的rRNA需要进行rRNA去除或者poly(A)⁺ RNA的富集。oligo(dT)的发卡夹板连接由于需要poly(A)尾序列的参与, 则无需考虑rRNA的问题。然而, 同样由于需要poly(A)尾序列的参与, 该策略无法捕获无poly(A)尾序列的RNA以及poly(A)尾末端有复杂非A残基的RNA转录本, 虽然mTAIL-seq和PAL-seq v2-4等方法通过加入多种夹板接头部分保留了末端为U残基的poly(A)尾^[23,41,42]。

在PAL-seq和TAIL-seq及其衍生版本之外, 另有一些基于二代测序平台的poly(A)尾检测技术出现(图1中绿色所示部分)。PAT-seq(poly(A)-test RNA sequencing)采用模板指导的末端延伸策略捕获完整poly(A)尾, 该技术利用Klenow聚合酶进行末端延伸代替adapter连接以提高转录本捕获效能^[54], 需要注意的是, 该方法直接用常规二代测序碱基读取程序读取poly(A)尾序列, 会在poly(A)长度估计时产生较大偏差; TED-seq(tail-end displaced sequencing)通过精确选择文库片段长度进行poly(A)尾长度估算^[46]; poly(A)-seq同样直接用常规二代测序碱基读取程序读取poly(A)尾序列^[32,45]。此外, 也出现了EnD-Seq(exonuclease degradation sequencing)^[44], 3'RACE-seq^[43], Nascent RNAend-Seq^[50], circ-TAIL-seq^[59]等一系列用于检测特定类型或单个尾巴较短的转录本的尾巴序列的方法, 这些方法能够检测非poly(A)尾。表1简要总结了这些基于二代测序平台的poly(A)尾检测方法, 包括各方法所需RNA最低起始量、poly(A)尾捕获策略、poly(A)尾序列检测情况和相应的数据分析算法等。

这些基于二代测序的转录组范围的poly(A)尾测序技术极大地促进了人们对于poly(A)尾的理解与认识。其中一个重要的发现是通过TAIL-seq发现部分poly(A)尾的末端存在非A残基, 包括U, G和C, 其中U倾向于加在非常短的poly(A)尾上^[30]。进一步研究表明,

表 1 基于二代测序方法的poly(A)尾检测技术
Table 1 Technologies for measuring RNA poly(A) tails by next-generation sequencing

检测方法	RNA最低起始量	Poly(A)尾 捕获方法	poly(A)尾序列 检测情况	数据分析算法	参考文献
TAIL-Seq	50 μg	3'端接头直接连接	可检测末端非A残基	相对荧光信号转换	[30]
mTAIL-seq	1 μg	3'端夹板接头连接	可检测末端U残基	相对荧光信号转换	[23]
PAL-seq	1 μg	3'端夹板接头连接	不能检测	相对荧光强度	[24]
PAL-seq v2	25 μg	3'端夹板接头连接	可检测末端U残基	相对荧光信号转换	[41]
PAL-seq v3	500 ng	3'端夹板接头连接	可检测末端U残基	相对荧光信号转换	[42]
PAL-seq v4	500 ng	3'端夹板接头连接	可检测末端U残基	相对荧光信号转换	[42]
TED-seq	N.A. ^{a)}	3'端接头直接连接	不能检测	DNA长度选择	[46]
PAT-seq	1 μg	末端延伸	可检测内部和末端非A残基	标准碱基读取	[54]
Poly(A)-seq	5.1 μg	3'端接头直接连接	可检测内部和末端非A残基	标准碱基读取	[45]
EnD-Seq	1.5 μg	3'端接头直接连接	可检测非A残基	AppEnD	[44]
3'RACE-seq	2~3 μg	3'端接头直接连接	可检测特定转录本中的非A残基	标准碱基读取	[43]
Nascent RNAend-Seq	600 ng	3'端接头直接连接	可检测特定新生转录本中的非A残基	标准碱基读取	[50]
circTAIL-seq	5 μg	RNA环化	可检测特定转录本中的非A残基	circTAIL-seq analyzer	[59]

a) N.A.代表RNA起始量未知(原始文献中未明确写明RNA起始量)

U残基可标记降解过程中的RNA poly(A)尾末端, 并促进其快速降解^[12]. G和C残基则抑制去尾复合物的活性, 帮助维持mRNA的poly(A)尾并提高相应mRNA的稳定性^[30]. 并且poly(A)尾末端的U残基含量在斑马鱼和非洲爪蟾的早期胚胎中表现高度的动态变化^[36]. 对果蝇卵母细胞的研究发现, 在其成熟过程中mRNA poly(A)尾的长度变化与其翻译效率变化表现出高度的正相关性, 表明poly(A)尾长度在卵母细胞翻译重塑中起着重要的调控作用^[23,77]. 在斑马鱼和非洲爪蟾的早期胚胎中也观察到了类似的poly(A)尾长度与翻译效率高度关联的现象^[24]. 然而, 在哺乳动物体细胞中, poly(A)尾的长度与翻译效率无明显相关性^[20,24,30], 当其poly(A)尾大于20 nt时, 其翻译效率即与poly(A)尾长度没有明显的正相关关系^[20]. 以上情况表明, poly(A)尾长度对mRNA翻译的调控有高度的细胞类型特异性.

总的来说, 基于二代测序平台的poly(A)尾检测技术极大地拓展了人们对于poly(A)尾动态与调控的认识. 然而二代测序平台无法准确测量长同聚物的问题并未得到根本解决, 因此, 采用常规序列读取程序的技术对poly(A)尾的检测准确度都较低. TAIL-seq和PAL-seq及其衍生技术拥有相对较准确的poly(A)长度测定能力, 然而其需要对测序程序进行定制, 无法从标准测序服务中获得, 因此, 只有少数实验室能开展这些

技术, 极大地限制了其应用. 尽管存在不少缺陷, 基于二代测序平台的poly(A)尾检测技术凭借低成本的优势仍然有着很好的生命力.

2.3 基于三代单分子测序的poly(A)尾检测技术

以PacBio单分子实时(single-molecule real-time, SMRT)测序和ONT纳米孔(Nanopore)测序为代表的三代单分子测序技术的兴起^[78], 为poly(A)尾检测提供了新契机. 目前, 三代单分子测序主要分为两种平行发展的技术路线: 一是基于SMRT原理的PacBio平台(表2); 二是基于测量核酸分子通过纳米孔时电流变化的ONT平台(表2). 与二代测序平台相比, PacBio和ONT平台在测定长同聚物序列时能够取得较好的性能表现, 虽然对同聚物序列测序的准确度仍会低于常规序列. Pac-Bio平台的HiFi模式通过将DNA分子环化, 实现对单分子的多次循环测序, 并生成多次测序的环形一致性序列(circular consensus sequence, CCS)实现对单分子序列的准确测定. 同时, 三代测序平台读长较长, 能够一次性测定转录本的完整序列; 且两种平台的文库构建方案高度兼容, 例如, 对FLEP-seq(full-length elongating and polyadenylated RNA sequencing)和FLEP-seq2的完整转录组文库在PacBio HiFi和ONT平台上进行测序分析, 发现结果有较好的一致性^[35,47]. 因此, 三代测序平台是目前较为理想的poly(A)尾测定平台.

表 2 基于三代单分子测序的poly(A)尾检测技术

Table 2 Technologies for measuring RNA poly(A) tails by third-generation single-molecule sequencing

检测方法	RNA最低起始量	Poly(A)尾捕获方法	Poly(A)尾序列检测情况	测序平台	参考文献
FLAM-seq	500 ng	G/I延伸	可检测内部非A残基	PacBio HiFi	[33]
FLEP-seq	3 μ g	3'端接头直接连接	可检测内部和末端非A残基	ONT或PacBio HiFi	[47]
FLEP-seq2	500 ng	3'端接头直接连接	可检测内部和末端非A残基	ONT或PacBio HiFi	[35]
PAIso-seq	0.5 ng	末端延伸	可检测内部非A残基	PacBio HiFi	[25,55,56]
PAIso-seq2	100 ng	3'端接头连接	可检测内部和末端非A残基	PacBio HiFi	[48]
PAPA-seq	100 ng	G/I延伸	可检测内部非A残基	PacBio HiFi	[58]
SM-PAT-seq	4 μ g	3'端夹板接头连接	可检测内部非A残基	PacBio HiFi	[49]
Nano3P-seq	500 ng	模板置换	可检测内部和末端非A残基	ONT	[34]
Direct RNA sequencing	100 ng	3'端夹板接头连接	算法依赖	ONT	[60-63]

2.3.1 基于PacBio HiFi平台的poly(A)尾检测技术

PacBio技术利用单分子实时测序, 将单个DNA聚合酶分子固定在微孔中, 并通过荧光标记在DNA合成时实时监测每一个核苷酸的加入^[78]。最早出现的两种技术是FLAM-seq(full-length mRNA sequencing)^[33]和PAIso-seq(poly(A) inclusive RNA isoform sequencing)^[25,55,56], 随后又有FLEP-seq(full-length elongating and polyadenylated RNA sequencing)^[47], FLEP-seq2^[35], PAIso-seq2^[48]等方法被开发出来。在PacBio HiFi平台上进行poly(A)尾检测时, 可直接采用标准的PacBio碱基识别算法, 获取完整poly(A)尾序列, 无需额外设计针对poly(A)尾的碱基读取算法。

PAIso-seq采用末端延伸策略捕获含有poly(A)尾的转录本, 以含有oligo(dT)和接头序列的引物为模板指导在poly(A)尾的末端加上一段接头序列, 去除末端延伸模板后, 利用该接头序列进行反转录、模板置换和PCR扩增, 获得包含完整poly(A)尾序列的全长转录组文库^[25,55,56](图2B)。FLAM-seq则先用oligo(dT)磁珠捕获poly(A)⁺ RNA, 再以G/I延伸策略捕获完整poly(A)尾, 继而进行反转录、模板置换和PCR扩增, 生成全长文库^[33](图2C)。PAIso-seq利用含有oligo(dT)的引物为末端延伸的模板以捕获含poly(A)尾的转录本, 避免了poly(A)⁺ RNA转录本富集过程导致的偏差, 同时提高了检测的灵敏度, 但由于这种末端延伸模板的结合位置有微小的随机性, 可能有一定概率在模板指导的末端延伸过程中在原poly(A)尾的3'端随机加入少量A碱基, 产生噪声。FLAM-seq采用的G/I延伸策略无法区分转录本是否有poly(A)尾, 所以FLAM-seq方法采

用了oligo(dT)磁珠进行poly(A)⁺转录本富集, 该富集过程使其更倾向于捕获有更长poly(A)尾的转录本, 导致在poly(A)尾长度检测上存在一定的偏差。受限于各自的转录本和完整poly(A)尾捕获策略, FLAM-seq与PAIso-seq两种方法均无法测定poly(A)尾末端的非A残基, 且不会捕获尾巴较短或尚未加尾的新生的转录本^[25,33,55,66]。

PAIso-seq2, FLEP-seq和FLEP-seq2采用3'端接头直接连接策略在所有转录本的3'端连接接头, 再依托接头序列进行反转录、模板置换和PCR扩增, 构建包含完整poly(A)尾序列在内的全长转录组文库^[35,47,48](图2A)。3'端接头直接连接策略对于RNA 3'端的序列组成没有选择性, 能够无偏差地测序各种类型的RNA序列, 无论其是否有poly(A)尾, 或者poly(A)尾是否有非A残基。但是3'端接头直接连接效率较低, 导致这些方法的灵敏度逊于PAIso-seq^[48]。同时由于3'端接头直接连接策略对于RNA 3'端的序列组成没有选择性, 细胞内高丰度的rRNA在此过程中同样可被捕获, 因此PAIso-seq2与FLEP-seq采用了不同的方法进行rRNA去除以降低rRNA量, 从而降低成本。PAIso-seq2在几轮PCR扩增后, 利用特异靶向各种rRNA的sgRNA通过CRISPR/Cas9降解rRNA反转录生成的cDNA, 然后再继续进行PCR并上机测序, 降低检测结果中rRNA来源的转录本^[48]。FLEP-seq则利用RiboMinus或者riboPOOL先从RNA样本中去除rRNA, 然后再进行3'端接头直接连接^[47]。然后, PAIso-seq2文库在PacBio HiFi平台进行测序, 获取包含完整poly(A)尾序列的全长转录组信息^[48]。FLEP-seq与FLEP-seq2的文库被成功在

ONT和PacBio HiFi平台测序^[35,47]。虽未进行直接比较,推测PAIso-seq2与FLEP-seq和FLEP-seq2的PacBio HiFi测序在poly(A)尾长度测量和非A残基检测方面的技术性能相当。

此外,还出现了PAPA-seq(poly(A)-PacBio sequencing)和SM-PAT-seq等方法(图1中蓝色所示部分)。PAPA-seq在原理与实现上与FLAM-seq高度相似,成功地揭示了小鼠精子发生过程中的poly(A)尾长度动态变化^[58]。SM-PAT-seq在poly(A)⁺富集的基础上采用含有oligo(dT)的3'端夹板接头连接策略捕获含有完整poly(A)尾的转录本,然后在PacBio HiFi平台上进行测序,很好地实现了全转录组poly(A)尾长的检测^[49]。

这些基于PacBio平台的poly(A)尾检测技术极大地推动了poly(A)尾的精准解析。其中最重要的发现是poly(A)尾内部广泛存在非A残基^[25,33,48,55]。自mRNA poly(A)尾被发现以来,研究者曾一直认为poly(A)尾是由一串纯A残基组成^[1,79],直到利用TAIL-seq技术发现poly(A)尾的末端存在非A残基^[30]。然而基于二代测序的poly(A)尾测序技术无法有效检测poly(A)尾内部的残基组成。基于PacBio HiFi的poly(A)尾测序技术PAIso-seq, PAIso-seq2和FLAM-seq等基于PacBio HiFi的方法发现在poly(A)尾的内部存在广泛的非A残基^[25,33,48,55]。U, C和G有着不一样的分布特征,且在发育过程中表现为动态变化^[27]。有意思的是, lncRNA的poly(A)尾中也分布有非A残基^[48]。如果是纯A的序列, poly(A)尾能编码的信息有限,仅A的长度;然而非A残基掺入的数量、位置、组合加上poly(A)尾的长度则可有庞大的信息编码空间,可能承载重要的RNA表观遗传调控信息,我们将其称为“poly(A)尾介导的表观遗传信息”,详见近期综述^[39]。

基于PacBio HiFi平台的poly(A)尾检测技术能够测定包含完整poly(A)尾的全长转录本,直接获得poly(A)尾的长度,且能够准确检测到poly(A)尾内部的非A残基等一系列先前无法检测到的poly(A)尾序列特征及其动态,极大地拓展了人们对于poly(A)尾的认识。除poly(A)尾外,全长转录本的信息可同时用于可变剪接与可变加尾(alternative polyadenylation, APA)分析等。结合更多RNA分析技术,这些技术可以用于染色体结合RNA的poly(A)尾分析^[47],含有m⁶A修饰RNA的poly(A)尾分析^[80],以及不同亚细胞组分中的RNA poly(A)尾分析^[48]等。这些技术目前面临的主要挑战包

括:文库构建过程中反转录导致转录本上的化学修饰信息丢失,文库构建过程中全长转录本cDNA扩增时PCR造成的偏差,测序中PacBio平台较高的单条cDNA测序成本等^[39,66]。

2.3.2 基于ONT平台的poly(A)尾检测技术

当核酸分子通过纳米孔时,会引起孔道电流的特征性波动,ONT纳米孔测序技术即基于对这些电信号解析进行核酸序列的测定^[81]。该平台的工作原理决定了其能够不经过PCR扩增,利用电压驱动单条核酸分子以从3'端到5'端的顺序通过纳米孔,直接采集蕴含着核酸分子序列及碱基化学修饰信息的电信号数据^[60,82],并通过机器学习算法从电信号数据中提取核酸序列和碱基化学修饰等信息。

目前,基于ONT平台可测序完整poly(A)尾的技术包括FLEP-seq^[47], FLEP-seq2^[35], Nano3P-seq(Nanopore 3 prime end-capture sequencing)^[34]和DRS(direct RNA sequencing)^[60]等。其中, FLEP-seq, FLEP-seq2和Nano3P-seq均对cDNA进行测序,而DRS则直接测序原始RNA转录本。一般来说,对原转录本RNA进行检测有助于对转录本的化学修饰状况进行全面、深入的挖掘,而对cDNA测序则可通过PCR扩增获得更好的检测灵敏度^[66]。

FLEP-seq和FLEP-seq2的测序文库构建已在前文基于PacBio HiFi平台的部分中介绍,在此不再重复。Nano3P-seq利用Ⅱ类内含子编码的反转录酶,通过一种特殊的模板置换策略进行完整poly(A)尾的捕获。该反转录酶只需要反转录引物有一个碱基与RNA配对就可进行反转录。该方法设计了一种特殊的类似夹板的接头,但其3'末端仅一个“N”(A, T, C或G)残基突出, N残基可与RNA的最3'末端碱基匹配,起始反转录,使得核酸合成匹配的序列从接头置换到目的RNA上,称为模板置换反转录。随后降解原RNA链,并在接头处退火互补的寡核苷酸并连接测序接头用于ONT测序包含完整poly(A)尾的全长转录本的cDNA^[34]。该方法利用N残基匹配RNA的最3'末端碱基,因此其对RNA 3'末端序列无选择性,可捕获所有转录本,这一点与PAIso-seq2, FLEP-seq和FLEP-seq2的性能相似,并且其也可较为准确地测定poly(A)尾的长度、poly(A)尾末端和内部的非A残基^[34],尽管其准确度可能会略低于基于PacBio平台的检测。

DRS以3'端夹板接头连接策略通过与poly(A)尾末端互补的oligo(dT)夹板接头连接捕获转录本, 随后进行反转录并连接测序接头, 最后在ONT平台测序cDNA/RNA杂交分子中的RNA链, 采集RNA分子穿过纳米孔时的电信号数据, 用于后续分析^[60]. 该捕获策略决定了其无法捕获3'末端为非A残基或无poly(A)尾的转录本. 尽管DRS测序的电信号数据中包含了完整的poly(A)尾序列通过纳米孔的电信号数据, 然而标准的ONT碱基读取程序无法正确解码poly(A)尾的长度或其序列^[83]. 因此, 针对ONT平台的poly(A)尾检测, 需要设计专用的算法. 此外, 对于RNA分子与cDNA分子的ONT测序数据, 需要分别设计不同的信号解析算法^[66].

从ONT原始电信号数据中提取poly(A)尾信息时, 需要首先将数据分割为测序接头、夹板接头、poly(A)尾以及转录本主体等区段, 并在原始电流信号数据中确定各分界点, 随后针对poly(A)尾对应的区段应用专门算法解析poly(A)尾的信息. 目前已有多款工具可直接从ONT转录本测序数据当中提取poly(A)尾信息.

Nanopolish-polyA是较早开发并广泛应用的poly(A)尾长分析算法^[61]. 该工具利用隐马尔可夫模型(HMM)划分电信号数据的不同区段, 然后根据可以确定长度的区段对应的电信号持续时长确定相应核酸分子通过纳米孔的速率, 结合poly(A)尾信号的持续时间, 推断相应转录本的poly(A)尾长. 该算法在DRS数据上可取得较好的效果, 但在cDNA测序数据中灵敏度和准确度较低^[61]. 该方法也被成功整合至诸如poreplex的DRS分析工具包中, 用于解析如SARS-CoV-2的转录组的poly(A)尾长度^[84].

TailfindR算法则通过电信号的不同区段之间的原始信号变化斜率(raw signal slope)来划分信号的不同区段^[62], 随后通过与Nanopolish-polyA类似的利用poly(A)序列通过纳米孔的时长, 及相应转录本通过纳米孔的平均速率的方式来估算poly(A)尾的长度. 该工具对RNA与cDNA的测序数据均适用, 但在cDNA数据分析中存在较高假阳性率^[62,66].

由于FLEP-seq与FLEP-seq2在ONT平台测定的是扩增后的cDNA序列, 已有工具效果欠佳, 为此开发了PolyAcaller这一poly(A)尾数据分析工具^[35,47]. 该工具先用常规碱基读取算法确定可读取序列, 从而提取常规碱基读取算法无法解析的poly(A)尾信号区段, 再对其进行poly(A)尾长度估算, 该工具在cDNA测序数据

中的poly(A)尾分析性能优于TailfindR^[35,47]. Nano3P-seq同样面对cDNA测序数据, 对TailfindR工具进行了针对性修改^[34], 并自主开发了PolyTailor^[85]. 值得注意的是, Nano3P-seq的上述两种方法不仅能估算poly(A)尾长, 还可以实现poly(A)尾内部的非A残基的测定^[34,85].

早期的DRS poly(A)尾分析工具均无法测定其中的非A残基. 近期发表的Ninetails工具则填补了这方面的空白^[63]. 该工具在Nanopolish-polyA等工具确定的poly(A)尾序列对应的信号区段基础上, 引入卷积神经网络(CNN), 对poly(A)尾相应区段的信号进行了深度解析, 能够鉴别、定量并定位poly(A)尾中的非A残基^[63].

需要注意的是, 基于PacBio HiFi平台的poly(A)尾测定技术可以在RNA或cDNA层面测定包含完整poly(A)尾的全长转录组信息, 而目前标准的ONT碱基读取程序无法直接获得poly(A)尾信息, 需要依赖上述专用的poly(A)尾分析算法, 这些算法仍然在快速发展的过程中. 尤其需要注意的是在测序芯片升级后需要重新评估其在新芯片产生数据中的适用性, 或迭代新的算法^[85]. 目前, ONT平台对poly(A)尾的检测准确性略逊于PacBio HiFi平台, 但在单条读段的成本和测序仪的可及性方面具有优势. 此外, ONT可直接测序原始RNA, 使其在同时获取poly(A)尾和转录本上碱基化学修饰信息方面具有很大的潜力^[66].

3 Poly(A)尾检测技术的开发方向和应用展望

本文从对含有完整poly(A)尾的转录本的捕获策略和核酸分子检测技术两个维度出发, 梳理了各类poly(A)尾捕获策略及不同时期的核酸分子检测技术的优势与局限, 并对主要poly(A)尾检测技术进行了深入讨论, 展示了该领域的技术演进. 自2014年至今, 基于高通量测序平台陆续开发出多种poly(A)尾检测技术(图1中绿色和蓝色部分所示), 表1和表2列举了已有的基于二代和三代测序平台的poly(A)尾检测方法的特性, 包括所需RNA起始量、poly(A)尾捕获策略、poly(A)尾序列组成检测能力及测序平台等.

随着RNA表观遗传调控网络的复杂性的不断揭示, poly(A)尾检测技术仍有多方面需求亟待满足, 包

括多模态RNA表观遗传修饰检测, 低成本和高通量单细胞检测等。

RNA上的修饰不是单一的, 包括此前认为RNA结构组分的5'端帽子、3'端尾巴、A-to-I编辑、碱基化学修饰(如m⁶A, m⁵C, ac⁴C)等。目前, 对这些修饰的研究多是针对单一修饰进行, 这些修饰之间如何协同调控RNA的代谢与功能是未来值得研究的重大问题。这方面的研究对这些修饰在单分子水平上高灵敏度的定量联合检测提出了要求。

目前, 基于PacBio HiFi和ONT三代单分子测序平台的技术在poly(A)尾检测中表现较好, 但单条cDNA的测序成本仍然偏高, 成为限制poly(A)尾功能与调控深入研究的重要技术瓶颈。未来亟需在测序化学、文

库制备和信号解析等方面实现技术突破, 以进一步显著降低单分子poly(A)尾检测的成本, 以推动poly(A)尾研究的广泛与深入的开展。

过去十年高通量单细胞转录组和空间转录组技术取得了飞速发展, 极大地推动了发育生物学、神经生物学、免疫学和疾病发生发展机制研究^[86]。然而, 目前这些单细胞和空间转录组技术都在文库构建过程中就已经丢弃了poly(A)尾信息。鉴于poly(A)尾中可能承载着丰富的重要RNA表观遗传调控信息, 开发兼具包含完整poly(A)尾信息的高通量单细胞和空间转录组技术, 可在单细胞和空间层面提供转录组定量之外全新层面的转录组调控信息, 将为解析细胞异质性和组织空间调控机制打开全新视野。

参考文献

- 1 Passmore L A, Collier J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat Rev Mol Cell Biol*, 2022, 23: 93–106
- 2 Tian B, Manley J L. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci*, 2013, 38: 312–320
- 3 Mandel C R, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci*, 2008, 65: 1099–1122
- 4 Neve J, Patel R, Wang Z, et al. Cleavage and polyadenylation: ending the message expands gene regulation. *RNA Biol*, 2017, 14: 865–890
- 5 Laishram R S. Poly(A) polymerase (PAP) diversity in gene expression—Star-PAP vs canonical PAP. *FEBS Lett*, 2014, 588: 2185–2197
- 6 Eckmann C R, Rammelt C, Wahle E. Control of poly(A) tail length. *WIREs RNA*, 2011, 2: 348–361
- 7 Park J, Kim M, Yi H, et al. Short poly(A) tails are protected from deadenylation by the LARP1-PABP complex. *Nat Struct Mol Biol*, 2023, 30: 330–338
- 8 Yang J, Bu J, Liu B, et al. MARTRE family proteins negatively regulate CCR4-NOT activity to protect poly(A) tail length and promote translation of maternal mRNA. *Nat Commun*, 2025, 16: 248
- 9 Yu S, Kim V N. A tale of non-canonical tails: gene regulation by post-transcriptional RNA tailing. *Nat Rev Mol Cell Biol*, 2020, 21: 542–556
- 10 Lim J, Kim D, Lee Y, et al. Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. *Science*, 2018, 361: 701–704
- 11 Lee Y, Leviansky Y, Jung Y, et al. Deadenylation kinetics of mixed poly(A) tails at single-nucleotide resolution. *Nat Struct Mol Biol*, 2024, 31: 826–834
- 12 Lim J, Ha M, Chang H, et al. Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell*, 2014, 159: 1365–1376
- 13 Brouze M, Czarnocka-Cieciura A, Gewartowska O, et al. TENT5-mediated polyadenylation of mRNAs encoding secreted proteins is essential for gametogenesis in mice. *Nat Commun*, 2024, 15: 5331
- 14 Liudkovska V, Krawczyk P S, Brouze A, et al. TENT5 cytoplasmic noncanonical poly(A) polymerases regulate the innate immune response in animals. *Sci Adv*, 2022, 8: eadd9468
- 15 Nousch M, Minasaki R, Eckmann C R. Polyadenylation is the key aspect of GLD-2 function in *C. elegans*. *RNA*, 2017, 23: 1180–1187
- 16 Krawczyk P S, Mazur M, Orzel W, et al. Re-adenylation by TENT5A enhances efficacy of SARS-CoV-2 mRNA vaccines. *Nature*, 2025, 641: 984–992
- 17 Liu Y, Nie H, Lu F. Dynamic RNA 3' uridylation and guanylation during mitosis. *iScience*, 2020, 23: 101402
- 18 Quan Y, Wang M, Xu C, et al. Cnot8 eliminates naïve regulation networks and is essential for naïve-to-formative pluripotency transition. *Nucleic Acids Res*, 2022, 50: 4414–4435
- 19 Weill L, Belloc E, Bava F A, et al. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat Struct Mol Biol*, 2012, 19: 577–585
- 20 Park J E, Yi H, Kim Y, et al. Regulation of poly(A) tail and translation during the somatic cell cycle. *Mol Cell*, 2016, 62: 462–471

- 21 Atkins C M, Davare M A, Oh M C, et al. Bidirectional regulation of cytoplasmic polyadenylation element-binding protein phosphorylation by Ca^{2+} /calmodulin-dependent protein kinase II and protein phosphatase 1 during hippocampal long-term potentiation. *J Neurosci*, 2005, 25: 5604–5610
- 22 Aslam N, Kubota Y, Wells D, et al. Translational switch for long-term maintenance of synaptic plasticity. *Mol Syst Biol*, 2009, 5: 284
- 23 Lim J, Lee M, Son A, et al. mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev*, 2016, 30: 1671–1682
- 24 Subtelny A O, Eichhorn S W, Chen G R, et al. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*, 2014, 508: 66–71
- 25 Liu Y, Nie H, Liu H, et al. Poly(A) inclusive RNA isoform sequencing (PAIso-seq) reveals wide-spread non-adenosine residues within RNA poly (A) tails. *Nat Commun*, 2019, 10: 5292
- 26 Xiong Z, Xu K, Lin Z, et al. Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nat Cell Biol*, 2022, 24: 968–980
- 27 Liu Y, Zhao H, Shao F, et al. Remodeling of maternal mRNA through poly(A) tail orchestrates human oocyte-to-embryo transition. *Nat Struct Mol Biol*, 2023, 30: 200–215
- 28 Liu Y, Tao W, Wu S, et al. Maternal mRNA deadenylation is defective in *in vitro* matured mouse and human oocytes. *Nat Commun*, 2024, 15: 5550
- 29 Wu Y K, Su R, Jiang Z Y, et al. ZAR1 and ZAR2 orchestrate the dynamics of maternal mRNA polyadenylation during mouse oocyte development. *Genome Biol*, 2025, 26: 120
- 30 Chang H, Lim J, Ha M, et al. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell*, 2014, 53: 1044–1052
- 31 Morgan M, Much C, DiGiacomo M, et al. mRNA 3' uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. *Nature*, 2017, 548: 347–351
- 32 Zhao T, Huan Q, Sun J, et al. Impact of poly(A)-tail G-content on *Arabidopsis* PAB binding and their role in enhancing translational efficiency. *Genome Biol*, 2019, 20: 189
- 33 Legnini I, Alles J, Karaiskos N, et al. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nat Methods*, 2019, 16: 879–886
- 34 Begik O, Diensthuber G, Liu H, et al. Nano3P-seq: transcriptome-wide analysis of gene expression and tail dynamics using end-capture nanopore cDNA sequencing. *Nat Methods*, 2023, 20: 75–85
- 35 Jia J, Lu W, Liu B, et al. An atlas of plant full-length RNA reveals tissue-specific and monocots-dicots conserved regulation of poly(A) tail length. *Nat Plants*, 2022, 8: 1118–1126
- 36 Chang H, Yeo J, Kim J, et al. Terminal uridylyltransferases execute programmed clearance of maternal transcriptome in vertebrate embryos. *Mol Cell*, 2018, 70: 72–82.e7
- 37 Liu Y, Jin J, Zhang Y, et al. Conservation and divergence of poly(A) tail regulation during the mammalian oocyte-to-embryo transition. *bioRxiv*, 2021, 458065
- 38 Viegas I J, de Macedo J P, Serra L, et al. N^6 -methyladenosine in poly(A) tails stabilize VSG transcripts. *Nature*, 2022, 604: 362–370
- 39 Liu J, Lu F. Beyond simple tails: poly(A) tail-mediated RNA epigenetic regulation. *Trends Biochem Sci*, 2024, 49: 846–858
- 40 Minasaki R, Rudel D, Eckmann C R. Increased sensitivity and accuracy of a single-stranded DNA splint-mediated ligation assay (sPAT) reveals poly(A) tail length dynamics of developmentally regulated mRNAs. *RNA Biol*, 2014, 11: 111–123
- 41 Eisen T J, Eichhorn S W, Subtelny A O, et al. The dynamics of cytoplasmic mRNA metabolism. *Mol Cell*, 2020, 77: 786–799.e10
- 42 Xiang K, Bartel D P. The molecular basis of coupling between poly(A)-tail length and translational efficiency. *eLife*, 2021, 10: e66493
- 43 Goldfarb K C, Cech T R. 3' terminal diversity of MRP RNA and other human noncoding RNAs revealed by deep sequencing. *BMC Mol Biol*, 2013, 14: 23
- 44 Welch J D, Slevin M K, Tatomer D C, et al. EnD-Seq and AppEnD: sequencing 3' ends to identify nontemplated tails and degradation intermediates. *RNA*, 2015, 21: 1375–1389
- 45 Yu F, Zhang Y, Cheng C, et al. Poly(A)-seq: a method for direct sequencing and analysis of the transcriptomic poly(A)-tails. *PLoS One*, 2020, 15: e0234696

- 46 Woo Y M, Kwak Y, Namkoong S, et al. TED-Seq identifies the dynamics of poly(A) length during ER stress. *Cell Rep*, 2018, 24: 3630–3641.e7
- 47 Long Y, Jia J, Mo W, et al. FLEP-seq: simultaneous detection of RNA polymerase II position, splicing status, polyadenylation site and poly(A) tail length at genome-wide scale by single-molecule nascent RNA sequencing. *Nat Protoc*, 2021, 16: 4355–4381
- 48 Liu Y, Nie H, Zhang Y, et al. Comprehensive analysis of mRNA poly(A) tails by PAIso-seq2. *Sci China Life Sci*, 2023, 66: 187–190
- 49 Mattijssen S, Iben J R, Li T, et al. Single molecule poly(A) tail-seq shows LARP4 opposes deadenylation throughout mRNA lifespan with most impact on short tails. *eLife*, 2020, 9: e59186
- 50 Roake C M, Chen L, Chakravarthy A L, et al. Disruption of telomerase RNA maturation kinetics precipitates disease. *Mol Cell*, 2019, 74: 688–700.e3
- 51 Salles F J, Strickland S. Analysis of poly(A) tail lengths by PCR: the PAT assay. *Methods Mol Biol*, 1999, 118: 441–448
- 52 Sallés F J, Darrow A L, O'Connell M L, et al. Isolation of novel murine maternal mRNAs regulated by cytoplasmic polyadenylation. *Genes Dev*, 1992, 6: 1202–1212
- 53 Jänicke A, Vancuylenberg J, Boag P R, et al. ePAT: a simple method to tag adenylated RNA to measure poly(A)-tail length and other 3' RACE applications. *RNA*, 2012, 18: 1289–1295
- 54 Harrison P F, Powell D R, Clancy J L, et al. PAT-seq: a method to study the integration of 3'-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA*, 2015, 21: 1502–1510
- 55 Liu Y, Zhang Y, Wang J, et al. Transcriptome-wide measurement of poly(A) tail length and composition at subnanogram total RNA sensitivity by PAIso-seq. *Nat Protoc*, 2022, 17: 1980–2007
- 56 Liu Y, Lu F, Wang J. Sequencing of transcriptome-wide poly(A) tails by PAIso-seq. *Methods Mol Biol*, 2024, 2723: 215–232
- 57 Bazzini A A, Lee M T, Giraldez A J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science*, 2012, 336: 233–237
- 58 Guo M, Luo C, Wang Z, et al. Uncoupling transcription and translation through miRNA-dependent poly(A) length control in haploid male germ cells. *Development*, 2022, 149: dev199573
- 59 Gazestani V H, Hampton M, Abrahante J E, et al. circTAIL-seq, a targeted method for deep analysis of RNA 3' tails, reveals transcript-specific differences by multiple metrics. *RNA*, 2016, 22: 477–486
- 60 Garalde D R, Snell E A, Jachimowicz D, et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods*, 2018, 15: 201–206
- 61 Workman R E, Tang A D, Tang P S, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods*, 2019, 16: 1297–1305
- 62 Krause M, Niazi A M, Labun K, et al. *tailfindr*: alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA*, 2019, 25: 1229–1241
- 63 Gumińska N, Matylla-Kulińska K, Krawczyk P S, et al. Direct profiling of non-adenosines in poly(A) tails of endogenous and therapeutic mRNAs with Ninetails. *Nat Commun*, 2025, 16: 2664
- 64 Sippel A E, Stavrianopoulos J G, Schutz G, et al. Translational properties of rabbit globin mRNA after specific removal of poly(A) with ribonuclease H. *Proc Natl Acad Sci USA*, 1974, 71: 4635–4639
- 65 Salles F J, Strickland S. Rapid and sensitive analysis of mRNA polyadenylation states by PCR. *Genome Res*, 1995, 4: 317–321
- 66 Brouze A, Krawczyk P S, Dziembowski A, et al. Measuring the tail: methods for poly(A) tail profiling. *WIREs RNA*, 2023, 14: e1737
- 67 Jiang X, Cheng Y, Zhu Y, et al. Maternal NAT10 orchestrates oocyte meiotic cell-cycle progression and maturation in mice. *Nat Commun*, 2023, 14: 3729
- 68 Palatnik C M, Storti R V, Jacobson A. Fractionation and functional analysis of newly synthesized and decaying messenger RNAs from vegetative cells of *Dictyostelium discoideum*. *J Mol Biol*, 1979, 128: 371–395
- 69 Jacobson A. Purification and fractionation of poly(A)⁺ RNA. *Methods Enzymol*, 1987, 152: 254–261
- 70 Meijer H A, Bushell M, Hill K, et al. A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells. *Nucleic Acids Res*, 2007, 35: e132
- 71 Meijer H A, de Moor C H. Fractionation of mRNA based on the length of the poly(A) tail. *Methods Mol Biol*, 2011, 703: 123–135
- 72 Du L, Richter J D. Activity-dependent polyadenylation in neurons. *RNA*, 2005, 11: 1340–1347
- 73 Shin J, Paek K Y, Ivshina M, et al. Essential role for non-canonical poly(A) polymerase GLD4 in cytoplasmic polyadenylation and carbohydrate metabolism. *Nucleic Acids Res*, 2017, 45: 6793–6804

- 74 Beilharz T H, Preiss T. Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA*, 2007, 13: 982–997
- 75 Goodwin S, McPherson J D, McCombie W R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, 2016, 17: 333–351
- 76 Picelli S, Björklund Å K, Faridani O R, et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*, 2013, 10: 1096–1098
- 77 Eichhorn S W, Subtelny A O, Kronja I, et al. mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos. *eLife*, 2016, 5: e16955
- 78 Monzó C, Liu T, Conesa A. Transcriptomics in the era of long-read sequencing. *Nat Rev Genet*, 2025, doi: 10.1038/s41576-025-00828-z
- 79 Edmonds M, Vaughan Jr. M H, Nakazato H. Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. *Proc Natl Acad Sci USA*, 1971, 68: 1336–1340
- 80 Wu S, Zhang Y, Yao L, et al. m⁶A-modified RNAs possess distinct poly(A) tails. *J Genet Genomics*, 2023, 50: 208–211
- 81 Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. *Nat Biotechnol*, 2016, 34: 518–524
- 82 Liu H, Begik O, Lucas M C, et al. Accurate detection of m⁶A RNA modifications in native RNA sequences. *Nat Commun*, 2019, 10: 4079
- 83 Rang F J, Kloosterman W P, de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol*, 2018, 19: 90
- 84 Kim D, Lee J Y, Yang J S, et al. The architecture of SARS-CoV-2 transcriptome. *Cell*, 2020, 181: 914–921.e10
- 85 Begik O, Prysycz L P, Niazi A M, et al. Nano3P-seq: charting the coding and non-coding transcriptome at single molecule resolution. *bioRxiv*, 2025, 624491
- 86 Wang J, Ye F, Chai H, et al. Advances and applications in single-cell and spatial genomics. *Sci China Life Sci*, 2025, 68: 1226–1282

Technologies for measuring RNA poly(A) tails

JIA Zheng^{1,2†}, LIU JingWen^{1,2†} & LU FaLong^{1,2*}

¹ Key Laboratory of Integrative Physiology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

² University of Chinese Academy of Sciences, Beijing 100049, China

† Contributed equally to this work

* Corresponding author; E-mail: jflu@genetics.ac.cn

Most eukaryotic mRNAs undergo non-templated polyadenylation at their 3' termini, generating poly(A) tails composed of adenosine residues. These tails not only facilitate mRNA export from the nucleus to the cytoplasm and enhance translation, but their length and base composition also tightly regulate mRNA stability. However, the homopolymeric nature of poly(A) tails has long hindered accurate sequence determination, limiting functional and regulatory investigations. In recent years, rapid advances in high-throughput sequencing have driven major progress in diverse poly(A) tail detection technologies, advancing our understanding of tail composition and regulation. This review systematically summarizes current developments in poly(A) tail analysis, including traditional molecular biology assays, next-generation sequencing-based methods, and third-generation single-molecule sequencing-based methods. We outline the underlying principles, sample preparation and capture strategies of each technique, and compare their performance and limitations in tail-length measurement, non-A residue detection, and sensitivity. The development of poly(A) tail detection technologies has not only revealed the role of tail length in post-transcriptional regulation but also established a robust technical foundation for exploring the epigenetic regulatory information encoded within non-templated poly(A) tail sequences.

poly(A) tail, translation, RNA stability, poly(A) tail non-A residues, poly(A) tail epigenetic regulation, RNA epigenetics

doi: [10.1360/SSV-2025-0143](https://doi.org/10.1360/SSV-2025-0143)