ADVANCED REVIEW

# Measuring the tail: Methods for poly(A) tail profiling

Aleksandra Brouze[1] | Paweł Szczepan Krawczyk[2] |
Andrzej Dziembowski[1,2,3] | Seweryn Mroczek[1,2]

[1]Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, Warsaw, Poland

[2]Laboratory of RNA Biology, International Institute of Molecular and Cell Biology, Warsaw, Poland

[3]Department of Embryology, Faculty of Biology, University of Warsaw, Warsaw, Poland

**Correspondence**
Andrzej Dziembowski, Laboratory of RNA Biology, International Institute of Molecular and Cell Biology, 02-109, Warsaw, Poland.
Email: adziembowski@iimcb.gov.pl

Seweryn Mroczek, Laboratory of RNA Biology, International Institute of Molecular and Cell Biology, 02-109 Warsaw, Poland.
Email: smroczek@iimcb.gov.pl

**Funding information**
Fundacja na Rzecz Nauki Polskiej; Narodowe Centrum Nauki, Grant/Award Numbers: UMO-2020/37/N/NZ2/02893, UMO-2020/38/E/NZ2/00372

**Edited by:** Jeff Wilusz, Editor-in-Chief

**Abstract**

The 3′-end poly(A) tail is an important and potent feature of most mRNA molecules that affects mRNA fate and translation efficiency. Polyadenylation is a posttranscriptional process that occurs in the nucleus by canonical poly(A) polymerases (PAPs). In some specific instances, the poly(A) tail can also be extended in the cytoplasm by noncanonical poly(A) polymerases (ncPAPs). This epitranscriptomic regulation of mRNA recently became one of the most interesting aspects in the field. Advances in RNA sequencing technologies and software development have allowed the precise measurement of poly(A) tails, identification of new ncPAPs, expansion of the function of known enzymes, discovery and a better understanding of the physiological role of tail heterogeneity, and recognition of a correlation between tail length and RNA translatability. Here, we summarize the development of polyadenylation research methods, including classic low-throughput approaches, Illumina-based genome-wide analysis, and advanced state-of-art techniques that utilize long-read third-generation sequencing with Pacific Biosciences and Oxford Nanopore Technologies platforms. A boost in technical opportunities over recent decades has allowed a better understanding of the regulation of gene expression at the mRNA level.

This article is categorized under:
    RNA Methods > RNA Analyses In Vitro and In Silico

**KEYWORDS**
3′-end tailing, poly(A) tail, regulation of gene expression, RNA, RNA sequencing

## 1 | INTRODUCTION

Polyadenylation is the process by which nontemplated adenosines are added to the 3′-end of mRNA, which occurs post-transcriptionally in the nucleus and is performed by canonical poly(A) polymerases. This modification affects almost all mRNA molecules except replication-dependent histone mRNA. The poly(A) tail is necessary for the transport of mRNA molecules from the nucleus to the cytoplasm (Fuke & Ohno, 2008) and enhances mRNA stability, as deadenylation is the first step in RNA decay pathways (Wang & Kiledjian, 2001) and translatability through an interaction between

poly(A)-binding protein 1 (PABPC) to the 5′ cap-binding translation initiation factor eukaryotic translation initiation factor eIF4G (Wakiyama et al., 2000; Wells et al., 1998; Xiang & Bartel, 2021).

It is now commonly known that poly(A) tails are not merely unchanging, stable features of the 3′-end of mRNA. The processes of polyadenylation, the regulation of poly(A) tail length, and its dynamic changes and modifications are controlled by a multitude of *cis*- and *trans*-regulatory mechanisms and enzymes (Figure 1). Consequently, 3′-end tailing is a much more complex and impactful process than previously thought.

Nuclear polyadenylation is preceded by endonucleolytic cleavage of the 3′-end of mRNA with signals that designate the cleavage site within the transcript body between the highly conserved AAUAAAA hexamer and a U- or GU-rich downstream sequence element (DSE). These elements serve as binding platforms for cleavage and polyadenylation specificity factor (CPSF) and cleavage specificity factor (CstF), respectively. After the recruitment of all necessary factors, including CPSF, CstF, cleavage factor I (CFI), CFII, and poly(A) polymerase, the cleavage of pre-mRNA and finally, 3′-end tailing occurs (Figure 1). Nascent homopolymers are then bound by poly(A) binding proteins (PABPs, PABPN1 in higher species; Proudfoot et al., 2002; Wahle & Rüegsegger, 1999).

However, more than half of human genes possess multiple polyadenylation sites (Tian et al., 2005). The usage of an alternative polyadenylation site (APA), constitutes an important regulatory mechanism. It can affect either the coding sequence, giving rise to different mRNA isoforms, or the 3′-untranslated region (UTR), which influences mRNA fate. Sequences within the 3′-UTR facilitate the binding of miRNAs or RNA binding proteins (Di Giammartino et al., 2011;
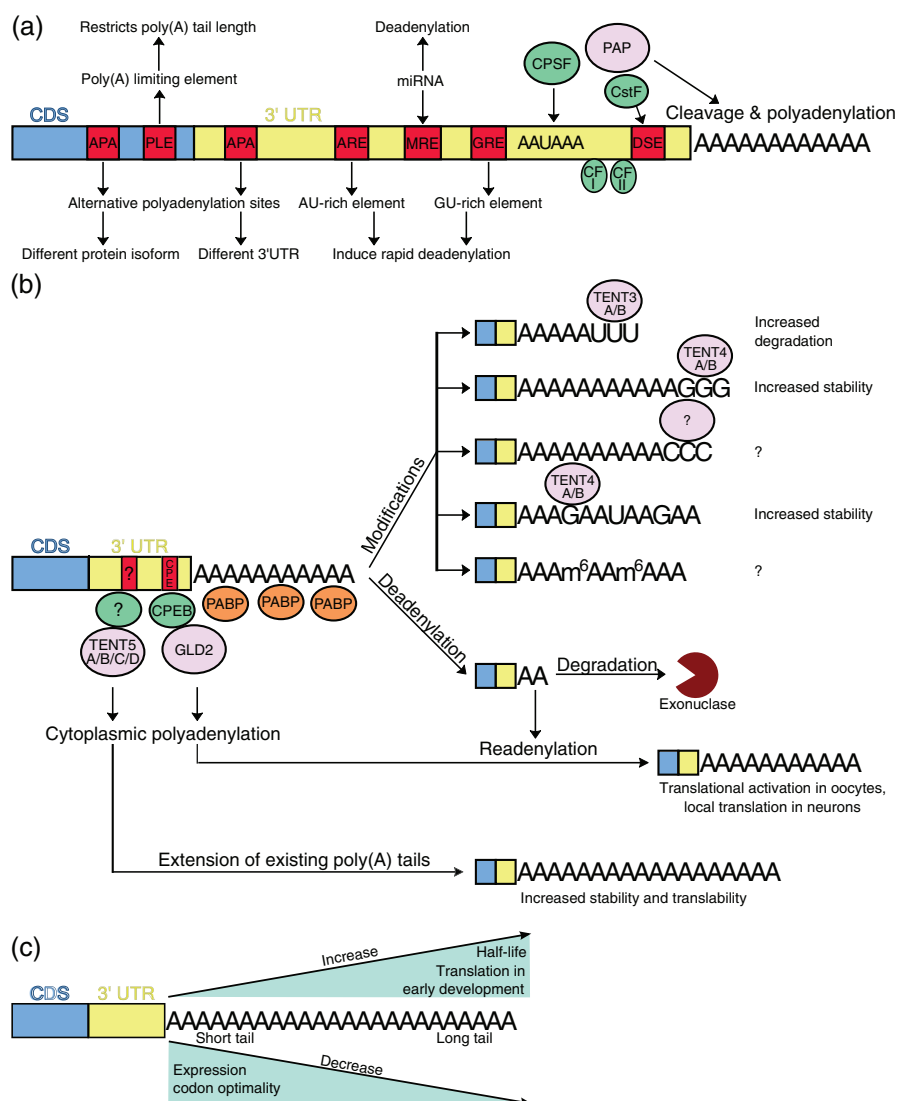


**FIGURE 1** The poly(A) tail is a dynamic feature of mRNA. (a) *Cis*- and *trans*-regulatory elements affect the poly(A) tail. (b) Possible fates of mRNA poly(A) tail in the cytoplasm. (c) Relationship between poly(A) tail length and different aspects of gene expression.

Tian & Manley, 2016). One of the best known examples of alternative polyadenylation is the switch from the distal to proximal poly(A) site in mRNA that encodes immunoglobulin M (IgM), resulting in a transition from the membrane-bound to secreted form of this protein during B cells differentiation to plasma cells (Alt et al., 1980; Early et al., 1980).

The choice of polyadenylation site and nuclear regulation of the mRNA poly(A) tail length depends on regulatory sequences within the transcript body. The presence of poly(A) limiting element (PLE) in the 5′ end of the terminal exon results in short poly(A) tails in albumin (Gupta et al., 1998) and transferrin (Gu et al., 1999) mRNAs in Xenopus that do not affect their decay rate. It also increases the efficiency of 3′-end processing (Peng et al., 2005). AU-rich elements (AREs) and GU-rich elements (GREs) in the 3′-UTR are linked with faster deadenylation and shortening of the poly(A) tail, leading to transcript degradation (C. Y. Chen et al., 2001; Vlasova et al., 2008; Xu et al., 1997). Similarly, miRNA binding to miRNA regulatory elements (MREs) in the 3′-UTR induces the deadenylation and degradation of mRNA (L. Wu et al., 2006), playing an important role in the clearance of maternal transcripts during zebrafish early development (Giraldez et al., 2006). Interestingly, miRNA- and ARE-mediated deadenylation can be connected. Dicer, a major component of miRNA processing machinery, is necessary for ARE-mediated mRNA degradation, probably due to the requirement of miRNA for ARE-RNA degradation. Indeed, miR16 targeted by ARE-binding protein tristetraprolin (TTP) is necessary for ARE-RNA turnover (Jing et al., 2005). Additionally, miRNA can interact with RNA-binding proteins recognizing nearby motifs by regulating their activity (like Cpeb1), or conversely, RBP can either enhance miRNA activity or inhibits it (like HuR; Ciafrè & Galardi, 2013).

The length of the poly(A) tail and its dynamic changes play a crucial role in the life of the mRNA molecule (Figure 1b). It must be sufficiently long to facilitate mRNA export to the cytoplasm (Dower et al., 2004; Tudek et al., 2018). During the lifetime of mRNA, the poly(A) tail shortens. The bulk mRNA degradation pathway in eukaryotic cells begins with tail shortening mainly by CCR4-NOT and PAN2/PAN3 deadenylases. Besides "canonical" poly(A) polymerases involved in cleavage-dependent polyadenylation of nascent transcripts in the nucleus, the human genome encodes 11 "noncanonical" poly(A) polymerases (ncPAPs) and poly(U) polymerases (PUPs). Contrary to the canonical poly(A) polymerases, they are involved in extending the already existing poly(A) tail independently of transcript cleavage, they vary in nucleotide and substrate specificity, and some of these enzymes reside in the cytoplasm (Figure 1b; Liudkovska & Dziembowski, 2021). In oocytes and during early embryonic development, a fraction of deadenylated transcripts is not degraded but rather stored in a translationally inactive state (Barnard et al., 2004; J. H. Kim & Richter, 2006). After appropriate stimulation, Gld2 poly(A) polymerase catalyzes re-adenylation and consequently the translational activation of these transcripts. This process is orchestrated by cytoplasmic polyadenylation element-binding protein (CPEB), which recognizes a cytoplasmic polyadenylation element (CPE) that is within the 3′-UTR of these transcripts (Richter, 2007; Villalba et al., 2011; Weill et al., 2012). In other cases, the extension of already existing poly(A) tails in the cytoplasm can increase stabilization and translation, such as in the case of immunoglobins or collagen transcripts in B cells (Bilska et al., 2020) and osteoblasts (Gewartowska et al., 2021), respectively. Tail elongation is catalyzed by proteins from the terminal nucleotidyltransferase 5 (TENT5) family of poly(A) polymerases, but the regulatory elements that drive this process remain unknown.

Finally, the poly(A) tail does not need to be composed only of adenosines. Recent studies revealed that 3′ ends of poly(A) tails are commonly uridylated (resulting in preferentially short poly[A] tails), guanylated (resulting in preferentially long poly[A] tails), or less frequently cytidilated (Chang et al., 2014), having diverse effects on mRNA stability. Uridylation by TENT3A and TENT3B (terminal uridylyltransferase 4 [TUT4] and TUT7) marks RNA for degradation, whereas guanylation by TENT4A/B protects mRNA from deadenylation, consequently increasing its stability (Lim et al., 2018). These modifications were also detected in the tail body (Y. Liu, Nie, Liu, & Lu, 2019b). Recently, a study of *Trypanosoma brucei* revealed the presence of methylated adenosines (m6A) within the poly(A) tail that protected them from deadenylation and stabilized the mRNA molecule (Viegas et al., 2022), thus revealing a new mechanism of the regulation of gene expression.

Poly(A) tails of steady-state mRNAs are shorter than previously assumed (Chang et al., 2014; Lima et al., 2017; Subtelny et al., 2014). Highly expressed transcripts with high codon optimality possess relatively short poly(A) tails (Lima et al., 2017). Poly(A) tail length correlates positively with half-lives, while with translation efficiency only in oocytes and early embryonic development (Chang et al., 2014; Lim et al., 2016; Subtelny et al., 2014; Figure 1c). Rapidly increasing interest in this area of research demands precise and high-throughput tools to dissect and understand the nature and dynamics of mRNA poly(A) tail lengths. Historically, the first approaches that were based on RNaseH/oligo (dT) cleavage, followed by Northern blot analysis or the polymerase chain reaction (PCR)-based amplification of poly(A) tails, provided information on only single transcripts and allowed only the observation of polyadenylation shift rather than subtle changes. These methods also did not provide any information about poly(A) tail length. The

development of high-throughput approaches, such as next-generation sequencing (NGS), opened new possibilities for polyadenylation studies. The most common current approaches are based on the sequencing of appropriately prepared mRNA 3′ end libraries on an Illumina platform, with nucleotide-level resolution. However, these estimations can be questionable because long homopolymer amplification is challenging, and PCR bias-prone short-read technology deprives us of information about very long poly(A) tails. Thus, single-molecule, long-read sequencing with Pacific BioSciences (PacBio) or Oxford Nanopore Technologies (ONT) platforms is recently gaining versatility and enabling non-PCR-biased poly(A) tail length estimations that can be accompanied by information about differential expression, splicing, and RNA modifications.

The present review provides details of the broad range of methods that are used for mRNA 3′-terminome analyses, from the single-transcript level to transcriptome-wide studies that utilize second- and third-generation sequencing.

## 2 | THE BEGINNINGS OF POLY(A) TAIL LENGTH ESTIMATION

The first method that was applied to study poly(A) tail length was the RNase H/oligo(dT) assay, originally designed based on rabbit globin mRNA (Sippel et al., 1974). RNase H specifically hydrolyzes phosphodiester bonds in RNA only in RNA:DNA hybrids (Sippel et al., 1974). Thus, oligo(dT) hybridization to the poly(A) tail results in the recognition of this duplex by RNase H and subsequent removal of the poly(A) tail (Murray & Schoenberg, 2008). The length of poly(A) tails is assessed by Northern blot by comparing samples that are treated with RNase H in the presence or absence of oligo(dT) (Figure 2a). The difference in electrophoretic mobility reflects poly(A) tail length (Murray & Schoenberg, 2008; Sallés et al., 1999). However, differences in poly(A) tail length in the case of larger mRNAs (>2 kb) can be difficult to detect. Thus, to enable higher resolution, RNA is hybridized, along with oligo(dT), with a second antisense deoxyoligonucleotide near the 3′ end to trim mRNA (Murray & Schoenberg, 2008; Sallés et al., 1999).

Another approach to study poly(A) tails is the PCR amplification of 3′ ends. Classic PCR-based assays for determining poly(A) tail lengths are the rapid amplification of cDNA ends poly(A) test (RACE-PAT; also called the poly [A] length assay or PCR poly[A] test; Salles et al., 1992; Sallés et al., 1999) and ligase-mediated poly(A) test (LM-PAT). In RACE-PAT, total RNA is incubated with an oligo(dT) anchor with a GC-rich primer sequence on its 5′ end. It hybridizes along with the entire poly(A) tail length, eliciting heterogeneous cDNA populations. RNA is reverse transcribed from an oligo(dT) anchor primer, and the poly(A) tail of the gene of interest is PCR amplified with gene-specific and oligo(dT) anchor primers (Figure 2b). The PCR products are analyzed by gel electrophoresis. Short poly(A) tails generate a band. In the case of longer poly(A) tails, amplicons with diverse poly(A) tail lengths create a smear on the gel (Murray & Schoenberg, 2008; Sallés et al., 1999). The LM-PAT (Sallés & Strickland, 1995) differs from RACE-PAT only in the oligo(dT) hybridization step design. Total RNA is hybridized with 5′-phosphorylated oligo(dT) molecules in the presence of T4 DNA ligase, resulting in full coverage of the poly(A) tail. Therefore, the oligo(dT) anchor ligates to the 5′ end of the 3′-most oligo(dT) (Figure 2c; Murray & Schoenberg, 2008; F. J. Sallés & Strickland, 1999; Sallés et al., 1999). Comparing these two methods, LM-PAT is more sensitive to subtle changes in poly(A) tail length (Sallés et al., 1999). Both the RNase H- and PCR-based approaches have advantages and limitations. Polymerase chain reaction-based methods are principally faster, less laborious, and require a smaller amount of starting material than the RNase H assay. They are also relatively insensitive to internal poly(A) priming (annealing of oligo(dT) primer to poly(A) stretch within transcript body). However, the DNA polymerase is unable to amplify the exact number of nucleotides in a long homopolymer stretch in each template molecule, mostly shortening the homopolymer length and generating products with a wide range of sizes (PCR bias), whereas the RNase H method provides direct information about RNA without reverse transcription (RT)- or PCR bias.

Modifications of the classic PAT assay are based on different primers that are anchored to the 3′ end of RNA. The extension poly(A) test (ePAT; Jänicke et al., 2012) is based on the extension of the 3′ end of RNA by Klenow polymerase on a DNA template (Figure 2d). The splint-mediated PAT (sPAT; Minasaki et al., 2014) utilizes splint ligation of the 3′ end of mRNA with an RNA anchor (Figure 2e). This approach is more sensitive than ePAT, requires less starting material and fewer amplification cycles, and is less biased toward shorter poly(A) tails (Minasaki et al., 2014). Other variations of the PAT assay include enzymatic G-tailing (Kusov et al., 2001) or G/I-tailing (Bazzini et al., 2012) of the 3′ end. The added tag serves as an anchor for a primer that consists of a C stretch followed by several Ts that are used for cDNA synthesis and PCR, accompanied by a gene-specific primer (Kusov et al., 2001). Moreover, the fluorescent labeling of one of the primers allows for separation of the PCR product by capillary electrophoresis, resulting in very high resolution, referred to as the high-resolution poly(A) test (Hire-PAT; Bazzini et al., 2012; Figure 2f).
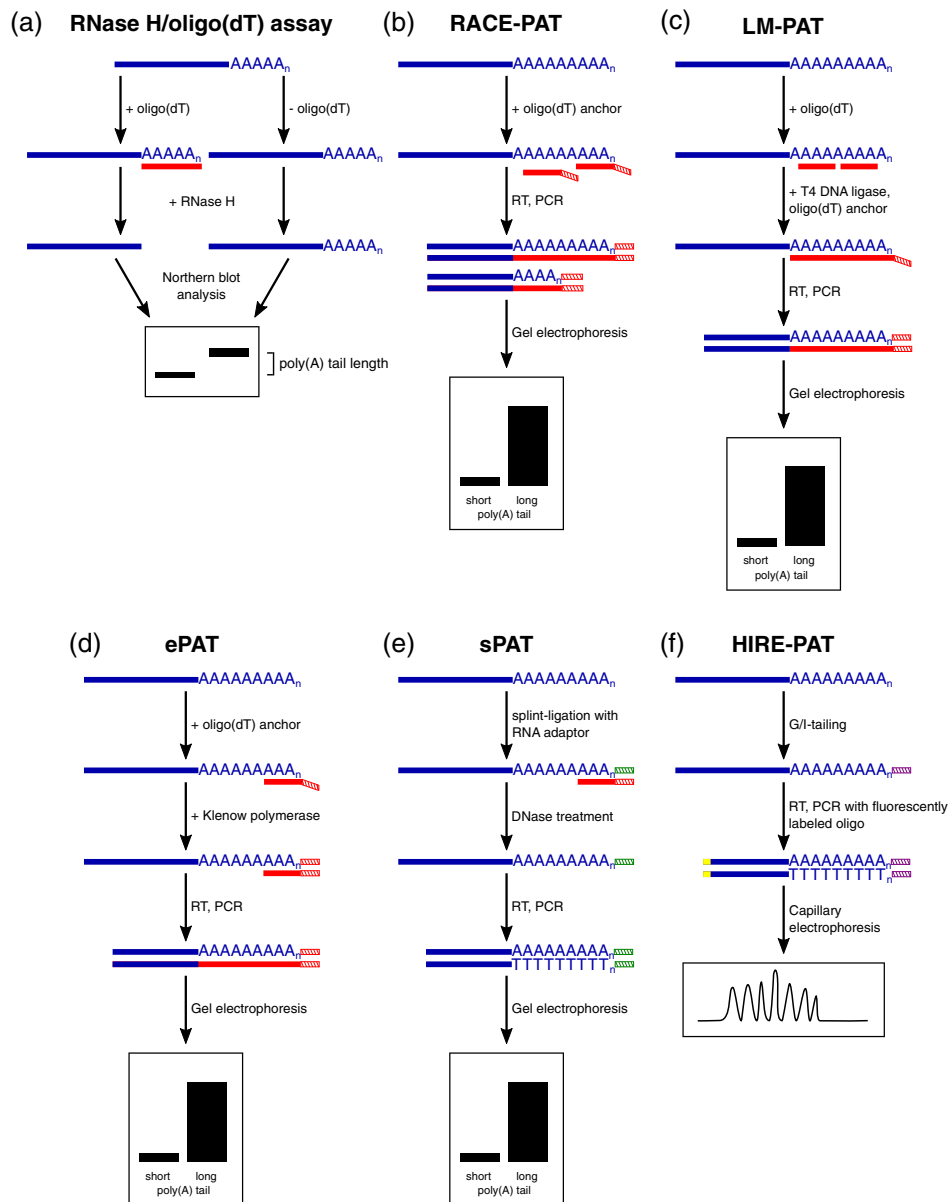
**FIGURE 2** Overview of classic methods used to study poly(A) tail length - (a) RNase H/oligo(dT) assay, and variations of PCR-based Poly(A) Test (b) RACE-PAT, (c) LM-PAT, (d) ePAT, (e) sPAT, (f) HIRE-PAT. The red line indicates the oligo(dT) sequence. Red hatched lines indicate the DNA 30 adapter sequence. Green hatched lines indicate the RNA 30 adapter sequence. Violet hatched lines indicate the G/I tag sequence. The yellow line indicates a fluorescent tag.

Different but less frequently used is a PCR-based approach called circular RT-PCR (cRT-PCR; Couttet et al., 1997). Total RNA is decapped, which allows for RNA circularization by 5′- and 3′-end ligation. After RT, nested PCR is performed to obtain products that cover the 3′–5′ junction, including poly(A) tails. The PCR products are visualized by gel electrophoresis in the presence of samples that are deadenylated at the beginning with RNase H to determine the poly(A) tail length (Couttet et al., 1997; Suh et al., 2006).

In conclusion, these methods contributed significantly to studies of poly(A) tails, thus opening a new, broad, and fascinating area of research. Although they are still commonly used for the validation of novel methods, single-transcript analyses, and pilot experiments, they have many limitations and restricted applications. They only allow single-transcript analyses of the poly(A) tail length shift, providing no precise poly(A) tail length or composition information. Consequently, the need for more accurate and higher throughput analyses led to the development of novel methods.

# 3 | FIRST POLYADENYLATION STUDIES AT THE TRANSCRIPTOME-WIDE SCALE WITH MICROARRAYS

The poly(A) tail is a distinctive feature of RNA that can be exploited for length-based separation. Palatnik et al. (1979) demonstrated RNA fractionation based on the difference in poly(A) tail length using poly(U) chromatography, followed by the thermal elution of RNA. Alternatively, oligo(dT) chromatography and elution with decreasing salt concentrations can be applied (Jacobson, 1987). Such approaches were initially used only with single-transcript analyses (Jacobson, 1987; Palatnik et al., 1979), but the development of microarray technology-facilitated analyses of polyadenylation status at the transcriptome-wide scale. RNA that is bound to biotinylated oligo(dT) and enriched on paramagnetic beads is eluted by buffers at decreasing salt concentrations. The polyadenylated and oligoadenylated fractions are then analyzed by microarrays (Meijer et al., 2007; Meijer & de Moor, 2011) to identify transcripts in each fraction. This method was used to show the polyadenylation of transcripts encoding cyclin B1, c-mos, and Aurora-A kinases during oocyte maturation in *Xenopus laevis* and in the mouse NIH3T3 fibroblast cell line (Meijer et al., 2007) and to study the effects of transcription dynamics on poly(A) tails (Slobodin et al., 2020). However, such a method deprives RNA with poly(A) tails that are shorter than 25 nt, whereas the abundance of short poly(A) tail transcripts constitutes 25% of expressed genes (Meijer et al., 2007). A comparative analysis of mRNA that was purified with cap-binding eIF4E protein and oligo(dT) revealed a population of transcripts that were purified only or preferentially using the cap with poly(A) tails that were shorter than 30 nt (Choi & Hagedorn, 2003).

Poly(U) chromatography with thermal elution coupled with microarray analysis was used to study CPEB-mediated cytoplasmic polyadenylation in neurons (Du & Richter, 2005) and identify transcripts that were polyadenylated by the cytoplasmic poly(A) polymerase GLD4 (Shin et al., 2017). Using this method, a positive correlation was found between poly(A) tail length and Pab1 and ribosome density in yeast (Beilharz & Preiss, 2007). To study differential polyadenylation at different cell cycle stages in HeLa cells (Novoa et al., 2010), a mixed approach was applied. Poly(U) chromatography with elution at low temperature selects RNAs with short poly(A) tails, whereas the oligo(dT) approach isolates all polyadenylated RNAs. These two fractions were compared (Novoa et al., 2010).

These methods are able to detect changes in poly(A) tail lengths in a transcriptome-wide manner, but they do not provide information about the exact poly(A) tail length. Thus, subtle changes in poly(A) tail length can be missed, and information about very short poly(A) tails is lacking. Moreover, the utilization of microarrays has several drawbacks, especially compared with RNA sequencing (RNA-Seq). Only genes to which probes are designed can be detected, and nonspecific hybridization and cross-hybridization impede the results (S. Zhao et al., 2014). Overall, although they made a step forward compared with previous single-transcript approaches, leading to many biologically relevant observations, they still do not meet the rising demands of the research community.

# 4 | POLY(A) TAIL LENGTH MEASUREMENT BY SEQUENCING

The development of NGS with the RNA-Seq technique raised new possibilities in transcriptomic studies. Starting from simple differential gene expression analysis at a transcriptome-wide scale, the applications of high-throughput sequencing are still rising, including polyadenylation status analyses. Here, we review methods of poly(A) tail length measurements using Illumina, PacBio, and ONT sequencing platforms.

## 4.1 | Illumina-based approaches: TAIL-Seq, PAL-Seq, and m-TAIL-Seq

Since the development of RNA-Seq in 2008, the Illumina platform has been a leading technology in the field. The basis of Illumina RNA-Seq is the generation of libraries of short cDNA fragments with adapters on one or both ends from mRNA-enriched or rRNA-depleted RNA, which are then amplified by PCR. cDNA molecules are clustered on a flow cell and sequenced from one end (single-end sequencing) or both ends (paired-end sequencing) by reading the fluorescence signal from a 3′-blocked fluorescently labeled nucleotide that is incorporated during DNA synthesis (so-called "sequencing-by-synthesis"; Stark et al., 2019). This approach is most commonly used for differential expression analysis, but library preparation and data analysis workflows are highly adjustable and compatible with various applications, such as protocols for poly(A) tail sequencing.

The first approaches to measure poly(A) tail lengths at genomic scale were published in 2014, including poly(A)-tail length profiling by sequencing (PAL-Seq; Subtelny et al., 2014) and TAIL-Seq (Chang et al., 2014). The latter was subsequently modified to become mRNA TAIL-Seq (mTAIL-Seq; Lim et al., 2016). Both PAL-Seq and TAIL-Seq rely on the preparation of sequencing libraries with intact mRNA 3′ ends, including the poly(A) tail, which is further sequenced on the Illumina platform. Although these approaches differ in their sequencing procedures and computational approaches that are used for the determination of poly(A) tail length, they utilize very similar experimental procedures (see overview in Figure 3). RNA is first ligated with a biotinylated 3′ DNA adapter and partially digested with RNase T1 that cleaves RNA after guanine residues, thereby leaving intact poly(A) tails. Subsequent pull-down on streptavidin beads allows the enrichment of only 3′ end fragments. Next, size selection by gel electrophoresis and ligation with a 5′ adapter provide complete templates for reverse transcription, library amplification, and sequencing on the Illumina platform (Chang et al., 2014; Lim et al., 2016; Subtelny et al., 2014). The order of steps slightly differs among these procedures, but one of the main differences between them is the form of the 3′ adapter. In TAIL-Seq, it is a single-stranded DNA adapter, thus introducing the rRNA-depletion step as necessary for the analysis of only poly(A) + molecules (Chang
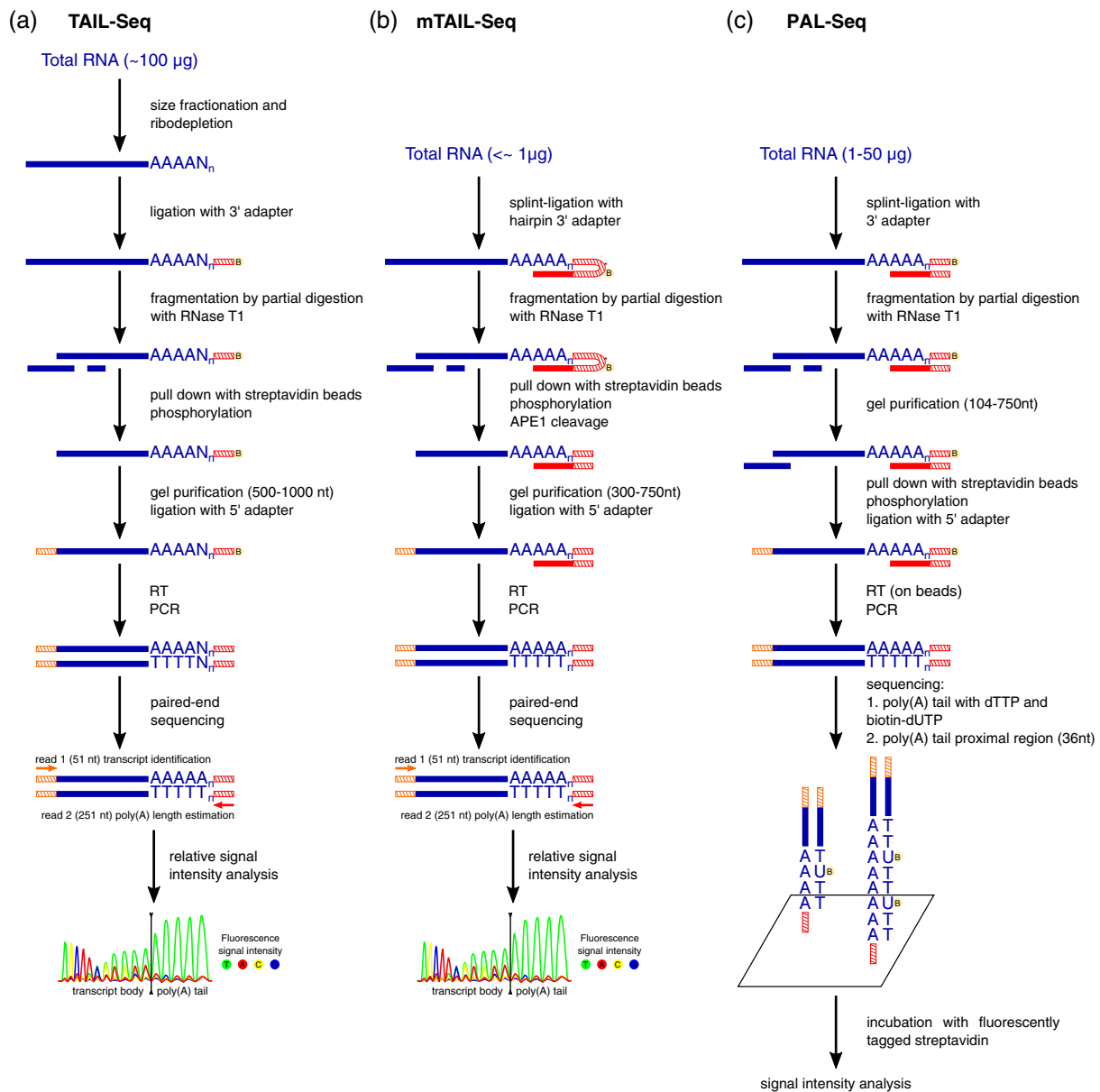


**FIGURE 3** Comparison of TAIL-Seq, mTAIL-Seq, and PAL-Seq protocols for mRNA poly(A) tail sequencing. The red line indicates the oligo(dT) sequence. Red hatched lines indicate the 3′ adapter sequence. Orange hatched lines indicate the 5′ adapter sequence. The "B" in the yellow circle indicates biotin. Asterisks indicate an abasic site in the 3′ adapter.

et al., 2014). In PAL-Seq, the adapter is also a single-stranded DNA, but a second DNA oligonucleotide that is complementary to both the poly(A) tail and 3′ adapter fragment facilitates efficient ligation (i.e., splint ligation) to exclusively mRNA molecules, thus eliminating the need for ribodepletion (Subtelny et al., 2014). In mTAIL-Seq, the adapter and splinter create one hairpin oligonucleotide that improves ligation efficiency. Additionally, two abasic sites that are introduced in this hairpin enable an efficient cut from streptavidin beads by apurinic/apyrimidinic endonuclease 1 (APE1; Lim et al., 2016).

A major difference between PAL-Seq and TAIL-Seq lies in the sequencing procedure and bioinformatic analysis. PAL-Seq is compatible only with the discontinued Illumina Genome Analyzer sequencer, whereas TAIL-Seq was successfully applied on Hiseq2000 and MiSeq machines. In the case of PAL-Seq on a flow cell, after cluster generation, a sequencing primer that is complementary to the 3′ adapter is extended with a mixture of dTTP and biotinylated dUTP to label the poly(A) tail, and standard Illumina chemistry is subsequently applied to sequence the fragment upstream of poly(A), thereby allowing transcript identification. Because of incubation with fluorochrome-conjugated streptavidin, the poly(A) tail length is estimated from the fluorescence signal intensity that is normalized to cluster density, assuming that the amount of incorporated biotin-dUTP increases proportionately to poly(A) tail length. In the original protocols for TAIL-Seq and mTAIL-Seq, paired-end sequencing allowed identification of the transcript from read 1 (51 nt) and examination of the poly(A) tail with read 2 (231 nt), but these read lengths can be customized. Standard Illumina sequencing fails when analyzing long homopolymer tracts. Therefore, the authors of TAIL-Seq developed an algorithm that measures the length of the poly(A) tail using fluorescence quantification. Importantly, this requires modification in the sequencer's control software, which allows storing intermediate CIF files that contain cluster fluorescence intensity information. When analyzing read 2, the polyadenylation site is detected by analyzing changes in relative fluorescence, defined as the ratio of the signal from thymidine (i.e., in the second read, the complementary cDNA strand is sequenced, thus obtaining the poly[T] homopolymer) to the signal from the different base. After the long poly(T) stretch, a signal from thymidine is still detectable even when a non-T base is sequenced, but the relative fluorescence changes, leading to the identification of the poly(A) tail start site and allowing calculation of the poly(A) tail length (Chang et al., 2014; Lim et al., 2016). In both approaches, accurate measurements are based on comparisons to signals from spike-ins with defined poly(A) tail lengths (Subtelny et al., 2014).

Although these methods provide nucleotide-resolution poly(A) tail length measurements on a genomic scale, they clearly have limitations (Table 1). Both reverse transcription and PCR amplification steps can introduce bias, leading to underestimations of the length of poly(A) tails. Because poly(A) spike-ins are generated by PCR or in vitro polyadenylation, they provide a range of poly(A) tail lengths rather than exact lengths. TAIL-Seq initially requires a large amount of RNA (∼100 μg; Lim et al., 2016) and expensive rRNA depletion (Chang et al., 2014). In mTAIL-Seq and PAL-Seq, the splint ligation step reduces the amount of starting material to as low as 100 ng (Lim et al., 2016) or 1–50 μg (Subtelny et al., 2014) of RNA and captures specifically poly(A) + mRNAs without a need for ribodepletion. However, a disadvantage of splint ligation is the absence of data on 3′ end modifications, such as uridylation and guanylation, which can still be identified with TAIL-Seq (Chang et al., 2014). Because very short and modified poly(A) tails are not captured, longer median poly(A) lengths are also obtained (Lim et al., 2016). As a result, there are discrepancies in median poly(A) tail lengths between TAIL-Seq and PAL-Seq (Table 2). Although all of these methods utilize the Illumina sequencing system, PAL-Seq requires manipulations on a flow cell, thereby introducing an additional imaging step, and also requires the now-discontinued Genome analyzer II sequencer, which in reality makes this method impractical. Both TAIL-Seq and mTAIL-Seq can be performed on the Illumina HiSeq or MiSeq platform. Because of the higher sensitivity of mTAIL-Seq, a lower sequencing depth from small-scale experiments on MiSeq can provide sufficient data while reducing the overall cost of the experiment (Lim et al., 2016). Importantly, both methods present significant bioinformatic challenges and are difficult to implement. For PAL-seq, raw signal intensities from the flow cell are collected and further processed to calculate poly(A) length for each cluster on the flow cell. In the case of TAIL-seq and its derivatives, raw intensities are analyzed by the provided software (tailseeker; https://github.com/hyeshik/tailseeker), which has multiple dependencies and is no longer maintained. There is also a known issue in the latest version of tailseeker that is related to the patterned imbalance of fluorescence signals among the channels and problems with signal normalization. To the best of our knowledge, these issues remain unresolved, as this software is not currently being developed.

The utilization of TAIL-Seq and PAL-Seq has dramatically expanded our knowledge of 3′ ends of mRNA. Poly(A) tails are shorter than previously considered (>200 nt; Weill et al., 2012), with a median poly(A) tail length of ∼60 nt (Chang et al., 2014) or in the range of 67–96 nt (Subtelny et al., 2014) in various mammalian cells (summarized in Table 2). Their lengths in orthologous genes are conserved among species (Subtelny et al., 2014). Indeed, shorter

**TABLE 1** Summary of major differences between the most common 3′-end sequencing approaches

| Feature | TAIL-Seq (Chang et al., 2014) | mTAIL-Seq (Lim et al., 2016) | PAL-Seq (Subtelny et al., 2014) | FLAM-Seq (Legnini et al., 2019) | PAIso-Seq (Y. Liu, Nie, et al., 2019) | DRS (Garalde et al., 2018) |
|---|---|---|---|---|---|---|
| RNA preparation | Ribodepletion | Not required | Not required | Poly(A) + selection | Not required | Poly(A) + selection |
| Amount of starting material | 100 µg | <1 µg | 1–50 µg | 500 ng–10 µg | ≤100 ng | 500 ng of polyadenylated RNA |
| Sequencing platform | Illumina HiSeq/MiSeq | Illumina HiSeq/MiSeq | Illumina Genome Analyzer | Pacific BioSciences | Pacific BioSciences | Oxford Nanopore Technologies |
| 3′ adapter addition | Ligation of biotinylated single-stranded DNA oligo | Splint ligation of biotinylated hairpin DNA oligo | Splint ligation of biotinylated single-stranded DNA oligo in the presence of splint DNA oligo | Enzymatically added G/I tail serving as anchor for oligo | 3′-end extension of template oligo containing T stretch | Splint ligation of DNA oligo containing T stretch |
| Fragmentation | Partial digestion with RNase T1 | Partial digestion with RNase T1 | Partial digestion with RNase T1 | None | None | None |
| PCR bias | + | + | + | + | + | − |
| RT bias | + | + | + | + | + | − |
| Oligo(dT) selection bias | − | − | − | + | − | + |
| Error rate | Low | Low | Low | Low | Low | High |
| Read length | Short | Short | Short | Long | Long | Long |
| Detection of modifications in 3′-end of poly(A) tail | + | − | − | − | − | − |
| Detection of modifications within poly(A) tail | − | − | − | + | + | − |

**TABLE 2** Comparison of results obtained with mRNA 3′-end sequencing methods, including median poly(A) tail length and correlations between poly(A) tail length and mRNA expression, half-life, and translational efficiency

| Model | Method | Median poly(A) | | Correlation with expression | Correlation with half-life | Correlation with translation efficiency |
| | | Total [nt] | Intragenic [nt] | | | |
|---|---|---|---|---|---|---|
| HeLa | TAIL-Seq (Chang et al., 2014) | 59 | 60 | | | $R_p = 0.058$noncano |
| | PAL-Seq (Subtelny et al., 2014) | 67.5 | 82.6 | $R_s = -0.053$ | $R_s = -0.048$ | $R_s = -0.1$ |
| | Poly(A)-Seq (Yu et al., 2020) | 95.5 | | | | |
| HeLa S3 | FLAM-Seq (Legnini et al., 2019) | 81 | 87 | $R_p = -0.18$ | $R_p = -0.13$ | $R_p = -0.1$ |
| HEK923T | PAL-Seq (Subtelny et al., 2014) | 75.3 | 88.4 | $R_s = -0.23$ | | $R_s = 0.07$ |
| HCT116 | Poly(A)-Seq (Yu et al., 2020) | 101.5 | | | | |
| K562 | DRS (Maier et al., 2020) | 94.75 | | | $R_s = 0.48$ | |
| Human iPSC | FLAM-Seq (Legnini et al., 2019) | 97 | 111.5 | $R_p = -0.19$ | | |
| Cerebral organoids | FLAM-Seq (Legnini et al., 2019) | 104 | 121.5 | $R_p = -0.32$ | | |
| NIH3T3 | TAIL-Seq (Chang et al., 2014) | 60 | 61 | $R_p = 0.076$ | $R_p = 0.364$ | $R_p = 0.009$ |
| | PAL-Seq (Subtelny et al., 2014) | 95.7 | 107.8 | $R_s = -0.23$ | $R_s = -0.16$ | $R_s = -0.04$ |
| Mouse liver | PAL-Seq (Subtelny et al., 2014) | 66.6 | 75.3 | $R_s = 0.085$ | | $R_s = 0$ |
| Mouse | Nano3P-Seq (Begik et al., 2021) | 100.7 | 106.4 | | | |
| Rat liver | PAIso-Seq (Y. Liu, Nie, Nie, et al., 2019) | 63 | | | | |
| *Arabidopsis* leaf | PAL-Seq (Subtelny et al., 2014) | 50.7 | 58.1 | | | |
| *Arabidopsis* | DRS (Parker et al., 2020) | 68 | | $R_s = -0.34$ | | |
| *Drosophila* S2 | PAL-Seq (Subtelny et al., 2014) | 50.4 | 63.3 | | | |
| *Drosophila* | mTAIL-Seq (Lim et al., 2016) | 60 | 58 | | | NA |
| Immature oocyte | | 75 | 76 | | | $R_s = 0.306$ |
| Mature oocyte | | 73 | 70 | | | $R_s = 0.638$ |
| Activated egg | | | | | | |

**TABLE 2** (Continued)

| Model | Method | Median poly(A) [nt] | | Correlation with expression | Correlation with half-life | Correlation with translation efficiency |
|---|---|---|---|---|---|---|
| | | Total [nt] | Intragenic [nt] | | | |
| *S. cerevisiae* | PAL-Seq (Subtelny et al., 2014) | 26.7 | 33.1 | $R_s = -0.44$ | $R_s = 0.048$<br>$R_s = -0.094$<br>$R_s = 0.044$<br>$R_s = -0.11$<br>$R_s = 0.045$<br>$R_s = -0.44$<br>$R_s = 0.23$<br>$R_s = -0.35$ | $R_s = -0.12$ |
| | DRS (Tudek et al., 2021) | 40 | | $R_p = -0.53$ | $R_p = -0,22$<br>$R_p = -0.39$ | |
| | Nano3P-Seq (Begik et al., 2021) | 22.3 | 24.8 | | | |
| *S. pombe* | PAL-Seq (Subtelny et al., 2014) | 28.2 | 32.8 | $R_s = -0.31$ | | $R_s = -0.15$ |
| Mouse germinal vesicle stage oocytes | PAIso-Seq (Y. Liu, Nie, et al., 2019) | 55.3 | 55.7 | | | |
| *Danio rerio* | Nano3P-Seq (Begik et al., 2021) | 74.5 | 63.6 | | | |
| *Danio rerio* embryo:<br>2 hpf<br>4 hpf<br>6 hpf | PAL-Seq (Subtelny et al., 2014) | 23.1<br>42.2<br>58.1 | 29.4<br>44<br>58.9 | $R_s = -0.13$<br>$R_s = -0.011$<br>$R_s = -0.25$ | | $R_s = 0.77$<br>$R_s = 0.62$<br>$R_s = 0.13$ |
| *X. laevis* embryo:<br>Stage 3–4<br>Stage 9<br>Stage 12–12.5 | PAL-Seq (Subtelny et al., 2014) | 21.7<br>35.4<br>44.5 | 25.4<br>38.1<br>49.2 | | | $R_s = 0.63$<br>$R_s = 0.65$<br>$R_s = 0.13$ |
| *C. elegans* L4 larval stage | mTAIL-Seq (Lima et al., 2017) | 57 | 82 | $R_s = -0.24$ | | $R = -0.32$<br>$F_{op}$ instead of TE |
| | FLAM-Seq (Legnini et al., 2019) | 51 | 59 | $R_p = -0.19$ | | |
| *C. elegans* adult stage | FLAM-Seq (Legnini et al., 2019) | 44.5 | 49.5 | $R_p = -0.21$ | | |

Abbreviations: DRS, Nanopore direct RNA sequencing; $F_{op}$, frequency of optimal codons; $R_p$, Pearson's correlation coefficient; $R_s$, Spearman's correlation coefficient; TE, translational efficiency.
*Note:* In the case of poly(A) tail length, in poly(A)-Seq, FLAM-Seq, and PAIso-Seq, the mean of medians for replicates is given.

poly(A) tails occur in highly expressed genes (Lima et al., 2017; Subtelny et al., 2014), and the length correlates positively with translation efficiency in maturing oocytes and early embryonic development until gastrulation (Eichhorn et al., 2016; Lim et al., 2016; Subtelny et al., 2014; Yang et al., 2020) while in somatic cells only when the poly(A) tail is shorter than 20 nt (Park et al., 2016). In other cases, such a correlation does not exist (Chang et al., 2014; Park et al., 2016; Subtelny et al., 2014), or it is even negative with the frequency of optimal codons in *C. elegans* (Lima et al., 2017; Table 2). Interestingly, a polyadenylation shift occurs during oocyte maturation in *Drosophila* (Lim et al., 2016) and *Xenopus laevis* (Yang et al., 2020) and during embryo development in *Danio rerio* and *X. laevis* (Subtelny et al., 2014). In addition to polyadenylation, 3′ ends of RNA are frequently uridylated with a preference for shorter poly(A) tails or less commonly guanylated with a preference for longer poly(A) tails, which correlates negatively and positively, respectively, with half-life (Chang et al., 2014; Eisen, Eichhorn, Subtelny, Lin, et al., 2020; Lim et al., 2018). Indeed, TAIL-Seq analysis revealed that TUT4 and TUT7 are enzymes that catalyze the uridylation of mRNAs with poly(A) tails <25 nt, which facilitates mRNA decay (Lim et al., 2014). This constitutes an important degradation mechanism of maternal transcripts in growing oocytes (Morgan et al., 2017) and early developmental stages in vertebrates but not invertebrates (Chang et al., 2018). Widespread uridylation was also detected in plants (Zuber et al., 2016). With the mTAIL-Seq approach, the deadenylases poly(A)-specific ribonuclease (PARN) and target of EGR1 (TOE1) do not act on the poly(A) tail of mRNA, contributing to nuclear small-noncoding RNA maturation (Son et al., 2018).

Recently, the novel PAL-Seq version 2 (v2) approach that combines the advantages of PAL-Seq and TAIL-Seq was published (Eisen, Eichhorn, Subtelny, Lin, et al., 2020). This approach utilizes splint ligation like PAL-Seq but is modified to enable the capture of poly(A) tails with terminal U. Poly(A) tails are measured by direct sequencing as in TAIL-Seq, without additional labeling steps, like in the original PAL-Seq approach. Application of this approach, combined with mRNA labeling with 5-ethynyl uridine in the time course, provided insights into poly(A) tail length dynamics in the cytoplasm and revealed their shortening toward a steady-state length over time and a positive correlation between poly(A) tail length and mRNA half-life after 2 h of labeling (pre-steady-state; Eisen, Eichhorn, Subtelny, Lin, et al., 2020), which was previously undetectable when steady-state levels of mRNA were analyzed (Chang et al., 2014; Eisen, Eichhorn, Subtelny, Lin, et al., 2020; Subtelny et al., 2014). PAL-Seq v2 was also used to study the dynamics of miRNA-mediated deadenylation, revealing its effect only on young mRNAs, whereas steady-state poly(A) tail lengths remained unaffected (Eisen, Eichhorn, Subtelny, & Bartel, 2020a). This method is still being improved in slightly differing variants (PAL-Seq v3 and v4; Xiang & Bartel, 2021). It was also combined with translating ribosome affinity purification (TRAP) to create poly(A) tail-length profiling following TRAP (PAL-TRAP) to study intragenic effects of poly(A) tail length on translational efficiency (Xiang & Bartel, 2021). A combination of these protocols was used to elucidate the role of PABPC1 in coupling mRNA poly(A) tail length with translational efficiency in frog oocytes (Xiang & Bartel, 2021).

The publication of these methods broadened our ability to study poly(A) tails. Data that have been generated with these methods changed our view of poly(A) tail length and its relationship to different aspects of RNA metabolism. However, in addition to the aforementioned disadvantages of PCR-biased estimations of tail length, these protocols are technically challenging and not commonly used among RNA-focused labs. The PAL-Seq protocol requires discontinued equipment, and further improvements to this protocol make it similar to TAIL-Seq. From our experience, library preparation for TAIL-Seq is not trivial and requires several steps of optimizations. It is also challenging to obtain high reproducibility between repetitions (Warkocki et al., 2018). Other problems include the requirement for a large amount of starting material and a costly ribodepletion step. Combined, this makes the protocol difficult, time-consuming, and expensive. Thus, in addition to the undeniable importance of these protocols, further developments in the field of poly(A) tail sequencing were essential for further progress in the field.

## 4.2 | Other Illumina-based methods

After these first achievements in genome-wide poly(A) tail length measurements, many other, albeit less common, methods to measure poly(A) tails at genomic scale based on sequencing with the Illumina platform were published. One of them is poly(A)-test RNA sequencing (PAT-Seq; Harrison et al., 2015). Figure 4a shows a general overview of PAT-Seq, which is similar to the previously described TAIL-Seq and PAL-Seq. It utilizes the Klenow fragment of DNA polymerase I to extend the 3′ end of RNA with dNTPs on a template of annealed biotinylated anchored adapter oligonucleotide instead of adapter ligation. It also employs standard Illumina sequencing procedures and directional
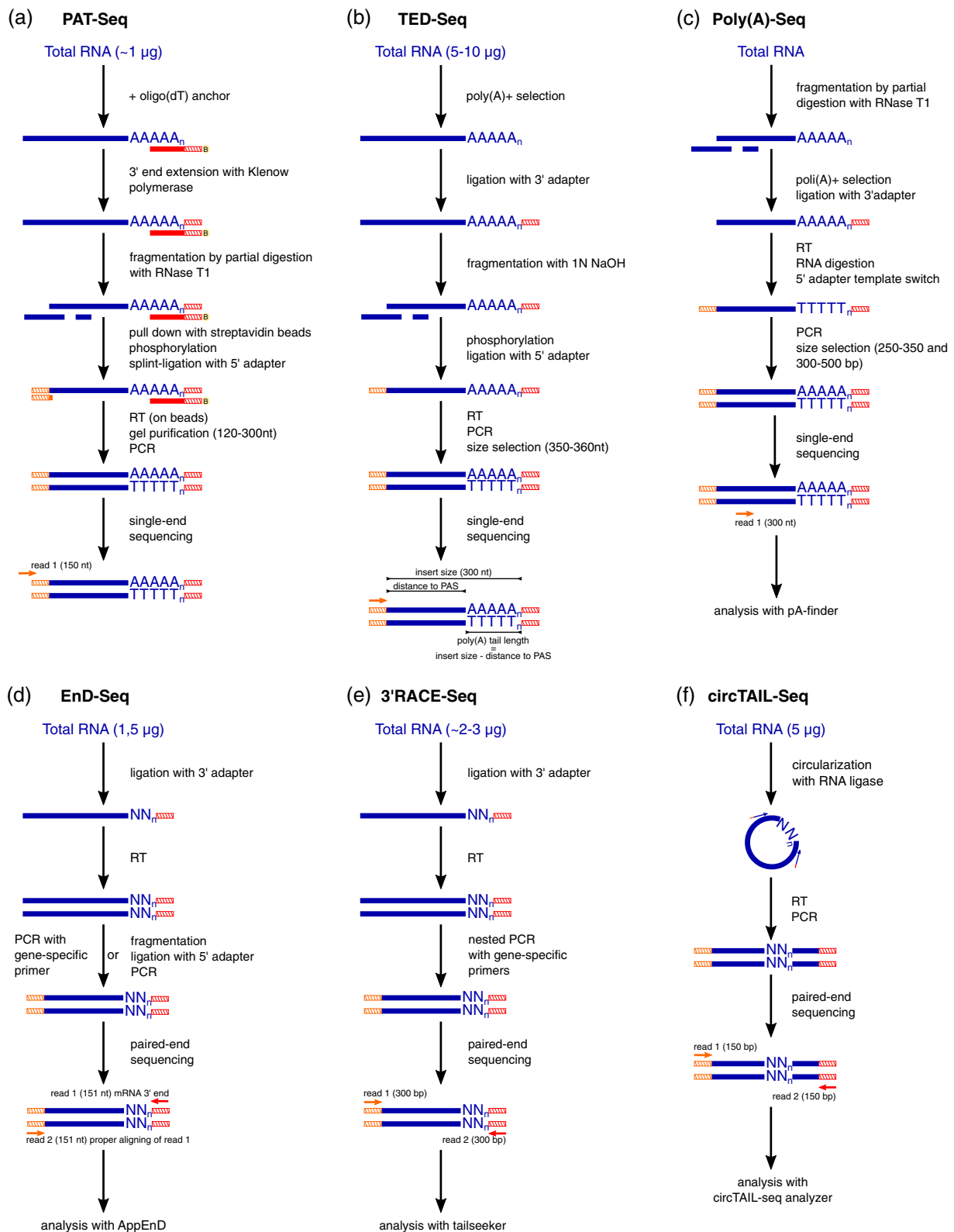
**FIGURE 4** Overview of other Illumina-based approaches to sequence RNA 3′ ends. The red line indicates the oligo(dT) sequence. Red hatched lines indicate the 3′ adapter sequence. Orange hatched lines indicate the 5′ adapter sequence. The "B" in the yellow circle indicates biotin attached to the 3′ adapter.

sequencing from the 5′ end as a solution to a problem of fluorescence signal blurring after a long homopolymer stretch. The poly(A) site is detected as the first non-templated adenosine in a tract that is longer than 3 A (Harrison et al., 2015).

The correlation between poly(A) tail data from PAT-Seq and PAL-Seq (Subtelny et al., 2014) is only 0.3339, with only a 1 nt difference in average poly(A) length. However, employed in this study, the sequencing of only 100 bases prevents the detection of poly(A) tails that are longer than 80 nt (Harrison et al., 2015). Even in a yeast model, where poly(A) tails are rather short (Subtelny et al., 2014), this can lead to the loss of information about the maximal poly(A) tail and restrict its applications. Improvements of this approach by utilizing 150 bp sequencing (Swaminathan et al., 2019) occurred in studies of *C. elegans* (Elewa et al., 2015). This method was also used as standard RNA-Seq for differential expression analysis (Beilharz et al., 2017; Harrison et al., 2015; Johnstone et al., 2018; Tucey et al., 2018), showing that data from PAT-Seq and RNA-Seq correlated well ($r = 0.7860$; Harrison et al., 2015).

Another method for transcriptome-wide poly(A) tail length profiling is tail-end displaced sequencing (TED-Seq; Woo et al., 2018). Similar to TAIL-Seq, poly(A) + RNA is subjected to ligation with a 3′ adapter, fragmentation, and ligation with a 5′ adapter, followed by RT-PCR. The most important step (i.e., size selection by gel excision) must be precisely performed, and libraries are sequenced using the standard Illumina platform (Figure 4b). The 5′ ends are mapped to a reference genome, and 3′ cleavage and polyadenylation sites are determined. The distance between the 5′ end and 3′ polyadenylation site is subtracted from the total size of the fragments, thus giving the poly(A) tail length. Thus, the analysis does not require sophisticated algorithms or modifications of the sequencing run, making it cost-efficient. In this approach, a 100-bp read was used to determine 3′ cleavage and the polyadenylation site, and a 300-bp read was used for poly(A) tail length estimations. To determine absolute poly(A) tail length, in vitro synthesized spike-ins of known poly(A) tail length were used for normalization (Woo et al., 2018). The major limitation of this protocol, however, is the size selection step, which must be done precisely and can directly affect the results. Although the resolution was determined to be ~10 bp (Woo et al., 2018), it may vary between laboratories. Because the recovery rate can differ between samples (Woo et al., 2018), this method cannot be used for quantitative analysis. Additionally, the poly(A) + selection step using oligo(dT) beads can introduce bias into RNAs with longer poly(A) tails (Park et al., 2016).

Poly(A) tail lengths in HEK293 cells, measured by TED-Seq, modestly correlate with data from PAL-Seq ($R = 0.27$; $R = 0.22$ for correlation between PAL-Seq and TAIL-Seq; Woo et al., 2018). This method was used for the analysis of poly(A) tails in the endoplasmic reticulum stress response, showing that under this specific condition, poly(A) tails of stress-responsive transcripts are lengthened, which positively correlates with the stability and translational de-repression of these mRNAs (Woo et al., 2018).

The most recent method of poly(A) tail sequencing with Illumina is poly(A)-Seq (Yu et al., 2020). This protocol starts with RNA fragmentation by partial digestion with RNAse T1. RNA fragments are subjected to the oligo(dT) selection of polyadenylated fragments, which are subsequently ligated to a single-stranded 3′ adapter. It is unlike TAIL-Seq and PAL-Seq protocols that start with the adapter ligation step. Authors that used poly(A)-Seq claim that ligation after fragmentation is more efficient. RNA is then reverse-transcribed, followed by second-strand synthesis based on template-switching for 5′ adapter ligation. cDNA is amplified and sequenced (Figure 4c) with a NextSeq 500 to obtain 300-nt-long single reads. Authors also developed pA-finder, a pipeline for poly(A) tail length analysis. It starts by searching for stretches of adenosines (9As and 6As closest to the 5′ and 3′ ends, respectively), and the region between them is considered to be a poly(A) tail. Only reads that contain such a fragment are subjected to further analysis. The sequence next to the 5′ end is mapped to the genome, and the poly(A) end position is determined by quality assessment.

Poly(A) tails that are measured using this method are longer than tails that are measured with TAIL-Seq and PAL-Seq (median poly[A] tail length for HeLa cells: 93–98 nt; median poly[A] tail length for HCT116 cells: 101–102 nt; median poly[A] tail length for HEK293 cells: 100–103 nt) (Yu et al., 2020). Nonetheless, this method, similar to all methods that utilize short reads) depletes the population of very long poly(A) tails (>150 nt). Correlations between library size and median poly(A) tail length were analyzed. Libraries of large fragments (350–450 nt) and small fragments (250–350 nt) yielded a significant difference in poly(A) tail lengths (~30–120 nt for large fragments vs. ~30–80 nt for small fragments), indicating that the size selection step has a profound impact on the outcome and can explain different lengths that are obtained with different methods. However, because poly(A)-Seq relies on oligo(dT) selection, it is biased toward longer poly(A) tails. Furthermore, in the pA-finder pipeline, preliminary poly(A) tails that are shorter than 10 nt are discarded, thus losing information about very short poly(A) tails (Yu et al., 2020). Poly(A)-Seq enables the detection of other nucleotides in poly(A) tails and has been used to detect guanosines in poly(A) tails of *Arabidopsis thaliana* samples (T. Zhao, Huan, et al., 2019).

All of the above described methods focus on the measurement of long poly(A) tails, and they either do not capture (Lim et al., 2016) or exclude from the analysis (Chang et al., 2014) poly(A) tails that are very short and non-poly (A) tails. Exonuclease degradation sequencing (EnD-Seq; Welch et al., 2015) is a sequencing protocol that detects 3′-end modifications of non-polyadenylated RNAs or decay intermediates, even 1 nt long. To preserve 3′ ends, total RNA is

ligated with the 3′ adapter, followed by cDNA synthesis from the primer antisense to the adapter. To increase sensitivity, this adapter can have a 3′ end that is designed to detect certain 3′ ends of interest. After fragmentation and 5′ adapter ligation, the library is amplified and sequenced (Figure 4d). Because of the utilization of Illumina paired-end sequencing, read 1 contains the adapter sequence and RNA 3′ while read 2 helps with proper alignment. To analyze the data, the application for mapping EnD-Seq data (AppEnD; Welch et al., 2015) computational method was developed. Unlike most approaches, AppEnD does not utilize the prealignment stripping of homopolymers from raw reads, identifying nontemplated 3′ end modifications as parts of reads that do not match the reference genome or the linker during alignment (i.e., soft-clipped reads).

Despite its application for short poly(A) tail detection, this method cannot be used for long poly(A) tail analysis. Importantly, improvements in sensitivity with specifically designed primers affect the results, in which the utilization of primers that end with 3 A cannot detect U-tails that are shorter than 3 nt (Welch et al., 2015).

EnD-Seq combined with the AppEnD approach was used mainly to study the role of uridylation in non-polyadenylated histone mRNA metabolism (Holmquist & Marzluff, 2019; Lackey et al., 2016; Meaux et al., 2018; Slevin et al., 2014; Welch et al., 2015). The AppEnD algorithm that was independently used in the EnD-Seq protocol was used to precisely measure poly(A) tail lengths of tails up to 15 nt based on data on *Drosophila* short capped RNAs (Henriques et al., 2013) and up to 30 nt from TAIL-Seq (Chang et al., 2014) raw data (Welch et al., 2015). Alternatively, EnD-Seq can be used for the deep analysis of specific transcripts (or groups of transcripts) if PCR with a gene-specific primer that contains a 5′ adapter sequence replaces standard adapter ligation (Welch et al., 2015).

Similar to EnD-Seq, 3′RACE-Seq (Goldfarb & Cech, 2013; Warkocki et al., 2018; Wilusz et al., 2011) allows the analysis of short 3′ termini of ncRNAs but only in a transcript-specific manner. This method utilizes a modified 3′RACE protocol to perform the deep sequencing of RNA 3′ ends. RNA is ligated to a 3′ adapter. After reverse transcription from an adapter-specific primer, libraries for Illumina sequencing are generated by nested PCR (Figure 4e). The detection of 3′-end modifications is based on filtering the reads by aligning both to the 3′ region of a gene and adapter sequence (Goldfarb & Cech, 2013). The 3′RACE-Seq approach was used in studies of RNase MRP 3′ (Goldfarb & Cech, 2013), tRNA and tRNA-like small RNAs (Wilusz et al., 2011), and human telomerase RNA (L. Chen et al., 2020; Gable et al., 2019) 3′ termini or combined with a bioinformatical pipeline based on TAIL-Seq (Chang et al., 2014) to study the role of uridylation by TUT4/TUT7 on LINE-1 retrotransposition (Warkocki et al., 2018).

A variant of 3′RACE-Seq was recently published: Nascent RNAend-Seq (Roake et al., 2019). This method combines nascent RNA-Seq (Rädle et al., 2013) and 3′ RACE-Seq (Goldfarb & Cech, 2013) to study changes in ncRNA 3′ tails during RNA maturation. Cells are treated with the nucleotide analog 4-thiouridine (4SU), which labels newly transcribed RNA. After thiol-specific biotinylation, nascent RNA is purified on streptavidin beads. RNA is then subjected to 3′ adapter ligation, followed by reverse transcription from a primer that contains an adapter sequence and PCR with a gene-specific primer and universal primer that contains the adapter sequence. Libraries are sequenced using the standard Illumina platform. This method has a strikingly different application from the ones described above, but it was successfully used to define the role of oligoadenylation by PAP-associated domain-containing protein 5 (PAPD5) and deadenylation by poly(A)-specific ribonuclease (PARN) in human telomerase RNA maturation (Roake et al., 2019).

Another approach to measure poly(A) tail length by targeted sequencing of the only transcript of interest was described as circTAIL-Seq (Gazestani et al., 2016). This method combines low-throughput circRT-PCR (Couttet et al., 1997) with NGS. Total RNA is circularized by RNA ligase, followed by RT-PCR with primers that are designed to obtain a 3′–5′ junction of a transcript of interest (Figure 4f). Because primers contain an adapter sequence, after these steps, a template for sequencing is generated. This method utilizes standard Illumina paired-end sequencing, and data are analyzed using an informatics workflow to allow estimations of poly(A) tail length and composition (Gazestani et al., 2016). circTAIL-Seq is used for single-transcript analysis and low-abundance mRNAs. Importantly, circTAIL-Seq was employed for mitochondrial transcript analyses from *Trypanosoma brucei*, where short (<30 nt) and heterogeneous poly(A) tails are typical (Gazestani et al., 2016; Gazestani et al., 2018). Therefore, circTAIL-Seq applications for longer poly(A) tails that consist of only adenosines is currently untested, especially because no poly(A) tail length spike-ins were used in the original study (Gazestani et al., 2016).

In 2018, a completely new approach to generate mRNA 3′ ends for sequencing that is based on tagmentation by Tn5 transposase was published (Hennig et al., 2018). This protocol utilizes a previously published Tn5 tagmentation approach (Adey et al., 2010; Picelli, Björklund, et al., 2014) that is used in the Nextera XT library preparation kit from Illumina. A hyperactive mutated version of Tn5 that is loaded with two different linker sequences cuts double-stranded DNA, thus ligating linker sequences to both ends of the cut. Linkers are composed of a mosaic end (Zhou et al., 1998) that is recognized by Tn5 and overhangs that are templates for index adapter primers for PCR. Tagmentation generates

short DNA fragments that are surrounded by known sequences that are used as templates for primers during amplification (Adey et al., 2010). cDNA is synthesized with an oligo(dT) primer that contains a 3′ adapter sequence, and the second strand is generated by a previously published template-switching protocol (Picelli, Faridani, et al., 2014). Tn5 that is loaded with only one linker sequence performs double-stranded cDNA tagmentation, leading to (*i*) mRNA 3′ end fragments that are surrounded by a 5′ adapter in the Tn5 linker and 3′ adapter in the oligo(dT) primer and (*ii*) fragments from the mRNA body that are surrounded only by linkers. Thus, PCR with primers for 5′ and 3′ adapters that omits the gap-filling step allows the specific amplification of only 3′-end sequences (Hennig et al., 2018).

This protocol is fast, robust, and inexpensive and requires only a low amount of starting material. Its main disadvantage is a lack of strand specificity. To date, this method has been used only for poly(A) site mapping. Nevertheless, this mRNA 3′ end library preparation protocol and its derivatives (e.g., transposase-mediated 3′-RNA-Seq [TM3′-Seq]; Pallares et al., 2020), combined with a good approach for sequencing and data analysis, can be useful for polyadenylation studies. However, no such protocol has yet been published. Thus, this approach requires a significant optimization effort. We tried performing 3′ end and standard RNA-Seq with Tn5. We found that although standard RNA-Seq works well, effective tagmentation and amplification of the mRNA 3′ end is challenging.

In summary, over the last 8 years, much progress has been made in the area of poly(A) tail length measurement techniques by designing protocols that are based on NGS. Various protocols have been created, most notably TAIL-Seq and PAL-Seq. Each of these techniques has its own advantages and disadvantages, with the common drawback of poly(A) tail estimation that is biased by the PCR amplification of long homopolymer stretches. However, these methods were designed based on the needs of specific laboratories, and some have never been used beyond the laboratory that developed them. Moreover, some of the methods were designed to analyze single transcripts or short poly(A) tails, thus limiting their application. Furthermore, the data that were generated with these protocols for sequencing, including TAIL-Seq and PAL-Seq, do not correlate well, raising the important problem of a lack of reproducibility between laboratories and questions about the reliability of estimations of these methods. When considering the technical challenges, cost, and time needed for library preparation, only a few laboratories have successfully performed poly(A) tail sequencing. Thus, there is a need for an easy-to-implement protocol that generates reproducible results.

## 4.3 | Poly(A) tail length measurement with the PacBio platform

Although Illumina-based approaches dominate in transcriptomic studies, two research groups published the first successful poly(A) tail profiling protocols using the PacBio third-generation sequencing platform: full-length mRNA sequencing (FLAM-Seq; Legnini et al., 2019) and poly(A) inclusive RNA isoform sequencing (PAIso-Seq; Y. Liu, Nie, et al., 2019). Unlike the Illumina sequencing-by-synthesis of clustered short cDNA fragments approach, PacBio features single-molecule, real-time (SMRT) sequencing (Eid et al., 2009) to generate full-length cDNA sequences with Iso-Seq technology (Gonzalez-Garay, 2016). The utilization of circular consensus sequencing by the multiple passing of a single template molecule generates long high-fidelity reads with high accuracy (even 99.8%; Wenger et al., 2019).

The FLAM-Seq (Legnini et al., 2019) and PAIso-Seq (Y. Liu, Nie, et al., 2019) protocols present very similar approaches. The PacBio platform, which can handle homopolymers better than Illumina, enables sequencing of the whole molecule of polyadenylated RNA, providing isoform-specific information about poly(A) tails. The main steps of both protocols are 3′-end extension, RT with template switching, cDNA amplification, circular adapter ligation, and PacBio sequencing (Figure 5). FLAM-Seq starts with the selection of a poly(A) + fraction of RNA, which is subsequently subjected to G/I tailing. RNA is reverse transcribed with a primer that consists of a PCR handle on its 5′ end, followed by an N stretch and three T pairing to the added G/I and poly(A) tail, respectively, in the presence of template-switching oligonucleotide (TSO) for second-strand synthesis. The created double-stranded cDNA is then amplified and subjected to PacBio sequencing (Legnini et al., 2019). PAIso-Seq does not require the selection of polyadenylated RNAs because the 3′ end is extended on a template of oligonucleotides that consist of a TSO sequence minus a triple G on the 5′ end and stretch of T on the 3′ end by Klenow polymerase. Inside the T stretch, two dU nucleotides are present, which enables oligonucleotide degradation with the USER enzyme before RT. Double-stranded cDNA is synthesized in the presence of primers that correspond to TSO minus triple G and with TSO with triple G, amplified, and subjected to PacBio sequencing (Y. Liu, Nie, et al., 2019). However, the PAIso-Seq protocol may lead to Klenow-dependent artifacts because the fill-in reaction may actually extend the measured poly(A) tails.

The major advantage of all PacBio-based approaches is the generation of long reads with high accuracy (error rate <0.8%; Legnini et al., 2019), thus enabling isoform-specific analysis. Moreover, because of PacBio's ability of
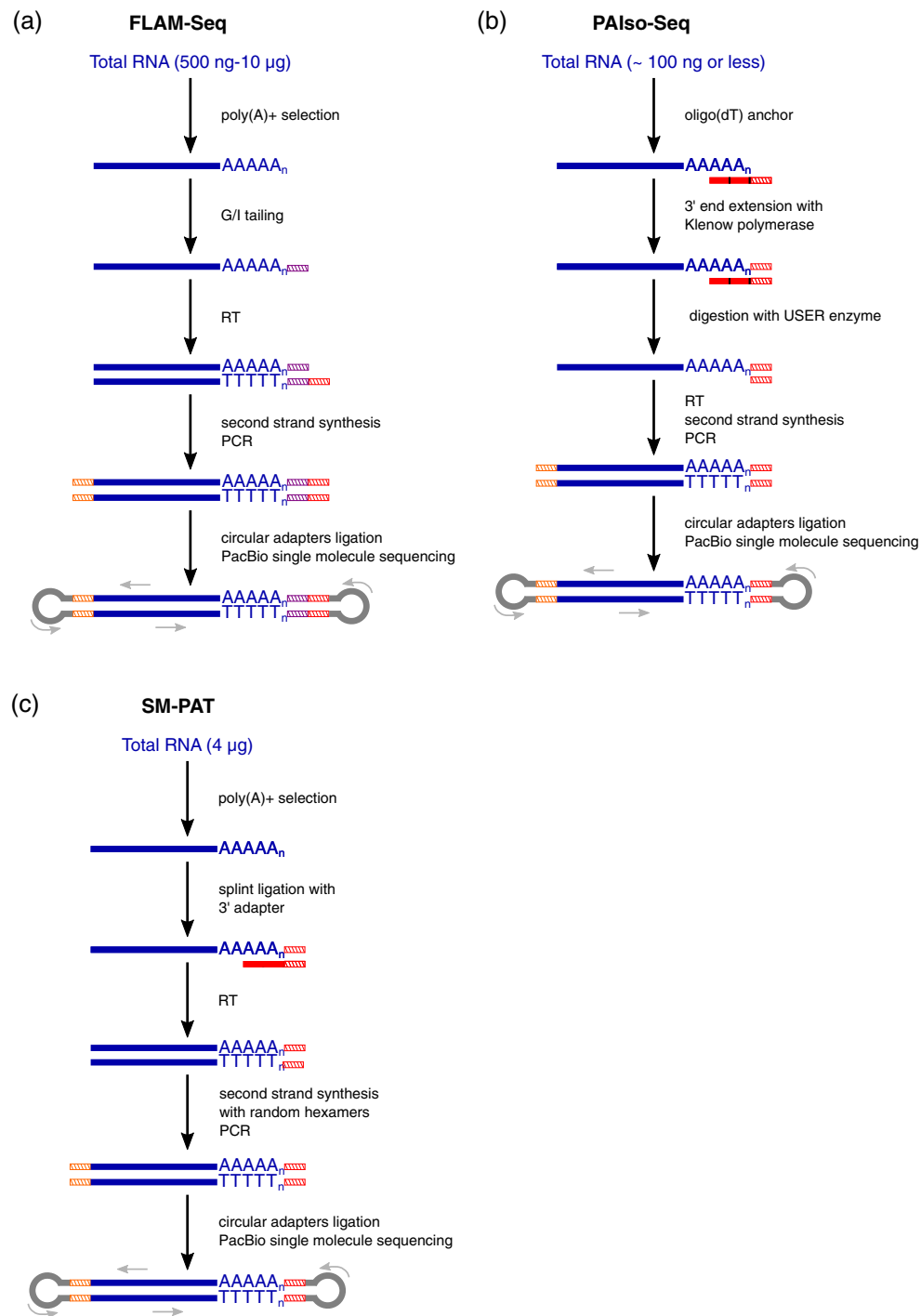
**FIGURE 5** Comparison of FLAM-Seq, PAIso-Seq, and SM-PAT protocols for mRNA poly(A) tail sequencing with the PacBio platform. Violet hatched lines indicate the G/I tag sequence. Red hatched lines indicate the 3′ adapter sequence. Orange hatched lines indicate the 5′ adapter sequence. The red line indicates the oligo(dT) sequence. dU nucleotides are marked in black.

homopolymer sequencing, it can be used to study the internal composition of poly(A) tails, which was impossible with previous approaches (e.g., TAIL-Seq allows the detection of only 3′-end modifications; Legnini et al., 2019; Y. Liu, Nie, et al., 2019). Non-A nucleotides were also shown to occur within poly(A) tails. However, based on the utilization of T-A pairing during RT (FLAM-Seq) or end extension (PAIso-Seq) to analyze only poly(A) + RNAs, neither method provides data about 3′-end modifications as TAIL-Seq does. FLAM-Seq involves a poly(A) + selection step with oligo(dT) beads that may introduce bias toward RNAs with longer poly(A) tails. This is avoided in PAIso-Seq, in which polyadenylated RNAs are selected by an end extension template oligonucleotide that contains a T stretch on the 3′ end. Moreover, this

improvement increases sensitivity, allowing less starting material with PAIso-Seq or even performing this protocol with material from a single cell (Y. Liu, Nie, et al., 2019). These features of these two methods are summarized in Table 1.

FLAM-Seq was used to analyze poly(A) tails in HeLa S3 cells, human induced pluripotent stem cells (iPSCs), iPSC-derived cerebral organoids, and *C. elegans* (L4 larval stage and adult). The obtained poly(A) tails are typically short and correlate well with mTAIL-Seq data from L4 larval stage *C. elegans* ($R = 0.63$; Lima et al., 2017), whereas the correlation of data from HeLa cells that were obtained with TAIL-Seq (Chang et al., 2014) and PAL-Seq (Subtelny et al., 2014) is rather low ($R = 0.03$ and $R = 0.24$, respectively; the correlation between TAIL-Seq and PAL-Seq: $R = 0.2$; see Table 2). Likewise and as postulated earlier (Lima et al., 2017), in highly expressed genes, poly(A) tails are shorter, whereas median poly(A) tail length correlates slightly negatively with mRNA half-life and translational efficiency (Legnini et al., 2019). With PAIso-Seq, mouse germinal vesicle-stage oocyte poly(A) tails were analyzed, showing a good correlation with TAIL-Seq data ($Rs = 0.65$; Morgan et al., 2017) and a positive correlation between poly(A) tails and protein abundance in oocytes, which was postulated previously (Lim et al., 2016; Subtelny et al., 2014; Yang et al., 2020; see Table 2).

Another single-molecule SM-PAT protocol for poly(A) sequencing with the PacBio platform was recently published (Mattijssen et al., 2020). Total RNA is subjected to the oligo(dT) selection of polyadenylated mRNAs, followed by splint ligation with a 3′ adapter. After RT, the second strand of cDNA is created with random hexamers that are attached to the adapter sequence. After PCR, adapters for SMRT sequencing are added, and the library is loaded onto SMRT Cell V2 and subjected to PacBio sequencing (Figure 5c). This protocol is very similar to FLAM-Seq and PAIso-Seq, differing in the 3′-adapter ligation and cDNA second-strand synthesis approaches. It was used to analyze the effect of La-related protein 4 (LARP4) on poly(A) tail length, revealing that this protein promotes the lengthening of poly(A) tails, protects them from deadenylation, and stabilizes mRNAs. Comparisons of the quantification of RNA abundance with this and other methods produced $R^2 = 0.58$ in the case of RNA-Seq and a very high $R^2 = 0.96$ in the case of Northern blot (Mattijssen et al., 2020).

In summary, with the development of third-generation sequencing technology, the next step in poly(A) tail length studies was made. Long-read sequencing removes the upper limits of poly(A) tail length detection and gives a more complete view of the mRNA molecule from its 5′ end to 3′ end. However, these methods are still biased by the PCR amplification of poly(A) tails during library preparation. Importantly, PacBio sequencing is quite expensive, and sequencers are not common among laboratories, thus decreasing its versatility and availability.

## 4.4 | Nanopore direct RNA sequencing in poly(A) tail length measurement

The latest approach for RNA sequencing and poly(A) tail length measurement is direct RNA sequencing (DRS) by ONT (Garalde et al., 2018). This approach is free from cDNA synthesis and PCR amplification, thus enabling the direct, real-time sequencing of a single RNA molecule and generating full-length, strand-specific reads. In addition to standard differential expression analysis, it allows studies of transcript isoforms and splicing (Depledge et al., 2019; Drexler et al., 2020; Roach et al., 2020; Workman et al., 2019; L. Zhao, Zhang, et al., 2019), RNA modifications (H. Liu, Begik, et al., 2019; Lorenz et al., 2020; Parker et al., 2020; Workman et al., 2019), and poly(A) tail length (D. Kim et al., 2020; Krause et al., 2019; Roach et al., 2020; Workman et al., 2019) at the single-molecule level. The sequencing library workflow (Figure 6) consists of splint ligation of the 3′ adapter to a polyadenylated RNA molecule, followed by the optional RT and ligation of a sequencing adapter that is attached to a motor protein. The library is loaded on a flow cell that is created by protein pores (nanopores) that are embedded in an array of synthetic membranes. After applying the voltage, RNA passes through the nanopore at a constant speed, driven by motor protein, in a 3′–5′ direction, creating a specific perturbation of current that is converted into a base sequence using a trained neural network algorithm (Garalde et al., 2018; Stark et al., 2019). Importantly, even if the cDNA strand is synthesized, only RNA is sequenced because of the motor protein that is attached to the sequencing adapter.

Compared with the previously described methods, the determination of poly(A) tail lengths is more challenging because standard base-calling tools are unable to call long homopolymers correctly. Thus, poly(A) tail lengths cannot be determined directly from sequence files. Therefore, a specific data analysis pipeline based on the raw current intensity signal is required using Nanopolish-polyA (https://github.com/jts/nanopolish; Workman et al., 2019) or TailfindR (https://github.com/adnaniazi/tailfindr; Krause et al., 2019). Both rely on proper segmentation of the raw current signal into a sequencing adapter, splint adapter, poly(A) tail, and eventually transcript body (Figure 6) but using different approaches. Nanopolish-polyA uses the hidden Markov model to find segment boundaries, whereas TailfindR looks for
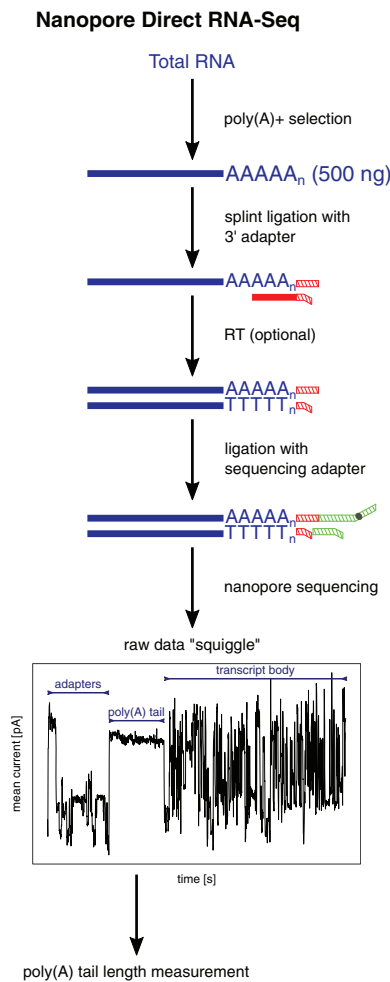
**Nanopore Direct RNA-Seq**



**FIGURE 6** Library preparation workflow for poly(A) tail length measurement with nanopore sequencing and exemplary raw data (squiggle) segmented into adapters, poly(A) tail, and transcript body sequence. The red line indicates the oligo(dT) sequence. Red hatched lines indicate the 3′ adapter sequence (splint adapter). Green hatched lines indicate the sequencing adapter with a motor protein attached (gray circle).

regions in the normalized current intensity signals with defined values. Having boundaries of poly(A) in the raw signal, the time it takes for the poly(A) tail to pass through a pore can be calculated. Additionally, based purely on the raw signal (for TailfindR) or alignment of the raw signal to the base-called and reference-mapped sequence (for Nanopolish-polyA), the speed at which the RNA passes through the pore can be calculated (deviations from the expected ∼70 nt/s are observed). Finally, by determining the dwell time of the poly(A) region in the pore and the average translocation speed, the length of the poly(A) tail can be calculated using the following formula: poly(A) length = poly(A) dwell time × average translocation speed.

The algorithm that is implemented in Nanopolish-polyA was tested on sequencing data from a human B cell lymphoma cell line, obtaining an average poly(A) size of 50 nt, which has been reported previously (Chang et al., 2014; Subtelny et al., 2014), and mean poly(A) tail length of 112 nt for nuclear transcripts. The poly(A) tail length can also be transcript isoform-specific (Workman et al., 2019) and its effect on RNA stability can be assessed for each isoform (Maier et al., 2020). This approach was also used in studies of *C. elegans* (Roach et al., 2020), showing similar poly(A) tail length distributions as previously published data from mTAIL-Seq (Lima et al., 2017) in the L4 stage. Moreover, in the larval stage, poly(A) tails are shorter than in adults and correlate negatively with gene expression, which is lost during development (Roach et al., 2020). Analyses of direct RNA-Seq data with Nanopolish-polyA were also performed to estimate poly(A) tail lengths in studies of *A. thaliana* (Parker et al., 2020), splicing in human K562 cell RNA with or without in vitro polyadenylation (Drexler et al., 2020), and cytoplasmic polyadenylation by TENT5C in mouse B cells (Bilska et al., 2020) and TENT5A in osteoblasts (Gewartowska

et al., 2021) and to show poly(A) tail shortening in human adenovirus type 5 with the progression of infection (Donovan-Banfield et al., 2020). It was also used to study the metabolism of poly(A) tails in the yeast *Saccharomyces cerevisiae* (Tudek et al., 2021; Turtola et al., 2021).

TailfindR is less widely used than Nanopolish-polyA, despite the fact that the correlation of results that are computed with these two methods is good ($R = 0.98$ for poly[A] tail lengths) (Krause et al., 2019). The discrepancies are probably caused by different approaches to calculating poly(A) boundaries. Both tools were also compared in poly(A) tail analyses of human HAP1 and HEK293 cell lines, giving comparable results with a modal poly(A) tail length that was similar to the one that was measured with TAIL-Seq (Chang et al., 2014; Soneson et al., 2019). TailfindR also has improvements in data analysis based on ONT cDNA sequencing, measuring both poly(A) and poly(T), which reduces the amount of input material and increases the number of reads. Moreover, the cDNA approach generated tail lengths even more precisely than RNA (Krause et al., 2019). Recently, Nanopore 3 Prime end-capture sequencing (Nano3P-Seq), another approach for poly(A) tail length profiling, was developed, relying on direct cDNA nanopore sequencing (Begik et al., 2021). In this protocol, the template-switching reverse transcriptase TGIRT is used to capture all, even non-polyadenylated, RNA species, and TailfindR is used to calculate poly(A) tail lengths. This method was used to study polyadenylation dynamics during the maternal-to-zygotic transition at the isoform-specific level (Begik et al., 2021). Interestingly, the authors of Nano3P-Seq noticed a consistent bias of 15 nt in TailfindR-calculated poly(A) lengths compared with poly(A) standards, thus obtaining results that required postprocessing.

Any long homopolymer at the 3′ end of RNA can be detected with DRS. This aspect was used by the authors of Nanopolish-detect-polyI and employed in the nano-COP protocol (Drexler et al., 2021). Nano-COP aims to analyze co-transcriptional processing by Nanopore sequencing. Because nascent RNA is not fully polyadenylated, RNA must be enzymatically tailed to perform sequencing. However, this approach makes it impossible to distinguish poly(A) tails of mature RNA from enzymatically added poly(A) tails. To overcome this issue, RNA can be tailed by poly(I), which allows the preparation of libraries using an appropriately designed adapter. Nanopolish-detect-polyI can then distinguish between poly(I) and poly(A) tails and measure the length of poly(A) tails (Drexler et al., 2021).

Recently, poly(A) tails of SARS-CoV-2 RNAs were measured with a not yet published poreplex tool (https://github.com/hyeshik/poreplex), showing median poly(A) tail lengths of 47 nt and a difference in poly(A) tail length between modified and unmodified RNAs (D. Kim et al., 2020). The authors utilized a similar approach as Nanopolish and TailfindR to measure the dwell time of the poly(A) tail. The dwell time was converted to a length by normalization to 1/30 of the dwell time of the 30-nt poly(A) tail from the ONT calibration strand that was added to the library (D. Kim et al., 2020). Thus, the normalization method is the main difference between these tools.

Direct RNA sequencing is still a novel method, and subsequent data analysis tools are still being developed. Computational pipelines allow the simultaneous analysis of various aspects of RNA biology. One example is the MasterOfPores (Cozzuto et al., 2020) pipeline for direct RNA-Seq data analysis. One of its modules, NanoTail, performs poly(A) tail length estimations with Nanopolish and TailfindR algorithms to generate data separately from both methods as well as computing correlations between the data that are obtained with them (Cozzuto et al., 2020). For the visualization and analysis of Oxford Nanopore direct RNA-Seq-based poly(A) predictions, the R package NanoTail (https://github.com/pbrigf-ibb/nanotail) was created, but it has not yet been published. It allows statistical comparisons of poly(A) lengths between multiple datasets and was used in several published works that employed DRS for poly(A) analysis (Bilska et al., 2020; Gewartowska et al., 2021).

Additionally, information about poly(A) tail length can be obtained from the sequencing of cDNA after amplification as shown in FLEP-Seq (Long et al., 2021). This protocol is used to study nascent transcripts by utilizing adapter ligation to capture both poly- and non-adenylated mRNAs, followed by cDNA synthesis, amplification, and nanopore sequencing. However, it requires dedicated PolyAcaller (Long et al., 2021) software to measure poly(A) tails as Nanopolish can be used only for RNA, and Tailfinder that has been designed for cDNA generates high false-positive rates. The developed protocol can also be used for PacBio sequencing. Poly(A) tails that are measured with both approaches show a good correlation ($R = 0.89$; Long et al., 2021).

Apart from the many advantages of Nanopore sequencing, its major disadvantages are the high error rate (Rang et al., 2018) and the requirement for a relatively large amount of starting material (500 ng of poly[A] + RNA) for direct RNA-Seq, which may be biased toward RNAs with longer poly(A) tails, depending on the poly(A) selection method used (Ibrahim et al., 2021; Tudek et al., 2021). Similar to mTAIL-Seq and PAL-Seq, because of the utilization of splint ligation to the 3′ adapter, it captures only polyadenylated RNAs, and a modified protocol is needed to study RNAs with non-poly(A) 3′ ends. Table 1 shows the major features of this method compared with other previously described methods.

Direct Nanopore sequencing was the first PCR-free method to study poly(A) tail length, representing the latest accomplishment in the development of these methods. The estimation of poly(A) tail length that is not limited by reading length and not biased by PCR amplification seems to be a solution to problems with the previous techniques. Additionally, the protocol for library preparation is short and easy to perform. The required equipment includes a portable and relatively inexpensive sequencer that is connected to a laptop computer, making this protocol easy to implement in most laboratories, resulting in the rising popularity of nanopore sequencing for poly(A) tail length measurements. Importantly, no special protocol for poly(A) tail analyses is needed, meaning that we can perform various analyses in one sequencing run, including differential expression, splicing isoforms, modifications, and co-transcriptional processing. Although there are still limitations of this approach, ONT is constantly improving the sequencing equipment and chemistry. Users are generating new versions of protocols for library preparation and computational analyses, making this technique very promising.

## 5 | CONCLUSION

Tailing of the 3′ end of mRNA is an important post-transcriptional modification that contributes significantly to the regulation of the expression of genetic information. Therefore, it requires an appropriate methodology to study it properly. During the last 50 years, we have witnessed fascinating developments in the methods that are used to study poly(A) tail length, starting with RNaseH- and PCR-based approaches, which opened this field of research. Their significant limitation with regard to throughput was overcome by the development of microarray technology and studying transcripts that are fractionated by their poly(A) tail length. However, none of these methods provided information about exact poly(A) tail lengths or possible modifications. A revolution occurred with the development of NGS. The first adjustments in sequencing protocols were made to enable studies of poly(A) tail length. Publications that were based on TAIL-Seq (Chang et al., 2014) and PAL-Seq (Subtelny et al., 2014) enhanced our knowledge of the 3′ terminome, providing more accurate, high-throughput data on poly(A) tail length and the detection of its modifications. Despite their undeniable impact on the field, these methods are characterized by the unavoidable bias of PCR amplification of long homopolymers, poly(A) tails of which lead to underestimations of poly(A) tail length, and notable technical challenges. Additionally, the measured lengths of poly(A) tails are limited to read length, leading to the loss of information about very long tails. These problems can be overcome by measuring poly(A) tail lengths using single-molecule third-generation sequencing. Protocols that are based on the PacBio platform, such as FLAM-Seq and PAIso-Seq, still introduce the PCR amplification of libraries. Oxford Nanopore Technologies offers a direct RNA or direct cDNA sequencing approach, leading to non-PCR-biased poly(A) tail length estimations. Because it generates long reads compared with short-read NGS, poly(A) tail measurements can be accompanied by analyses of other aspects of RNA biology, such as splicing isoforms and base modifications.

In conclusion, the massive area of 3′ terminome research requires more exploration. Further developments in poly(A) tail sequencing still need to be made. The utilization of Nanopore sequencing is gaining popularity, and we predict further improvements in protocols based on this technology.

**AUTHOR CONTRIBUTIONS**
**Aleksandra Brouze:** Conceptualization (equal); funding acquisition (equal); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Paweł Szczepan Krawczyk:** Writing – original draft (equal). **Andrzej Dziembowski:** Conceptualization (supporting); supervision (supporting); writing – original draft (supporting). **Seweryn Mroczek:** Conceptualization (equal); funding acquisition (equal); supervision (lead); writing – original draft (equal).

## CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

*Aleksandra Brouze* https://orcid.org/0000-0002-0531-8451
*Paweł Szczepan Krawczyk* https://orcid.org/0000-0001-9531-1298
*Andrzej Dziembowski* https://orcid.org/0000-0001-8492-7572
*Seweryn Mroczek* https://orcid.org/0000-0001-5026-1741

## RELATED WIREs ARTICLES

Polyadenylation and beyond: Emerging roles for noncanonical poly(A) polymerases
Functions and mechanisms of RNA tailing by metazoan terminal nucleotidyltransferases
Control of poly(A) tail length

## FURTHER READING

Nicholson, A. L., & Pasquinelli, A. E. (2019). Tales of detailed poly(A) tails. Trends in Cell Biology, 29(3), 191–200. https://doi.org/10.1016/J.TCB.2018.11.002

Passmore, L. A., & Coller, J. (2022). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. Nature Reviews Molecular Cell Biology, 23, 93–106. https://doi.org/10.1038/s41580-021-00417-y

Yu, S., & Kim, V. N. (2020). A tale of non-canonical tails: Gene regulation by post-transcriptional RNA tailing. Nature Reviews Molecular Cell Biology, 21(9), 542–556. https://doi.org/10.1038/s41580-020-0246-8

## REFERENCES

Adey, A., Morrison, H. G., Asan, X., Xun, X., Kitzman, J. O., Turner, E. H., Stackhouse, B., MacKenzie, A., Caruccio, N. C., Zhang, X., & Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology*, *11*(12), R119. https://doi.org/10.1186/gb-2010-11-12-r119

Alt, F. W., Bothwell, A. L. M., Knapp, M., Siden, E., Mather, E., Koshland, M., & Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. *Cell*, *20*(2), 293–301. https://doi.org/10.1016/0092-8674(80)90615-7

Barnard, D. C., Ryan, K., Manley, J. L., & Richter, J. D. (2004). Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell*, *119*(5), 641–651. https://doi.org/10.1016/j.cell.2004.10.029

Bazzini, A. A., Lee, M. T., & Giraldez, A. J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science*, *336*(6078), 233–237. https://doi.org/10.1126/science.1215704

Begik, O., Liu, H., Delgado-tejedor, A., Giraldez, A. J., Beaudoin, J., Mattick, J. S., & Novoa, E. M. (2021). Nano3P-seq: transcriptome-wide analysis of gene expression and tail dynamics using end-capture nanopore sequencing.

Beilharz, T. H., Harrison, P. F., Miles, D. M., See, M. M., Le, U. M. M., Kalanon, M., Curtis, M. J., Hasan, Q., Saksouk, J., Margaritis, T., Holstege, F., Geli, V., & Dichtl, B. (2017). Coordination of cell cycle progression and mitotic spindle assembly involves histone H3 lysine 4 methylation by set1/COMPASS. *Genetics*, *205*(1), 185–199. https://doi.org/10.1534/genetics.116.194852

Beilharz, T. H., & Preiss, T. (2007). Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA*, *13*(7), 982–997. https://doi.org/10.1261/rna.569407

Bilska, A., Kusio-Kobiałka, M., Krawczyk, P. S., Gewartowska, O., Tarkowski, B., Kobyłecki, K., Nowis, D., Golab, J., Gruchota, J., Borsuk, E., Dziembowski, A., & Mroczek, S. (2020). Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(a) polymerase TENT5C. *Nature Communications*, *11*(2031), 1–17. https://doi.org/10.1038/s41467-020-15835-3

Chang, H., Lim, J., Ha, M., & Kim, V. N. (2014). TAIL-seq: Genome-wide determination of poly(a) tail length and 3′ end modifications. *Molecular Cell*, *53*(6), 1044–1052. https://doi.org/10.1016/j.molcel.2014.02.007

Chang, H., Yeo, J., Kim, J., Kim, H., Lim, J., Lee, M., Kim, H. H., Ohk, J., Jeon, H. Y., Lee, H., Jung, H., Kim, K. W., & Kim, V. N. (2018). Terminal Uridylyltransferases execute programmed clearance of maternal transcriptome in vertebrate embryos. *Molecular Cell*, *70*(1), 72–82.e7. https://doi.org/10.1016/j.molcel.2018.03.004

Chen, C. Y., Gherzi, R., Ong, S. E., Chan, E. L., Raijmakers, R., Pruijn, G. J. M., Stoecklin, G., Moroni, C., Mann, M., & Karin, M. (2001). AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell*, *107*(4), 451–464. https://doi.org/10.1016/S0092-8674(01)00578-5

Chen, L., Roake, C. M., Galati, A., Bavasso, F., Micheli, E., Saggio, I., Schoeftner, S., Cacchione, S., Gatti, M., Artandi, S. E., & Raffa, G. D. (2020). Loss of human TGS1 Hypermethylase promotes increased telomerase RNA and telomere elongation. *Cell Reports*, *30*(5), 1358–1372.e5. https://doi.org/10.1016/j.celrep.2020.01.004

Choi, Y. H., & Hagedorn, C. H. (2003). Purifying mRNAs with a high-affinity elF4E mutant identifies the short 3′ poly(a) end phenotype. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(12), 7033–7038. https://doi.org/10.1073/pnas.1232347100

Ciafrè, S. A., & Galardi, S. (2013). microRNAs and RNA-binding proteins. *RNA Biology*, *10*(6), 934–942. https://doi.org/10.4161/RNA.24641

Couttet, P., Fromont-Racine, M., Steel, D., Pictet, R., & Grange, T. (1997). Messenger RNA deadenylylation precedes decapping in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(11), 5628–5633. https://doi.org/10.1073/pnas.94.11.5628

Cozzuto, L., Liu, H., Pryszcz, L. P., Pulido, T. H., Delgado-Tejedor, A., Ponomarenko, J., & Novoa, E. M. (2020). MasterOfPores: A workflow for the analysis of Oxford Nanopore direct RNA sequencing datasets. *Frontiers in Genetics*, *11*, 1–11. https://doi.org/10.3389/fgene.2020.00211

Depledge, D. P., Srinivas, K. P., Sadaoka, T., Bready, D., Mori, Y., Placantonakis, D. G., Mohr, I., & Wilson, A. C. (2019). Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nature Communications*, *10*(1), 754. https://doi.org/10.1038/s41467-019-08734-9

Di Giammartino, D. C., Nishida, K., & Manley, J. L. (2011). Mechanisms and consequences of alternative polyadenylation. *Molecular Cell*, *43*(6), 853–866. https://doi.org/10.1016/j.molcel.2011.08.017

Donovan-Banfield, I., Turnell, A. S., Hiscox, J. A., Leppard, K. N., & Matthews, D. A. (2020). Deep splicing plasticity of the human adenovirus type 5 transcriptome drives virus evolution. *Communications Biology*, *3*(1), 1–14. https://doi.org/10.1038/s42003-020-0849-9

Dower, K., Kuperwasser, N., Merrikh, H., & Rosbash, M. (2004). A synthetic a tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. *RNA*, *10*(12), 1888–1899. https://doi.org/10.1261/RNA.7166704

Drexler, H. L., Choquet, K., & Churchman, L. S. (2020). Splicing kinetics and coordination revealed by direct nascent RNA sequencing through Nanopores. *Molecular Cell*, *77*(5), 985–998.e8. https://doi.org/10.1016/j.molcel.2019.11.017

Drexler, H. L., Choquet, K., Merens, H. E., Tang, P. S., Simpson, J. T., & Churchman, L. S. (2021). Revealing nascent RNA processing dynamics with nano-COP. *Nature Protocols*, *16*(3), 1343–1375. https://doi.org/10.1038/s41596-020-00469-y

Du, L., & Richter, J. D. (2005). Activity-dependent polyadenylation in neurons. *RNA*, *11*(9), 1340–1347. https://doi.org/10.1261/rna.2870505

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., & Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell*, *20*(2), 313–319. https://doi.org/10.1016/0092-8674(80)90617-0

Eichhorn, S. W., Subtelny, A. O., Kronja, I., Kwasnieski, J. C., Orr-Weaver, T. L., & Bartel, D. P. (2016). mRNA poly(a)-tail changes specified by deadenylation broadly reshape translation in drosophila oocytes and early embryos. *eLife*, *5*, 1–24. https://doi.org/10.7554/eLife.16955

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., … Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*(5910), 133–138. https://doi.org/10.1126/science.1162986

Eisen, T. J., Eichhorn, S. W., Subtelny, A. O., & Bartel, D. P. (2020). MicroRNAs cause accelerated decay of short-tailed target mRNAs. *Molecular Cell*, *77*(4), 775–785.e8. https://doi.org/10.1016/j.molcel.2019.12.004

Eisen, T. J., Eichhorn, S. W., Subtelny, A. O., Lin, K. S., McGeary, S. E., Gupta, S., & Bartel, D. P. (2020). The dynamics of cytoplasmic mRNA metabolism. *Molecular Cell*, *77*(4), 786–799.e10. https://doi.org/10.1016/j.molcel.2019.12.005

Elewa, A., Shirayama, M., Kaymak, E., Harrison, P. F., Powell, D. R., du, Z., Chute, C. D., Woolf, H., Yi, D., Ishidate, T., Srinivasan, J., Bao, Z., Beilharz, T. H., Ryder, S. P., & Mello, C. C. (2015). POS-1 promotes Endo-mesoderm development by inhibiting the cytoplasmic polyadenylation of neg-1 mRNA. *Developmental Cell*, *34*(1), 108–118. https://doi.org/10.1016/j.devcel.2015.05.024

Fuke, H., & Ohno, M. (2008). Role of poly (a) tail as an identity element for mRNA nuclear export. *Nucleic Acids Research*, *36*(3), 1037–1049. https://doi.org/10.1093/nar/gkm1120

Gable, D. L., Gaysinskaya, V., Atik, C. C., Talbot, C. C., Jr., Kang, B., Stanley, S. E., Pugh, E. W., Amat-Codina, N., Schenk, K. M., Arcasoy, M. O., Brayton, C., Florea, L., & Armanios, M. (2019). ZCCHC8, the nuclear exosome targeting component, is mutated in familial pulmonary fibrosis and is required for telomerase RNA maturation. *Genes and Development*, *33*(19–20), 1381–1396. https://doi.org/10.1101/gad.326785.119

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., Jordan, M., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, E. J., Jayasinghe, L., Wright, C., … Turner, D. J. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, *15*(3), 201–206. https://doi.org/10.1038/nmeth.4577

Gazestani, V. H., Hampton, M., Abrahante, J. E., Salavati, R., & Zimmer, S. L. (2016). CircTAIL-seq, a targeted method for deep analysis of RNA 3′ tails, reveals transcript-specific differences by multiple metrics. *RNA*, *22*(3), 477–486. https://doi.org/10.1261/rna.054494.115

Gazestani, V. H., Hampton, M., Shaw, A. K., Salavati, R., & Zimmer, S. L. (2018). Tail characteristics of Trypanosoma brucei mitochondrial transcripts are developmentally altered in a transcript-specific manner. *International Journal for Parasitology*, *48*(2), 179–189. https://doi.org/10.1016/j.ijpara.2017.08.012

Gewartowska, O., Aranaz-Novaliches, G., Krawczyk, P. S., Mroczek, S., Kusio-Kobiałka, M., Tarkowski, B., Spoutil, F., Benada, O., Kofroňová, O., Szwedziak, P., Cysewski, D., Gruchota, J., Szpila, M., Chlebowski, A., Sedlacek, R., Prochazka, J., & Dziembowski, A. (2021). Cytoplasmic polyadenylation by TENT5A is required for proper bone formation. *Cell Reports*, *35*(3), 109015. https://doi.org/10.1016/j.celrep.2021.109015

Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., van Dongen, S., Inoue, K., Enright, A. J., & Schier, A. F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science*, *312*(5770), 75–79. https://doi.org/10.1126/science.1122689

Goldfarb, K. C., & Cech, T. R. (2013). 3′ terminal diversity of MRP RNA and other human noncoding RNAs revealed by deep sequencing. *BMC Molecular Biology*, *14*, 23. https://doi.org/10.1186/1471-2199-14-23

Gonzalez-Garay, M. L. (2016). Introduction to isoform sequencing using Pacific Biosciences technology (Iso-Seq). In J. Wu (Ed.), *Transcriptomics and Gene Regulation. Translational Bioinformatics* (pp. 141–160). Springer. https://doi.org/10.1007/978-94-017-7450-5_5

Gu, H., Gupta, D., & Schoenberg, D. R. (1999). The poly(a)-limiting element is a conserved cis-acting sequence that regulates poly(a) tail length on nuclear pre-mRNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(16), 8943–8948. https://doi.org/10.1073/pnas.96.16.8943

Gupta, J. D., Gu, H., Chernokalskaya, E., Gao, X., & Schoenberg, D. R. (1998). Identification of two cis-acting elements that independently regulate the length of poly(a) on Xenopus albumin pre-mRNA. *RNA*, *4*(7), 766–776. https://doi.org/10.1017/s1355838298971837

Harrison, P. F., Powell, D. R., Clancy, J. L., Preiss, T., Boag, P. R., Traven, A., Seemann, T., & Beilharz, T. H. (2015). PAT-seq: A method to study the integration of 3′-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA*, *21*(8), 1502–1510. https://doi.org/10.1261/rna.048355.114

Hennig, B. P., Velten, L., Racke, I., Tu, C. S., Thoms, M., Rybin, V., Besir, H., Remans, K., & Steinmetz, L. M. (2018). Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3*: *Genes*, *genomes*. *Genetics*, *8*(1), 79–89. https://doi.org/10.1534/g3.117.300257

Henriques, T., Gilchrist, D. A., Nechaev, S., Bern, M., Muse, G. W., Burkholder, A., Fargo, D. C., & Adelman, K. (2013). Stable pausing by rna polymerase II provides an opportunity to target and integrate regulatory signals. *Molecular Cell*, *52*(4), 517–528. https://doi.org/10.1016/j.molcel.2013.10.001

Holmquist, C. E., & Marzluff, W. F. (2019). Determining degradation intermediates and the pathway of 3′ to 5′ degradation of histone mRNA using high-throughput sequencing. *Methods*, *155*, 104–115. https://doi.org/10.1016/j.ymeth.2018.11.001

Ibrahim, F., Oppelt, J., Maragkakis, M., & Mourelatos, Z. (2021). TERA-Seq: True end-to-end sequencing of native RNA molecules for transcriptome characterization. *Nucleic Acids Research*, *1–18*, e115. https://doi.org/10.1093/nar/gkab713

Jacobson, A. (1987). Purification and fractionation of poly(a)+ RNA. *Methods in Enzymology*, *152*, 254–261.

Jänicke, A., Vancuylenberg, J., Boag, P. R., Traven, A., & Beilharz, T. H. (2012). ePAT: A simple method to tag adenylated RNA to measure poly(a)-tail length and other 3′ RACE applications. *RNA*, *18*(6), 1289–1295. https://doi.org/10.1261/rna.031898.111

Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., di Padova, F., Lin, S. C., Gram, H., & Han, J. (2005). Involvement of MicroRNA in AU-rich element-mediated mRNA instability. *Cell*, *120*(5), 623–634. https://doi.org/10.1016/j.cell.2004.12.038

Johnstone, C. N., Pattison, A. D., Gorringe, K. L., Harrison, P. F., Powell, D. R., Lock, P., Baloyan, D., Ernst, M., Stewart, A. G., Beilharz, T. H., & Anderson, R. L. (2018). Functional and genomic characterisation of a xenograft model system for the study of metastasis in triple-negative breast cancer. *Disease Models & Mechanisms*, *11*(5), dmm032250. https://doi.org/10.1242/dmm.032250

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell*, *1–8*, 914–921.e10. https://doi.org/10.1016/j.cell.2020.04.011

Kim, J. H., & Richter, J. D. (2006). Opposing polymerase-Deadenylase activities regulate cytoplasmic polyadenylation. *Molecular Cell*, *24*(2), 173–183. https://doi.org/10.1016/j.molcel.2006.08.016

Krause, M., Niazi, A. M., Labun, K., Torres Cleuren, Y. N., Müller, F. S., & Valen, E. (2019). Tailfindr: Alignment-free poly(a) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA*, *25*(10), 1229–1241. https://doi.org/10.1261/rna.071332.119

Kusov, Y. Y., Shatirishvili, G., Dzagurov, G., & Verena, G.-M. (2001). A new G-tailing method for the determination of the poly(a) tail length applied to hepatitis a virus RNA. *Nucleic Acids Research*, *29*(12), e57. https://doi.org/10.1093/nar/29.12.e57

Lackey, P. E., Welch, J. D., & Marzluff, W. F. (2016). TUT7 catalyzes the uridylation of the 3′ end for rapid degradation of histone mRNA. *RNA*, *22*(11), 1673–1688. https://doi.org/10.1261/rna.058107.116

Legnini, I., Alles, J., Karaiskos, N., Ayoub, S., & Rajewsky, N. (2019). FLAM-seq: Full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nature Methods*, *16*(9), 879–886. https://doi.org/10.1038/s41592-019-0503-y

Lim, J., Ha, M., Chang, H., Kwon, S. C., Simanshu, D. K., Patel, D. J., & Kim, V. N. (2014). Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell*, *159*(6), 1365–1376. https://doi.org/10.1016/j.cell.2014.10.055

Lim, J., Kim, D., Lee, Y. S., Ha, M., Lee, M., Yeo, J., Chang, H., Song, J., Ahn, K., & Kim, V. N. (2018). Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. *Science*, *361*(6403), 701–704. https://doi.org/10.1126/science.aam5794

Lim, J., Lee, M., Son, A., Chang, H., & Kim, V. N. (2016). mTAIL-seq reveals dynamic poly(a) tail regulation in oocyte-to-embryo development. *Genes and Development*, *30*(14), 1671–1682. https://doi.org/10.1101/gad.284802.116

Lima, S. A., Chipman, L. B., Nicholson, A. L., Chen, Y. H., Yee, B. A., Yeo, G. W., Coller, J., & Pasquinelli, A. E. (2017). Short poly(a) tails are a conserved feature of highly expressed genes. *Nature Structural and Molecular Biology*, *24*(12), 1057–1063. https://doi.org/10.1038/nsmb.3499

Liu, H., Begik, O., Lucas, M. C., Ramirez, J. M., Mason, C. E., Wiener, D., Schwartz, S., Mattick, J. S., Smith, M. A., & Novoa, E. M. (2019). Accurate detection of m6A RNA modifications in native RNA sequences. *Nature Communications*, *10*(1), 1–9. https://doi.org/10.1038/s41467-019-11713-9

Liu, Y., Nie, H., Liu, H., & Lu, F. (2019). Poly(a) inclusive RNA isoform sequencing (PAIso−seq) reveals wide-spread non-adenosine residues within RNA poly(a) tails. *Nature Communications*, *10*(1), 5292. https://doi.org/10.1038/s41467-019-13228-9

Liudkovska, V., & Dziembowski, A. (2021). Functions and mechanisms of RNA tailing by metazoan terminal nucleotidyltransferases. *WIREs RNA*, *12*(2), e1622. https://doi.org/10.1002/WRNA.1622

Long, Y., Jia, J., Mo, W., Jin, X., & Zhai, J. (2021). FLEP-seq: Simultaneous detection of RNA polymerase II position, splicing status, poly-adenylation site and poly(a) tail length at genome-wide scale by single-molecule nascent RNA sequencing. *Nature Protocols*, *16*(9), 4355–4381. https://doi.org/10.1038/s41596-021-00581-7

Lorenz, D. A., Sathe, S., Einstein, J. M., & Yeo, G. W. (2020). Direct RNA sequencing enables m6A detection in endogenous transcript isoforms at base-specific resolution. *RNA*, *26*(1), 19–28. https://doi.org/10.1261/rna.072785.119

Maier, K. C., Gressel, S., Cramer, P., & Schwalb, B. (2020). Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Research*, *30*(9), 1332–1344. https://doi.org/10.1101/GR.257857.119

Mattijssen, S., Iben, J. R., Li, T., Coon, S. L., & Maraia, R. J. (2020). Single molecule poly(a) tail-seq shows larp4 opposes deadenylation throughout mrna lifespan with most impact on short tails. *eLife*, *9*, 1–18. https://doi.org/10.7554/ELIFE.59186

Meaux, S. A., Holmquist, C. E., & Marzluff, W. F. (2018). Role of oligouridylation in normal metabolism and regulated degradation of mammalian histone mRNAs. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *373*(1762), 20180170. https://doi.org/10.1098/rstb.2018.0170

Meijer, H. A., Bushell, M., Hill, K., Gant, T. W., Willis, A. E., Jones, P., & de Moor, C. H. (2007). A novel method for poly(a) fractionation reveals a large population of mRNAs with a short poly(a) tail in mammalian cells. *Nucleic Acids Research*, *35*(19), e132. https://doi.org/10.1093/nar/gkm830

Meijer, H. A., & de Moor, C. H. (2011). Fractionation of mRNA based on the length of the poly(a) tail. *Methods in Molecular Biology*, *703*(2), 123–135. https://doi.org/10.1007/978-1-59745-248-9

Minasaki, R., Rudel, D., & Eckmann, C. R. (2014). 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 Biology*, *11*(2), 111–123. https://doi.org/10.4161/rna.27992

Morgan, M., Much, C., DiGiacomo, M., Azzi, C., Ivanova, I., Vitsios, D. M., Pistolic, J., Collier, P., Moreira, P. N., Benes, V., Enright, A. J., & O'Carroll, D. (2017). mRNA 3′ uridylation and poly(a) tail length sculpt the mammalian maternal transcriptome. *Nature*, *548*(7667), 347–351. https://doi.org/10.1038/nature23318

Murray, E. L., & Schoenberg, D. R. (2008). Assays for determining poly(a) tail length and the polarity of mRNA decay in mammalian cells. In *Methods in Enzymology* (Vol. 448, 1st ed., pp. 483–504). Springer. https://doi.org/10.1016/S0076-6879(08)02624-4

Novoa, I., Gallego, J., Ferreira, P. G., & Mendez, R. (2010). Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nature Cell Biology*, *12*(5), 447–456. https://doi.org/10.1038/ncb2046

Palatnik, C. M., Storti, R. V., & Jacobson, A. (1979). Fractionation and functional analysis of newly synthesized and decaying messenger RNAs from vegetative cells of Dictyostelium discoideum. *Journal of Molecular Biology*, *128*(3), 371–395. https://doi.org/10.1016/0022-2836(79)90093-7

Pallares, L. F., Picard, S., & Ayroles, J. F. (2020). Tm3'Seq: A tagmentation-mediated 3′ sequencing approach for improving scalability of RNAseq experiments. *G3: Genes, Genomes, Genetics*, *10*(1), 143–150. https://doi.org/10.1534/g3.119.400821

Park, J. E., Yi, H., Kim, Y., Chang, H., & Kim, V. N. (2016). Regulation of poly(a) tail and translation during the somatic cell cycle. *Molecular Cell*, *62*(3), 462–471. https://doi.org/10.1016/j.molcel.2016.04.007

Parker, M. T., Knop, K., Sherwood, A. V., Schurch, N. J., Mackinnon, K., Gould, P. D., Hall, A. J., Barton, G. J., & Simpson, G. G. (2020). Nanopore direct RNA sequencing maps the complexity of arabidopsis mRNA processing and m6A modification. *eLife*, *9*, 1–35. https://doi.org/10.7554/eLife.49658

Peng, J., Murray, E. L., & Schoenberg, D. R. (2005). The poly(a)-limiting element enhances mRNA accumulation by increasing the efficiency of pre-mRNA 3 0 processing. *RNA*, *11*(6), 958–965. https://doi.org/10.1261/rna.2020805

Picelli, S., Björklund, A. K., Reinius, B., Sagasser, S., Winberg, G., & Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research*, *24*(12), 2033–2040. https://doi.org/10.1101/gr.177881.114

Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using smart-seq2. *Nature Protocols*, *9*(1), 171–181. https://doi.org/10.1038/nprot.2014.006

Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell*, *108*(4), 501–512. https://doi.org/10.1016/S0092-8674(02)00617-7

Rädle, B., Rutkowski, A., Ruzsics, Z., Friedel, C. C., Koszinowski, U. H., & Dölken, L. (2013). Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. *Journal of Visualized Experiments*, *78*, 1–11. https://doi.org/10.3791/50195

Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, *19*(1), 1–11. https://doi.org/10.1186/s13059-018-1462-9

Richter, J. D. (2007). CPEB: A life in translation. *Trends in Biochemical Sciences*, *32*(6), 279–285. https://doi.org/10.1016/j.tibs.2007.04.004

Roach, N. P., Sadowski, N., Alessi, A. F., Timp, W., Taylor, J., & Kim, J. K. (2020). The full-length transcriptome of C. elegans using direct RNA sequencing. *Genome Research*, *30*(2), 299–312. https://doi.org/10.1101/gr.251314.119

Roake, C. M., Chen, L., Chakravarthy, A. L., Ferrell, J. E., Raffa, G. D., & Artandi, S. E. (2019). Disruption of telomerase RNA maturation kinetics precipitates disease. *Molecular Cell*, *74*(4), 688–700.e3. https://doi.org/10.1016/j.molcel.2019.02.033

Salles, F. J., Darrow, A. L., O'Connell, M. L., & Strickland, S. (1992). Isolation of novel murine maternal mRNAs regulated by cytoplasmic polyadenylation. *Genes and Development*, *6*(7), 1202–1212. https://doi.org/10.1101/gad.6.7.1202

Sallés, F. J., & Strickland, S. (1999). Analysis of poly(a) tail lengths by PCR: The PAT assay. *Methods in Molecular Biology (Clifton, N.J.)*, *118*, 441–448. https://doi.org/10.1385/1-59259-676-2:441

Sallés, F. J., Richards, W. G., & Strickland, S. (1999). Assaying the polyadenylation state of mRNAs. *Methods: A companion to Methods in Enzymology*, *17*(1), 38–45. https://doi.org/10.1006/meth.1998.0705

Sallés, F. J., & Strickland, S. (1995). Rapid and sensitive analysis of mRNA polyadenylation states by PCR. *Genome Research*, *4*(6), 317–321. https://doi.org/10.1101/gr.4.6.317

Shin, J., Paek, K. Y., Ivshina, M., Stackpole, E., & Richter, J. D. (2017). Essential role for non-canonical poly(a) polymerase GLD4 in cytoplasmic polyadenylation and carbohydrate metabolism. *Nucleic Acids Research*, *45*(11), 6793–6804. https://doi.org/10.1093/nar/gkx239

Sippel, A. E., Stavrianopoulos, J. G., Schutz, G., & Feigelson, P. (1974). Translational properties of rabbit globin mRNA after specific removal of poly(a) with ribonuclease H. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(11), 4635–4639. https://doi.org/10.1073/pnas.71.11.4635

Slevin, M. K., Meaux, S., Welch, J. D., Bigler, R., Miliani de Marval, P. L., Su, W., Rhoads, R. E., Prins, J. F., & Marzluff, W. F. (2014). Deep sequencing shows multiple oligouridylations are required for 3′ to 5′ degradation of histone mRNAs on polyribosomes. *Molecular Cell*, *53*(6), 1020–1030. https://doi.org/10.1016/j.molcel.2014.02.027

Slobodin, B., Bahat, A., Sehrawat, U., Becker-Herman, S., Zuckerman, B., Weiss, A. N., Han, R., Elkon, R., Agami, R., Ulitsky, I., Shachar, I., & Dikstein, R. (2020). Transcription dynamics regulate poly ( a ) tails and expression of the RNA degradation machinery to balance mRNA levels article transcription dynamics regulate poly ( a ) tails and expression of the RNA degradation machinery to balance mRNA levels. *Molecular Cell*, *78*, 434–444.e5. https://doi.org/10.1016/j.molcel.2020.03.022

Son, A., Park, J. E., & Kim, V. N. (2018). PARN and TOE1 constitute a 3′ end maturation module for nuclear non-coding RNAs. *Cell Reports*, *23*(3), 888–898. https://doi.org/10.1016/j.celrep.2018.03.089

Soneson, C., Yao, Y., Bratus-Neuenschwander, A., Patrignani, A., Robinson, M. D., & Hussain, S. (2019). A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nature Communications*, *10*(1), 1–14. https://doi.org/10.1038/s41467-019-11272-z

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: The teenage years. *Nature Reviews Genetics*, *20*(11), 631–656. https://doi.org/10.1038/s41576-019-0150-2

Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H., & Bartel, D. P. (2014). Poly(a)-tail profiling reveals an embryonic switch in translational control. *Nature*, *508*(1), 66–71. https://doi.org/10.1038/nature13007

Suh, N., Jedamzik, B., Eckmann, C. R., Wickens, M., & Kimble, J. (2006). The GLD-2 poly(a) polymerase activates gld-1 mRNA in the Caenorhabditis elegans germ line. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(41), 15108–15112. https://doi.org/10.1073/pnas.0607050103

Swaminathan, A., Harrison, P. F., Preiss, T., & Beilharz, T. H. (2019). PAT-Seq: A method for simultaneous quantitation of gene expression, poly(a)-site selection and poly(a)-length distribution in yeast transcriptomes. In S. Oliver & J. Cawstrillo (Eds.), *Yeast Systems Biology. Methods in Molecular Biology* (Vol. 2049, pp. 141–164). Springer. https://doi.org/10.1111/j.1567-1364.2007.00346.x

Tian, B., Hu, J., Zhang, H., & Lutz, C. S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Research*, *33*(1), 201–212. https://doi.org/10.1093/nar/gki158

Tian, B., & Manley, J. L. (2016). Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology*, *18*(1), 18–30. https://doi.org/10.1038/nrm.2016.116

Tucey, T. M., Verma, J., Harrison, P. F., Snelgrove, S. L., Lo, T. L., Scherer, A. K., Barugahare, A. A., Powell, D. R., Wheeler, R. T., Hickey, M. J., Beilharz, T. H., Naderer, T., & Traven, A. (2018). Glucose homeostasis is important for immune cell viability during Candida challenge and host survival of systemic fungal infection. *Cell Metabolism*, *27*(5), 988–1006.e7. https://doi.org/10.1016/j.cmet.2018.03.019

Tudek, A., Krawczyk, P. S., Mroczek, S., Tomecki, R., Turtola, M., Matylla-Kulińska, K., Jensen, T. H., & Dziembowski, A. (2021). Global view on the metabolism of RNA poly(a) tails in yeast *Saccharomyces cerevisiae*. *Nature Communications*, *12*(1), 4951. https://doi.org/10.1038/s41467-021-25251-w

Tudek, A., Lloret-Llinares, M., & Heick Jensen, T. (2018). The multitasking polyA tail: Nuclear RNA maturation, degradation and export. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *373*(1762), 20180169. https://doi.org/10.1098/RSTB.2018.0169

Turtola, M., Manav, M. C., Kumar, A., Tudek, A., Mroczek, S., Krawczyk, P. S., Dziembowski, A., Schmid, M., Passmore, L. A., Casañal, A., & Jensen, T. H. (2021). Three-layered control of mRNA poly(A) tail synthesis in *Saccharomyces cerevisiae*. *Genes & Development*, *35*, 1290–1303. https://doi.org/10.1101/gad.348634.121

Viegas, I. J., de Macedo, J. P., Serra, L., de Niz, M., Temporão, A., Silva Pereira, S., Mirza, A. H., Bergstrom, E., Rodrigues, J. A., Aresta-Branco, F., Jaffrey, S. R., & Figueiredo, L. M. (2022). N6-methyladenosine in poly(a) tails stabilize VSG transcripts. *Nature*, *604*, 362–370. https://doi.org/10.1038/s41586-022-04544-0

Villalba, A., Coll, O., & Gebauer, F. (2011). Cytoplasmic polyadenylation and translational control. *Current Opinion in Genetics and Development*, *21*(4), 452–457. https://doi.org/10.1016/j.gde.2011.04.006

Vlasova, I. A., Tahoe, N. M., Fan, D., Larsson, O., Rattenbacher, B., Sternjohn, J. R., Vasdewani, J., Karypis, G., Reilly, C. S., Bitterman, P. B., & Bohjanen, P. R. (2008). Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1. *Molecular Cell*, *29*(2), 263–270. https://doi.org/10.1016/j.molcel.2007.11.024

Wahle, E., & Rüegsegger, U. (1999). 3′-end processing of pre-mRNA in eukaryotes. *FEMS Microbiology Reviews*, *23*(3), 277–295. https://doi.org/10.1016/S0168-6445(99)00008-X

Wakiyama, M., Imataka, H., & Sonenberg, N. (2000). Interaction of eIF4G with poly(a)-binding protein stimulates translation and is critical for Xenopus oocyte maturation. *Current Biology*, *10*(18), 1147–1150. https://doi.org/10.1016/S0960-9822(00)00701-6

Wang, Z., & Kiledjian, M. (2001). Functional link between the mammalian exosome and mRNA Decapping. *Cell*, *107*(6), 751–762. https://doi.org/10.1016/S0092-8674(01)00592-X

Warkocki, Z., Krawczyk, P. S., Adamska, D., Bijata, K., Garcia-Perez, J. L., & Dziembowski, A. (2018). Uridylation by TUT4/7 restricts Retrotransposition of human LINE-1s. *Cell*, *174*(6), 1537–1548.e29. https://doi.org/10.1016/j.cell.2018.07.022

Weill, L., Belloc, E., Bava, F. A., & Méndez, R. (2012). Translational control by changes in poly(a) tail length: Recycling mRNAs. *Nature Structural and Molecular Biology*, *19*(6), 577–585. https://doi.org/10.1038/nsmb.2311

Welch, J. D., Slevin, M. K., Tatomer, D. C., Duronio, R. J., Prins, J. F., & Marzluff, W. F. (2015). EnD-Seq and AppEnD: Sequencing 3′ ends to identify nontemplated tails and degradation intermediates. *RNA*, *21*(7), 1375–1389. https://doi.org/10.1261/rna.048785.114

Wells, S. E., Hillner, P. E., Vale, R. D., & Sachs, A. B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. *Molecular Cell*, *2*(1), 135–140. https://doi.org/10.1016/S1097-2765(00)80122-7

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., Hall, R. J., Concepcion, G. T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N. D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C. S., Phillippy, A. M., Schatz, M. C., Myers, G., DePristo, M., ... Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, *37*(10), 1155–1162. https://doi.org/10.1038/s41587-019-0217-9

Wilusz, J. E., Whipple, J. M., Phizicky, E. M., & Sharp, P. A. (2011). tRNAs marked with CCACCA are targeted for degradation. *Science*, *334*(6057), 817–821. https://doi.org/10.1126/science.1213671

Woo, Y. M., Kwak, Y., Namkoong, S., Kristjánsdóttir, K., Lee, S. H., Lee, J. H., & Kwak, H. (2018). TED-Seq identifies the dynamics of poly(A) length during ER stress. *Cell Reports*, *24*(13), 3630–3641.e7. https://doi.org/10.1016/j.celrep.2018.08.084

Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., Zuzarte, P. C., Gilpatrick, T., Payne, A., Quick, J., Sadowski, N., Holmes, N., de Jesus, J. G., Jones, K. L., Soulette, C. M., Snutch, T. P., Loman, N., Paten, B., Loose, M., ... Timp, W. (2019). Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods*, *16*(12), 1297–1305. https://doi.org/10.1038/s41592-019-0617-2

Wu, L., Fan, J., & Belasco, J. G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(11), 4034–4039. https://doi.org/10.1073/pnas.0510928103

Xiang, K., & Bartel, D. P. (2021). The molecular basis of coupling between poly(A)-tail length and translational efficiency. *eLife*, *10*, 1–39. https://doi.org/10.7554/ELIFE.66493

Xu, N., Chen, C. Y., & Shyu, A. B. (1997). Modulation of the fate of cytoplasmic mRNA by AU-rich elements: Key sequence features controlling mRNA deadenylation and decay. *Molecular and Cellular Biology*, *17*(8), 4611–4621. https://doi.org/10.1128/mcb.17.8.4611

Yang, F., Wang, W., Cetinbas, M., Sadreyev, R. I., & Blower, M. D. (2020). Genome-wide analysis identifies cis-acting elements regulating mRNA polyadenylation and translation during vertebrate oocyte maturation. *RNA*, *26*(3), 324–344. https://doi.org/10.1261/rna.073247.119

Yu, F., Zhang, Y., Cheng, C., Wang, W., Zhou, Z., Rang, W., Yu, H., Wei, Y., Wu, Q., & Zhang, Y. (2020). Poly(A)-seq: A method for direct sequencing and analysis of the transcriptomic poly(A)-tails. *PLoS One*, *15*(6), 1–21. https://doi.org/10.1371/journal.pone.0234696

Zhao, L., Zhang, H., Kohnen, M. V., Prasad, K. V. S. K., Gu, L., & Reddy, A. S. N. (2019). Analysis of transcriptome and epitranscriptome in plants using pacbio iso-seq and nanopore-based direct RNA sequencing. *Frontiers in Genetics*, *10*, 1–14. https://doi.org/10.3389/fgene.2019.00253

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One*, *9*(1), e78644. https://doi.org/10.1371/journal.pone.0078644

Zhao, T., Huan, Q., Sun, J., Liu, C., Hou, X., Yu, X., Silverman, I. M., Zhang, Y., Gregory, B. D., Liu, C. M., Qian, W., & Cao, X. (2019). Impact of poly(A)-tail G-content on Arabidopsis PAB binding and their role in enhancing translational efficiency. *Genome Biology*, *20*(1), 1–12. https://doi.org/10.1186/s13059-019-1799-8

Zhou, M., Bhasin, A., & Reznikoff, W. S. (1998). Molecular genetic analysis of transposase-end DNA sequence recognition: Cooperativity of three adjacent base-pairs in specific interaction with a mutant Tn5 transposase. *Journal of Molecular Biology*, *276*(5), 913–925. https://doi.org/10.1006/jmbi.1997.1579

Zuber, H., Scheer, H., Ferrier, E., Sement, F. M., Mercier, P., Stupfler, B., & Gagliardi, D. (2016). Uridylation and PABP cooperate to repair mRNA Deadenylated ends in Arabidopsis. *Cell Reports*, *14*(11), 2707–2717. https://doi.org/10.1016/j.celrep.2016.02.060