UC Irvine

UC Irvine Previously Published Works

Title

PAS-seq 2: A fast and sensitive method for global profiling of polyadenylated RNAs

Permalink

https://escholarship.org/uc/item/8gw8h4bt

Authors

Yoon, Yoseop
Soles, Lindsey V
Shi, Yongsheng

Publication Date

2021

DOI

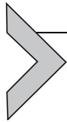
10.1016/bs.mie.2021.03.013

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library University of California



CHAPTER TWO

PAS-seq 2: A fast and sensitive method for global profiling of polyadenylated RNAs

Yoseop Yoon, Lindsey V. Soles, and Yongsheng Shi*
Department of Microbiology and Molecular Genetics, School of Medicine, University of California, Irvine, CA, United States
*Corresponding author: e-mail address: yongshes@uci.edu

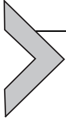
Contents

1. Introduction 26
2. Materials 27
   2.1 Solutions 27
   2.2 Reagents and equipment 27
   2.3 Primer sequences 28
3. Methods 28
   3.1 Isolation of total RNA from biological samples (1–1.5 h (hrs)) 28
   3.2 Isolation of poly(A) RNA (1–1.5 h) 28
   3.3 RNA fragmentation (1.5 h) 29
   3.4 Reverse transcription (3–3.5 h) 30
   3.5 Purification of cDNA (0.5 h) 31
   3.6 PCR amplification and size selection (3.5 h) 31
4. Data analysis 32
5. Troubleshooting tips 32
Acknowledgments 33
References 33

Abstract

Alternative polyadenylation (APA) is a widespread phenomenon in eukaryotes that contributes to regulating gene expression and generating proteomic diversity. APA plays critical roles in development and its mis-regulation has been implicated in a wide variety of human diseases, including cancer. To study APA on the transcriptome-wide level, numerous deep sequencing methods that capture 3′ end of mRNAs have been developed in the past decade, but they generally require a large amount of hands-on time and/or high RNA input. Here, we introduce PAS-seq 2, a fast and sensitive method for global and quantitative profiling of polyadenylated RNAs. Compared to our original PAS-seq, this method takes less time and requires much lower total RNA input due to improvement in the reverse transcription process. PAS-seq 2 can be applied to both APA and differential gene expression analyses.

Methods in Enzymology, Volume 655
ISSN 0076-6879
https://doi.org/10.1016/bs.mie.2021.03.013
Copyright © 2021 Elsevier Inc.
All rights reserved.



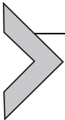
1. Introduction

Alternative polyadenylation (APA) allows a single gene to generate multiple transcript isoforms through the use of alternative poly(A) sites (PAS). Approximately 70% of mammalian genes are subjected to APA (Shi, 2012; Tian & Manley, 2017). APA plays important roles in gene regulation (Mittleman et al., 2020; Sandberg, Neilson, Sarma, Sharp, & Burge, 2008), protein localization (Berkovits & Mayr, 2015), cell differentiation (Zhu et al., 2018), immune response (Rogers et al., 1980), development and evolution (Yoon et al., 2019). Aberrant APA regulation has been linked to many diseases, including cancer and neurological disorders (Gruber & Zavolan, 2019).

APA studies have been facilitated by deep sequencing-based methods for transcriptome-wide profiling of polyadenylated RNAs. These methods can not only map PAS globally, but also quantify their relative usage frequency, thereby allowing comparison of differential PAS usage among different cell types, tissues, and biological conditions. For sequencing library construction, most methods use an oligo(dT) primer to selectively reverse transcribe (RT) polyadenylated RNAs (Derti et al., 2012; Fox-Walsh, Davis-Turak, Zhou, Li, & Fu, 2011; Fu et al., 2011; Lianoglou, Garg, Yang, Leslie, & Mayr, 2013; Martin, Gruber, Keller, & Zavolan, 2012; Sanfilippo, Miura, & Lai, 2017; Shepard et al., 2011; Yoon & Brem, 2010; Zhou et al., 2016). While oligo(dT) primer could also bind internal A-rich sequences during RT, a phenomenon called internal priming, it can be mitigated by computational filtering and use of annotated PAS database for data analyses (Derti et al., 2012; Shepard et al., 2011). Several other methods circumvent the problem of internal priming by linking a 3′ adapter for RT to polyadenylated RNAs (Hoque et al., 2013; Jan, Friedman, Ruby, & Bartel, 2011; Zheng, Liu, & Tian, 2016), or enriching polyadenylated RNAs that bind poly(A) binding proteins (Hwang et al., 2016). However, these methods tend to be laborious compared to oligo(dT) primer-based methods. The large number of steps could also introduce biases that may prevent accurate quantification (Spies, Burge, & Bartel, 2013). Finally, most methods developed so far require high RNA input, up to 10–20 μg of total RNA or 1 μg of total RNA.

Here, we introduce PAS-Seq 2, an improved version of PAS-seq (Shepard et al., 2011), that allows library construction from 100 ng of total RNA. The original PAS-seq combines RT and adapter linking into a single

step and has been successfully used in many APA studies, but requires 5–10 μg of total RNA input (Brumbaugh et al., 2018; Huang et al., 2017; Wang et al., 2020; Zhu et al., 2018). In PAS-seq 2, we adopted the reverse transcription protocol from SMART-seq 2 (Picelli et al., 2014), which is intended for samples with low RNA input such as single cells. Specifically, PAS-seq 2 uses a template switch oligo (TSO) that contains a locked nucleic acid (LNA) at the 3′ end, thereby enhancing TSO annealing to untemplated 3′ extension of the cDNA. Also, the protocol includes betaine and magnesium chloride in the RT reaction, which together can increase cDNA yield (Picelli et al., 2014). In our hands, PAS-seq 2 libraries can be successfully generated from 100 ng total RNA. This protocol may work with even lower RNA input if the fragmentation step is optimized.



2. Materials

2.1 Solutions

1. 1 M Magnesium Chloride ($MgCl_2$)
2. 100% Ethanol (EtOH)
3. 3 M Sodium acetate (NaOAc)
4. Betaine, 5 M solution (Thermo Scientific AAJ77507UCR)
5. 10 mM dNTP mix
6. 100% isopropyl alcohol
7. RNase-free water
8. Distilled water

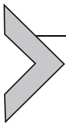
2.2 Reagents and equipment

1. TRIzol™ Reagent (Invitrogen 15596026)
2. RNA fragmentation reagents (Invitrogen AM8740)
3. NEBNext® Poly(A) mRNA magnetic isolation module (NEB E7940)
4. NanoDrop®
5. Heat block or water bath
6. −80 °C freezer
7. Temperature-controlled tabletop centrifuge
8. SuperScript™ III first-strand synthesis system (Invitrogen 18080051)
9. AMPure XP (Beckman-Coulter A63880)
10. Magnetic stand
11. Thermal cycler

12. Phusion® high-fidelity PCR master mix with HF buffer (NEB M0531S)
13. RNaseOUT™ recombinant ribonuclease inhibitor (Invitrogen 1077 7019)
14. GeneJET gel extraction kit (Thermo Scientific K0691)
15. GlycoBlue™ Coprecipitant (Invitrogen AM9516)

2.3 Primer sequences

1. Template switch oligo (TSO) containing locked nucleic acid (LNA): CTACACGACGCTCTTCCGATCTCATrGrG+G
2. PAS-seq 2 oligo(dT): GTGACTGGAGTTCAGACGTGTGCTCT TCCGATCTTTTTTTTTTTTTTTTTTTTV (V: A/C/G)
3. TruSeq universal adapter: AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCT
4. TruSeq indexed adapter: CAAGCAGAAGACGGCATACGAGAT**[index]**GTGACTGGA GTTCAGACGTGTGCTCTTCCGATC.



3. Methods

3.1 Isolation of total RNA from biological samples (1–1.5 h (hrs))

1. Isolate total RNA from the biological sample using TRIzol™ Reagent, as per manufacturer's instructions. During the final steps, wash pellet twice instead of once with 75% EtOH. Air-dry and resuspend the pellet with 30 μL RNase-free water. Incubate in a water bath or heat block set at 55–60 °C for 10–15 min. Determine RNA concentration using NanoDrop® and store the samples at −80 °C until use (Fig. 1)

3.2 Isolation of poly(A) RNA (1–1.5 h)

1. Dilute 100 ng—5 μg of total RNA in 50 μL of RNase-free H_2O
2. Isolate poly(A) RNA from total RNA using NEBNext® Poly(A) mRNA magnetic isolation module, as per manufacturer's instructions. Elute the poly(A) RNA using 17 μL of the Tris buffer included in the kit
3. Add 28 μL H_2O to the eluted mRNA, making the final volume of 45 μL
4. Place tube on ice

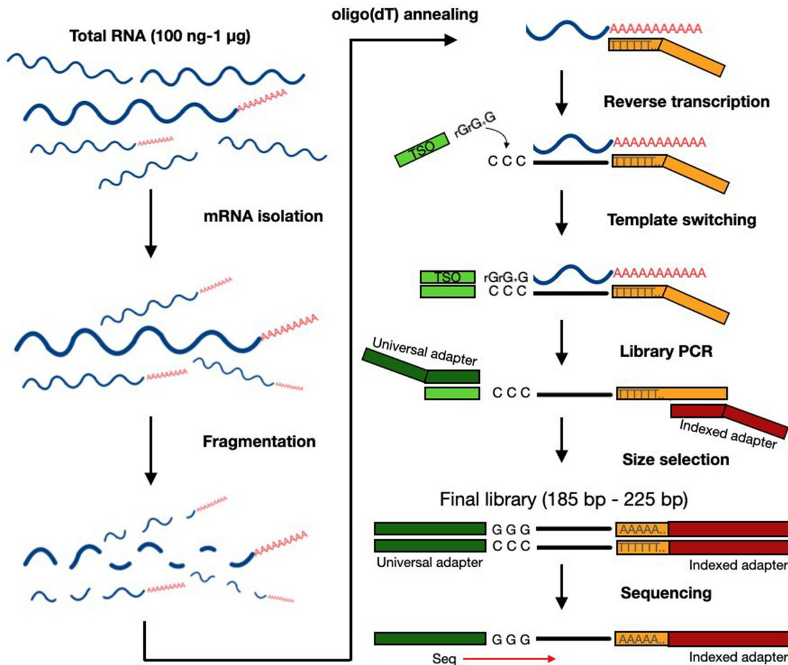


Fig. 1 Overview of PAS-seq 2. TSO: template switch oligo.

3.3 RNA fragmentation (1.5 h)

1. For 1 μg of starting total RNA, add 5 μL of 10 × RNA fragmentation reagent (final volume 50 μL), mix well, and incubate at 70 °C in a thermal cycler for 5 min (min).
2. Add 5.5 μL of Stop solution (10 ×), mix well, and place tube on ice for 2 min
3. Add following mixture to the tube:
   - 144.5 μL H_2O
   - 20 μL 3 M NaOAc
   - 1 μL GlycoBlue™
   - 660 μL ice cold 100% EtOH
4. Precipitate for 1 h at −80 °C
5. Centrifuge at full speed (13,000 rpm) in 4 °C for 30 min
6. Wash pellet with 1 mL of ice cold 75% EtOH. Centrifuge at full speed (13,000 rpm) in 4 °C for 5 min
7. Dissolve the pellet in 5 μL H_2O

3.4 Reverse transcription (3–3.5 h)

1. Set up oligo(dT) primer binding reaction in a PCR tube on ice:

5 µL	Fragmented poly(A) RNA (from step 3.3.7)
2 µL	10 mM dNTP mix
2 µL	10 µM PAS-seq2 oligo(dT) primer
9 µL	**Total**

2. Incubate at 72 °C for 3 min, and put it back to ice
3. Prepare the following mixture in a master mix:

1 µL	SuperScript III reverse transcriptase
4 µL	5 × First-strand buffer
1 µL	100 mM DTT
0.5 µL	RNaseOUT RNase inhibitor
4 µL	5 M Betaine
0.12 µL	1 M $MgCl_2$
0.2 µL	100 µM template switch oligo containing LNA
0.18 µL	RNase-free H_2O
11 µL	**Total**

4. Add 11 µL of mixture from step 3.4.3 to 9 µL of reaction from step 3.4.1 to obtain a final reaction volume of 20 µL. Mix well
5. Incubate the reaction in a thermal cycler with a heated lid, using the following program:
   - 1: 42 °C, 90 min
   - 2: 50 °C, 2 min
   - 3: 42 °C, 2 min
   - 4: Repeat 2–3 for 9 times
   - 5: 72 °C, 15 min
   - 6: 4 °C, hold

*Addition of betaine permits cycle 2–4 by stabilizing reverse transcriptase in higher temperature and promotes unfolding of RNA secondary structures.

6. Add 30 μL of H_2O to obtain a final volume of 50 μL

3.5 Purification of cDNA (0.5 h)

1. Purify 50 μL of cDNA from step 3.4.6. using AMPure XP, as per manufacturer's instructions. Do not over-dry the beads
2. In the final step, elute cDNA with 25 μL of H_2O
3. Transfer 23 μL of the supernatant (cDNA) to a new tube

3.6 PCR amplification and size selection (3.5 h)

1. Prepare the following mixture:

23 μL	cDNA
25 μL	2 × Phusion Hifi PCR master mix
1 μL	10 μM TruSeq universal adapter
1 μL	10 μM TruSeq indexed adapter
50 μL	**Total**

2. Incubate the reaction in a thermal cycler with a heated lid, using the following program:
   - 1: 98 °C, 30 s
   - 2: 98 °C, 10 s
   - 3: 66 °C, 30 s
   - 4: 72 °C, 20 s
   - 5: Repeat step 2–4 19 times
   - 6: 72 °C, 5 min
   - 7: 4 °C, hold
3. Run all PCR products on a 2.5% agarose gel (80 V, 2.5 h).
4. Gel extract 185–225 base-pair (bp) band and submit samples to the sequencing core etc. for quality check prior to sequencing (Fig. 2)
5. Perform sequencing (100 bp single read) of the libraries as per manufacturer's instructions. Add Phix control or multiplex with other libraries to add diversity

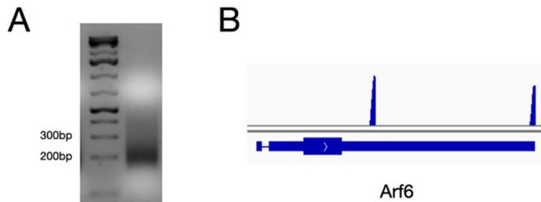
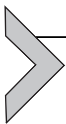
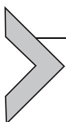


Fig. 2 (A) PAS-seq 2 library. After PCR amplification (Section 3.6), the reaction was resolved on a 2.5% agarose gel. 185–225 bp fragments are selected for sequencing. (B) An example of PAS-seq 2 data. Bigwig file was generated from PAS-seq 2 reads. 3′UTR region of Arf6 gene is shown in the figure. Each peak corresponds alternative cleavage/polyadenylation site for Arf6 gene.



4. Data analysis

1. Trim the sequences by removing the first six nucleotides (CATGGG—from the template switch oligo) and consecutive As (from the oligo(dT) primer sequence) with Cutadapt (Martin, 2011).
2. Align trimmed reads to reference genome using aligner programs such as STAR (Dobin et al., 2013).
3. To filter out internal priming events, remove reads that are mapped to genomic sequences that are immediately followed by 6-consecutive As, or seven As out of 10 nucleotides
4. Compare the 3′ ends of each read to a list of annotated PAS based on published databases (e.g., APADB, PolyA_DB) (Muller et al., 2014; Wang, Nambiar, Zheng, & Tian, 2018). Generate a count table by assigning reads to known PAS if they are mapped within ±40 nt from the site
5. APA analysis is performed with edgeR using the "exon" mode to obtain statistical significance values (false discovery rate (FDR)) for each PAS
6. Genes are considered as APA genes if at least one PAS reaches statistical difference (FDR <0.05) and its percent usage within the gene is different by more than 15 between samples
7. For further bioinformatic analyses, please refer to (Brumbaugh et al., 2018; Shepard et al., 2011).



5. Troubleshooting tips

1. Isolating poly(A) RNA (Section 3.2) from total RNA helps to remove reads from rRNAs, but this protocol works with total RNA as well
2. RNA fragmentation step should be optimized based on the concentration of total RNA input. The current fragmentation protocol is based on

1 μg of total RNA input. Add Stop solution immediately after incubating the samples with RNA fragmentation buffer according to the protocol. Both incubation time and amount of fragmentation buffer can affect average library size. Libraries smaller than 200 bp may contain primer dimers and very short reads

3. Addition of GlycoBlue™ is useful for detecting small pellets during the RNA fragmentation step
4. For gel extraction of PCR amplified libraires, make sure to run the gel sufficiently to remove primer dimers (around 150 bp) from 185 to 225 bp fragments. Check with bioanalyzer to make sure that sample sizes are appropriate and primer dimers are absent from the library
5. PAS-seq 2 reads contain six identical sequences (CATGGG) at the beginning. Therefore, adding spike-in controls such as Phix is recommended. Alternatively, libraries can be multiplexed with other samples to increase diversity at the start of reads
6. While we found that 100 ng of total RNA input was sufficient for generating PAS-seq 2 library, even lower RNA input might be sufficient
7. It is important to perform PAS-seq analysis with biological replicates

Acknowledgments

This study was supported by the following grants: NIH GM090056 and GM128441. We thank the UCI GHTF for sequencing.

References

Berkovits, B. D., & Mayr, C. (2015). Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature*, *522*(7556), 363–367. https://doi.org/10.1038/nature14321.

Brumbaugh, J., Di Stefano, B., Wang, X., Borkent, M., Forouzmand, E., Clowers, K. J., et al. (2018). Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling. *Cell*, *172*(3), 629–631. https://doi.org/10.1016/j.cell.2017.12.035.

Derti, A., Garrett-Engele, P., Macisaac, K. D., Stevens, R. C., Sriram, S., Chen, R., et al. (2012). A quantitative atlas of polyadenylation in five mammals. *Genome Research*, *22*(6), 1173–1183. https://doi.org/10.1101/gr.132563.111.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Fox-Walsh, K., Davis-Turak, J., Zhou, Y., Li, H., & Fu, X. D. (2011). A multiplex RNA-seq strategy to profile poly(A+) RNA: Application to analysis of transcription response and 3′ end formation. *Genomics*, *98*(4), 266–271. https://doi.org/10.1016/j.ygeno.2011.04.003.

Fu, Y., Sun, Y., Li, Y., Li, J., Rao, X., Chen, C., et al. (2011). Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Research*, *21*(5), 741–747. https://doi.org/10.1101/gr.115295.110.

Gruber, A. J., & Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. *Nature Reviews. Genetics*, *20*, 599–614. https://doi.org/10.1038/s41576-019-0145-z.

Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., et al. (2013). Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nature Methods*, *10*(2), 133–139. https://doi.org/10.1038/nmeth.2288.

Huang, C., Shi, J., Guo, Y., Huang, W., Huang, S., Ming, S., et al. (2017). A snoRNA modulates mRNA 3′ end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Research*, *45*(15), 8647–8660. https://doi.org/10.1093/nar/gkx651.

Hwang, H. W., Park, C. Y., Goodarzi, H., Fak, J. J., Mele, A., Moore, M. J., et al. (2016). PAPERCLIP identifies microRNA targets and a role of CstF64/64tau in promoting non-canonical poly(A) site usage. *Cell Reports*, *15*(2), 423–435. https://doi.org/10.1016/j.celrep.2016.03.023.

Jan, C. H., Friedman, R. C., Ruby, J. G., & Bartel, D. P. (2011). Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature*, *469*(7328), 97–101. https://doi.org/10.1038/nature09616.

Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S., & Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes & Development*, *27*(21), 2380–2396. https://doi.org/10.1101/gad.229328.113.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. *[S.l.]*, *17*(1), 102226–126089. https://doi.org/10.14806/ej.17.1.200.

Martin, G., Gruber, A. R., Keller, W., & Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Reports*, *1*(6), 753–763. https://doi.org/10.1016/j.celrep.2012.05.003.

Mittleman, B. E., Pott, S., Warland, S., Zeng, T., Mu, Z., Kaur, M., et al. (2020). Alternative polyadenylation mediates genetic regulation of gene expression. *eLife*, *9*. https://doi.org/10.7554/eLife.57492.

Muller, S., Rycak, L., Afonso-Grunz, F., Winter, P., Zawada, A. M., Damrath, E., et al. (2014). APADB: A database for alternative polyadenylation and microRNA regulation events. *Database (Oxford)*, *2014*, bau076. https://doi.org/10.1093/database/bau076.

Picelli, S., Faridani, O. R., Bjorklund, A. 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.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., et al. (1980). Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin mu chain. *Cell*, *20*(2), 303–312. https://doi.org/10.1016/0092-8674(80)90616-9.

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., & Burge, C. B. (2008). Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science*, *320*(5883), 1643–1647. https://doi.org/10.1126/science.1155390.

Sanfilippo, P., Miura, P., & Lai, E. C. (2017). Genome-wide profiling of the 3′ ends of polyadenylated RNAs. *Methods*, *126*, 86–94. https://doi.org/10.1016/j.ymeth.2017.06.003.

Shepard, P. J., Choi, E.-A., Lu, J., Flanagan, L. A., Hertel, K. J., & Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA (New York)*, *17*(4), 761–772. https://doi.org/10.1261/rna.2581711.

Shi, Y. (2012). Alternative polyadenylation: New insights from global analyses. *RNA*, *18*(12), 2105–2117. https://doi.org/10.1261/rna.035899.112.

Spies, N., Burge, C. B., & Bartel, D. P. (2013). 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Research*, *23*(12), 2078–2090. https://doi.org/10.1101/gr.156919.113.

Tian, B., & Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors [review]. *Nature Reviews Molecular Cell Biology*, *18*(1), 18–30. https://doi.org/10.1038/nrm.2016.116.

Wang, R., Nambiar, R., Zheng, D., & Tian, B. (2018). PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Research*, *46*(D1), D315–D319. https://doi.org/10.1093/nar/gkx1000.

Wang, X., Hennig, T., Whisnant, A. W., Erhard, F., Prusty, B. K., Friedel, C. C., et al. (2020). Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. *Nature Communications*, *11*(1), 293. https://doi.org/10.1038/s41467-019-14109-x.

Yoon, O. K., & Brem, R. B. (2010). Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA*, *16*(6), 1256–1267. https://doi.org/10.1261/rna.2038810.

Yoon, Y., Klomp, J., Martin-Martin, I., Criscione, F., Calvo, E., Ribeiro, J., et al. (2019). Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms. *eLife*, *8*. https://doi.org/10.7554/eLife.46711.

Zheng, D., Liu, X., & Tian, B. (2016). 3'READS+, a sensitive and accurate method for 3′ end sequencing of polyadenylated RNA. *RNA (New York)*, *22*(10), 1631–1639. https://doi.org/10.1261/rna.057075.116.

Zhou, X., Li, R., Michal, J. J., Wu, X. L., Liu, Z., Zhao, H., et al. (2016). Accurate profiling of gene expression and alternative polyadenylation with whole transcriptome termini site sequencing (WTTS-Seq). *Genetics*, *203*(2), 683–697. https://doi.org/10.1534/genetics.116.188508.

Zhu, Y., Wang, X., Forouzmand, E., Jeong, J., Qiao, F., Sowd, G. A., et al. (2018). Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation. *Molecular Cell*, *69*(1), 62–74. e64 https://doi.org/10.1016/j.molcel.2017.11.031.