SCIENCE CHINA
Life Sciences

•LETTER TO THE EDITOR•

January 2023 Vol.66 No.1: 187–190
https://doi.org/10.1007/s11427-022-2186-8

Comprehensive analysis of mRNA poly(A) tails by PAIso-seq2

Yusheng Liu[1,2†*], Hu Nie[2,3†], Yiwei Zhang[2,4†], Falong Lu[2,3*] & Jiaqiang Wang[4*]

[1]*College of Life Science, Northeast Forestry University, Harbin 150040, China;*
[2]*State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China;*
[3]*University of Chinese Academy of Sciences, Beijing 100049, China;*
[4]*College of Life Science, Northeast Agricultural University, Harbin 150030, China*

Received August 7, 2022; accepted August 23, 2022; published online August 29, 2022

Citation: Liu, Y., Nie, H., Zhang, Y., Lu, F., and Wang, J. (2023). Comprehensive analysis of mRNA poly(A) tails by PAIso-seq2. Sci China Life Sci 66, 187–190. https://doi.org/10.1007/s11427-022-2186-8

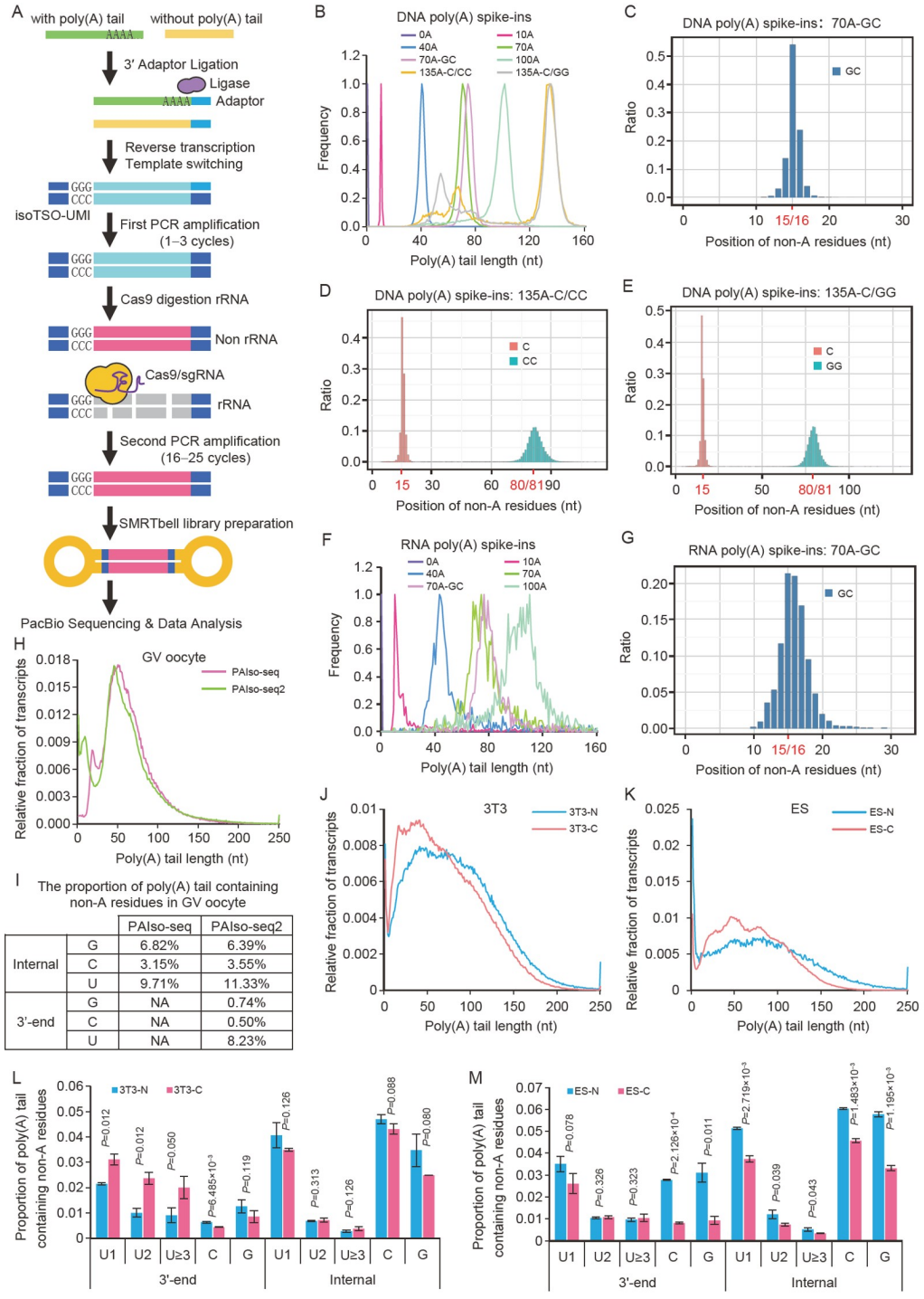
Dear Editor,

Poly(A) tails are added to the 3′-end of most mRNAs (Colgan and Manley, 1997; Yu and Kim, 2020). The regulatory roles of poly(A) tails have long been underestimated due to technical difficulties in analyzing homopolymers (Chang et al., 2014; Subtelny et al., 2014). Several methods were recently developed to accurately measure their length (Chang et al., 2014; Legnini et al., 2019; Liu et al., 2019; Liu et al., 2022; Long et al., 2021; Subtelny et al., 2014). Poly(A) inclusive RNA isoform sequencing (PAIso-seq), TAIL-seq and FLAM-seq, reveal that poly(A) tails contain U, C, and G residues internally and at their 3′-ends (Chang et al., 2014; Legnini et al., 2019; Liu et al., 2019). U or G residues at the 3′-end of poly(A) tails promote RNA degradation or stabilization, respectively (Lim et al., 2014; Lim et al., 2018). However, there is no method for transcriptome-wide measurement of 3′-end and internal non-A residues of poly(A) tails simultaneously. Here, we developed a new method called PAIso-seq2 that allows transcriptome-wide measurement of the poly(A) tail length as well as internal and 3′-end non-A residues together with the full-length cDNA.

PAIso-seq2 employs a direct 3′-end adaptor ligation to preserve all poly(A) tail information (Figure 1A). The adaptor-ligated RNA is then reverse transcribed with a primer complementary to the 3′-adaptor followed by 5′-end tagging through template-switching using an oligo with a unique molecular identifier (isoTSO-UMI). Traditional rRNA removal by hybridization or RNase H digestion are not suitable for low input samples, while Cas9 mediated cDNA removal of abundant species after several round of PCR amplification is suitable for any amount of input sample (Gu et al., 2016). The cDNA is then amplified by a second round of PCR and used for PacBio sequencing in HiFi mode.

To validate the performance of PAIso-seq2, we sequenced a pool of barcoded synthetic cDNA and mRNA spike-ins with different poly(A) tails (Figure S1A, B in Supporting Information), and recovered a similar number of reads for each of the spike-ins (Figure S1C, D in Supporting Information). We observed sharp peaks at or very close to the expected length for the DNA spike-ins (Figure 1B). For the 135 nt tails, secondary peaks of smaller size were observed (Figure 1B), likely resulting from recombination of long homopolymers in *E. coli* (Trepotec et al., 2019). Non-A residues were included within the poly(A) sequence to evaluate the performance of PAIso-seq2 in sequencing them, and were identified at the expected positions (Figure 1C–E). For the *in vitro* transcribed (IVT) mRNA spike-ins, we observed peaks close to the expected sizes (Figure 1F). The reduced resolution of peaks for IVT mRNA spike-ins compared to the DNA spike-ins is likely due to the inaccuracy of IVT in

†Contributed equally to this work
*Corresponding authors (Yusheng Liu, email: liuys1126@foxmail.com; Falong Lu, email: fllu@genetics.ac.cn; Jiaqiang Wang, email: wangjiaqiang@neau.edu.cn)



		PAIso-seq	PAIso-seq2
Internal	G	6.82%	6.39%
	C	3.15%	3.55%
	U	9.71%	11.33%
3′-end	G	NA	0.74%
	C	NA	0.50%
	U	NA	8.23%

Figure 1 The principle and validation of PAIso-seq2. A, Flowchart for PAIso-seq2 method. B and F, Histograms (bin size=1 nt) of poly(A) length for each DNA spike-in (B, detected length: 0, 10, 40, 70, 74, 99, 131, and 131 nt, respectively) or RNA spike-in (F, detected length: 0, 13, 46, 75, 79, and 103 nt, respectively) normalized by the maximum value. C–E and G, Proportion of measured non-A residues at the indicated positions in 70A-GC (C), 135A-C/CC (D), 135A-C/GG (E) DNA spike-ins, and 70A-GC RNA spike-ins (G). The numbers in magenta represent the expected position of the non-A residues. H, Histograms of poly(A) tail length mouse GV oocytes: PAIso-seq (Liu et al., 2019) and PAIso-seq2 (this study). Histograms (bin size=1 nt, 1–250 nt, tails longer than 250 nt are included in the 250 nt bin) are normalized to cover the same area. I, Proportion of transcripts containing non-A residues in GV oocytes detected by PAIso-seq and PAIso-seq2. J and K, Global distribution of poly(A) tail lengths of all transcripts in the 3T3 (J) or ES (K) nuclear (N) and the cytoplasmic (C) fractions. Histograms (bin size=1 nt, 1–250 nt, tails longer than 250 nt are included in the 250 nt bin) are normalized to cover the same area. L, M, Proportion of transcripts containing non-A residues at 3′-end (left) and Internal (right) positions in the 3T3 (L) or ES (M) nuclear (N) and the cytoplasmic (C) fractions. The U residues were further divided according to the length of the longest consecutive U (1, 2, and ≥3). Error bars indicate the standard error of the mean from two replicates. The ratio of CCS reads containing non-A residues shows the number of reads with poly(A) tails containing the indicated non-A residues divided by the total number of reads with poly(A) tails. The differences between the N and C fractions are tested by Student's *t* test. The *P* values are shown for each comparison.

synthesis of the desired length of poly(A) tail from DNA templates as revealed by liquid chromatography-mass spectrometry analysis (Beverly et al., 2018). In addition, the non-A residues within the mRNA poly(A) spike-ins were detected around the expected sites (Figure 1G). Together, these results confirm that PAIso-seq2 can accurately measure both the length and non-A residues of the poly(A) tails.

Next, we measured the mRNA poly(A) tails of 3T3 cells (3T3-T: 3T3 total RNA; 3T3-N: 3T3 nuclear RNA; 3T3-C: 3T3 cytoplasmic RNA) and mouse embryonic stem (ES) cells (ES-T: ES total RNA; ES-N: ES nuclear RNA; ES-C: ES cytoplasmic RNA). Cytoplasmic and nuclear RNA was isolated following a published method (Hu et al., 2015). The PAIso-seq2 data showed good correlations between replicates on gene expression level and poly(A) tail length (Figure S2A, B in Supporting Information), and good coverage of full-length transcripts (Figure S2C, D in Supporting Information). U, C, and G residues were present at the 5′-ends, 3′-ends, and internal positions of poly(A) tails (Figure S3A, B, D in Supporting Information), while the proportion of transcripts harboring non-A residues was comparable between 3T3 and ES cells (Figure S3C in Supporting Information). We investigated the distribution of single, double, and oligo (three or more consecutive) non-A residues at the three different positions (5′-end, 3′-end, or internal that non-A residues are separated by at least one A residue from both ends) in poly(A) tails. C and G existed largely as single residues, whereas U was often present in a consecutive manner (Figure S3E, F in Supporting Information). Therefore, C and G residues are not separated to different lengths in the subsequent analysis. As the level of non-A residues at the 5′-ends is relatively low, we combined these with those internal ones and call them together as internal ones in the subsequent analysis. Examination of the length of poly(A) tails bearing non-A residues revealed that tails with U residues are relatively short, while tails with C and G residues are slightly depleted in the short fraction (Figure S3G, H in Supporting Information).

We also performed PAIso-seq2 analysis of mouse germinal vesicle (GV) oocytes, the further development of which relies on poly(A) tail-mediated post-transcriptional regulation, and compared the data to the previous PAIso-seq datasets (Liu et al., 2019). A very similar pattern was found among the poly(A) tails longer than 25 nt, whereas PAIso-seq2 captured a much greater number of tails <20 nt compared to PAIso-seq (Figure 1H). Besides the poly(A) tail length, the level of internal non-A residues quantified by these two methods was largely comparable with 3′-end non-A residues quantified by PAIso-seq2 but not PAIso-seq (Figure 1I).

The difference in poly(A) tails between cytoplasmic and nuclear mRNA is largely unknown. The nuclear-cytoplasmic RNA separation was successful as confirmed by the distribution of known nuclear or cytoplasmic enriched RNA (Figure S4 in Supporting Information). Interestingly, the nuclear transcripts from both 3T3 and ES cells had obviously longer poly(A) tails than cytoplasmic ones (Figure 1J, K). Moreover, the non-A residues showed cell type-specific global differences between the nuclear and cytoplasmic fractions (Figure 1L, M).

In summary, PAIso-seq2 reveals that non-A residues can occur at the internal and 3′-end parts of poly(A) tails (Figure S5A in Supporting Information), and poly(A) tails are differentially regulated in terms of both length and non-A residues in the nucleus and the cytoplasm (Figure S5B in Supporting Information), suggesting important regulation and function of poly(A) tails in these compartments. As PAIso-seq2 can capture both polyadenylated and non-polyadenylated transcripts, it will likely be applicable to explore the potential post-transcriptional regulation of 3′-ends of the large catalog of non-coding RNAs (Saw et al., 2021). PAIso-seq2 is the first method for comprehensively analyzing full-length transcriptome together with complete poly(A) tail sequences, which provides rich information of the transcriptome, including gene expression level, alternative transcription starting sites, alternative splicing, alternative polyadenylation sites, novel long non-coding RNAs, poly(A) tail length, as well as internal and 3′-end non-A residues within poly(A) tails. The relationship among the above information within a full-length poly(A) tail-inclusive cDNA and the underlying mechanisms are interesting themes that warrant further exploration in the future. Among all the current transcriptome-wide poly(A) analyzing methods, PAIso-seq2 is the most comprehensive, while PAIso-seq is the most sensitive (Figure S6 in Supporting Information), which will be powerful tools for exploring the poly(A) tail mediated post-transcriptional regulations.

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported by the National Key Research and Development Program of China (2018YFA0107001, 2020YFA0804000), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24020203), the CAS Project for Young Scientists in Basic Research (YSBR-012), the National Natural Science Foundation of China (31970588, 32170606), the Natural Science Foundation of Heilongjiang province (YQ2020C003), the China Postdoctoral Science Foundation (2020M670516, 2020T130687), and the State Key Laboratory of Molecular Developmental Biology. We thank Ying Liu for her technical assistance in cell culturing.*

References

Beverly, M., Hagen, C., and Slack, O. (2018). Poly A tail length analysis of *in vitro* transcribed mRNA by LC-MS. Anal Bioanal Chem 410, 1667–1677.

Chang, H., Lim, J., Ha, M., and Kim, V.N. (2014). TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. Mol Cell 53, 1044–1052.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev 11, 2755–2766.
Gu, W., Crawford, E.D., O'Donovan, B.D., Wilson, M.R., Chow, E.D., Retallack, H., and DeRisi, J.L. (2016). Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol 17, 41.
Hu, S.B., Xiang, J.F., Li, X., Xu, Y., Xue, W., Huang, M., Wong, C.C., Sagum, C.A., Bedford, M.T., Yang, L., et al. (2015). Protein arginine methyltransferase CARM1 attenuates the paraspeckle-mediated nuclear retention of mRNAs containing IR*Alu*s. Genes Dev 29, 630–645.
Legnini, I., Alles, J., Karaiskos, N., Ayoub, S., and Rajewsky, N. (2019). FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. Nat Methods 16, 879–886.
Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J., and Kim, V.N. (2014). Uridylation by TUT4 and TUT7 marks mRNA for degradation. Cell 159, 1365–1376.
Lim, J., Kim, D., Lee, Y.S., Ha, M., Lee, M., Yeo, J., Chang, H., Song, J., Ahn, K., and Kim, V.N. (2018). Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. Science 361, 701–704.
Liu, Y., Nie, H., Liu, H., and Lu, F. (2019). Poly(A) inclusive RNA isoform sequencing (PAIso-seq) reveals wide-spread non-adenosine residues within RNA poly(A) tails. Nat Commun 10, 5292.
Liu, Y., Zhang, Y., Wang, J., and Lu, F. (2022). Transcriptome-wide measurement of poly(A) tail length and composition at subnanogram total RNA sensitivity by PAIso-seq. Nat Protoc, doi: 10.1038/s41596-022-00704-8.
Long, Y., Jia, J., Mo, W., Jin, X., and Zhai, J. (2021). FLEP-seq: simultaneous detection of RNA polymerase II position, splicing status, polyadenylation site and poly(A) tail length at genome-wide scale by single-molecule nascent RNA sequencing. Nat Protoc 16, 4355–4381.
Saw, P.E., Xu, X., Chen, J., and Song, E.W. (2021). Non-coding RNAs: the new central dogma of cancer biology. Sci China Life Sci 64, 22–50.
Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature 508, 66–71.
Trepotec, Z., Geiger, J., Plank, C., Aneja, M.K., and Rudolph, C. (2019). Segmented poly(A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life. RNA 25, 507–518.
Yu, S., and Kim, V.N. (2020). A tale of non-canonical tails: gene regulation by post-transcriptional RNA tailing. Nat Rev Mol Cell Biol 21, 542–556.

SUPPORTING INFORMATION

The supporting information is available online at https://doi.org/10.1007/s11427-022-2186-8. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.