

Nano3P-seq: transcriptome-wide analysis of gene expression and tail dynamics using end-capture nanopore cDNA sequencing

Received: 15 October 2021

Accepted: 3 November 2022

Published online: 19 December 2022

 Check for updates

Oguzhan Begik^{1,2,3}, Gregor Diensthuber¹, Huanle Liu¹, Anna Delgado-Tejedor^{1,4}, Cassandra Kontur⁵, Adnan Muhammad Niazi⁶, Eivind Valen^{6,7}, Antonio J. Giraldez⁵, Jean-Denis Beaudoin⁸, John S. Mattick^{2,9} & Eva Maria Novoa^{1,4}✉

RNA polyadenylation plays a central role in RNA maturation, fate, and stability. In response to developmental cues, polyA tail lengths can vary, affecting the translation efficiency and stability of mRNAs. Here we develop Nanopore 3' end-capture sequencing (Nano3P-seq), a method that relies on nanopore cDNA sequencing to simultaneously quantify RNA abundance, tail composition, and tail length dynamics at per-read resolution. By employing a template-switching-based sequencing protocol, Nano3P-seq can sequence RNA molecule from its 3' end, regardless of its polyadenylation status, without the need for PCR amplification or ligation of RNA adapters. We demonstrate that Nano3P-seq provides quantitative estimates of RNA abundance and tail lengths, and captures a wide diversity of RNA biotypes. We find that, in addition to mRNA and long non-coding RNA, polyA tails can be identified in 16S mitochondrial ribosomal RNA in both mouse and zebrafish models. Moreover, we show that mRNA tail lengths are dynamically regulated during vertebrate embryogenesis at an isoform-specific level, correlating with mRNA decay. Finally, we demonstrate the ability of Nano3P-seq in capturing non-A bases within polyA tails of various lengths, and reveal their distribution during vertebrate embryogenesis. Overall, Nano3P-seq is a simple and robust method for accurately estimating transcript levels, tail lengths, and tail composition heterogeneity in individual reads, with minimal library preparation biases, both in the coding and non-coding transcriptome.

RNA molecules are subject to multiple co- and post-transcriptional modifications, shaping them to their final mature form¹. Polyadenylation of RNA is one such modification, which is known to affect the stability and translation efficiency of the RNA molecule^{2–4} and to play an essential role in determining the fate of RNA molecules in a wide range of biological processes^{5,6}.

One context in which polyadenylation has been shown to play a major role in determining RNA fate and decay is vertebrate embryogenesis⁶. Indeed, in the first hours after fertilization, vertebrate embryos undergo major cellular reprogramming, a process known as the maternal-to-zygotic transition (MZT)⁷. During the MZT, maternally inherited RNA and proteins are responsible for activation of the zygotic

genome and are later replaced by the zygotic program^{8,9}. Because the MZT begins in a transcriptionally silent embryo, this transition relies heavily on post-transcriptional regulatory mechanisms⁷, including modulation of the polyadenylation status of the RNA molecules^{6,10}. Therefore, characterizing the dynamics of RNA polyadenylation is key to understanding how these modifications regulate the fate and function of RNA molecules.

In the past few years, several transcriptome-wide methods have become available for studying the dynamics of polyadenylated tails (polyA tails) based on next-generation sequencing (NGS), such as PAL-seq or TAIL-seq^{10,11}. While these methods have been successfully employed to characterize the dynamics of polyA tail lengths in various contexts, they have several caveats: (i) they provide a limited perspective on isoform–tail relationships owing to the short-read-length nature of NGS-based technologies; (ii) they do not provide single-molecule resolution; (iii) they are severely affected by PCR amplification biases; and (iv) they can only measure tail lengths that are shorter than the read length.

To overcome these limitations, the direct RNA sequencing (dRNA-seq) platform offered by Oxford Nanopore Technologies (ONT) has been proposed as a means to study both the transcriptome and polyA tail lengths simultaneously^{12,13}. To sequence native RNAs using dRNA-seq, polyA-tailed RNA molecules are ligated to a 3' adapter that contains an oligo(dT) overhang. Consequently, dRNA-seq libraries capture the full-length polyA tail; however, ligation occurs only on RNA molecules that anneal to the oligo(dT) overhang, thus exclusively capturing polyA transcripts with tail lengths greater than 10 nucleotides. A variation of this method consisting of *in vitro* poly(I)-tailing the transcriptome before library preparation has been proposed for studying nascent RNAs using dRNA-seq^{14,15}, thus capturing both polyadenylated and non-polyadenylated mRNAs. However, major limitations to this variation include low sequencing yields compared with standard dRNA-seq (10–30%)¹⁴ and a lack of tools to distinguish poly(I) and poly(A) signals; therefore, polyA tail length information is lost in these datasets^{14,15}. An alternative approach for studying the transcriptome using nanopore technologies is direct cDNA sequencing (dcDNA-seq), but this approach cannot sequence the polyA⁺ transcriptome, nor can it capture polyA tail length information. Overall, both dRNA-seq and dcDNA-seq nanopore library preparation protocols are limited to the sequencing of polyadenylated transcripts and thus cannot provide a comprehensive view of both polyadenylated and deadenylated RNA molecules, in addition to being unable to capture RNA molecules with other types of RNA tails (for example, polyuridine).

Here we present a novel method that employs nanopore sequencing to simultaneously obtain per-isoform transcriptome abundance and tail lengths in full-length individual reads, with minimal library preparation steps, which we term Nanopore 3'-end-capture sequencing (Nano3P-seq). Notably, Nano3P-seq uses template switching to initiate reverse transcription and, therefore, does not require 3' end adapter ligation steps, PCR amplification, or second-strand cDNA synthesis. We demonstrate that Nano3P-seq can capture any type of RNA molecule regardless of its 3' sequence, including polyA-tailed and non-tailed RNA. Moreover, we show that Nano3P-seq can accurately quantify RNA abundances in both the coding and non-coding transcriptome, as well as quantify the polyA tail lengths and tail composition of individual RNA molecules in a highly reproducible manner.

Results

Nano3P-seq robustly captures both polyadenylated and non-polyadenylated RNA

Because nanopore sequencing is typically limited to the analysis of polyA⁺ RNA molecules (Fig. 1a), previous efforts have opted to perform *in vitro* polyadenylation reactions of the total RNA to capture non-polyadenylated RNA in the sequencing run¹⁶. While this option can capture any given transcript present in the sample, it also leads to

a loss of polyA tail length information. Therefore, we reasoned that by coupling template switching to cDNA nanopore sequencing, we would simultaneously capture the polyA⁺ and polyA⁻ transcriptome, while retaining polyA tail length information from each individual RNA molecule (Fig. 1a).

To assess the ability of Nano3P-seq to sequence both polyA⁺ and polyA⁻ RNA, we first sequenced two synthetic RNAs, one lacking a polyA tail and one that had been *in vitro* polyadenylated (Methods) (Extended Data Fig. 1a–c). Our results show that Nano3P-seq efficiently captures both polyadenylated and non-polyadenylated RNA molecules, as well as the diversity of polyA tail lengths in individual RNAs (Extended Data Fig. 1c). We then examined the performance of Nano3P-seq on *in vivo* samples, and sequenced total RNA samples from mouse brain, previously enriched in nuclear and mitochondrial content via subcellular fractionation to increase the content of non-coding RNAs¹⁷ (Methods). We confirmed that Nano3P-seq captured RNA biotypes that are typically polyadenylated (that is, mRNA, long intervening non-coding RNAs (lincRNA), and processed transcript) as well as non-polyadenylated (that is rRNA, miscellaneous RNA (miscRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA)), the majority of them being rRNA, mRNA, and snRNA (Fig. 1b and Extended Data Fig. 1d). In addition, our results confirmed that polyA tail length information was retained in individual reads. Specifically, the majority of reads corresponding to mRNAs had polyA tails (Fig. 1c and Extended Data Fig. 1e, f), whereas non-coding RNAs such as snoRNAs (Extended Data Fig. 1f) and snRNAs did not have polyA tails (Extended Data Fig. 1g), as expected.

To assess the accuracy and reproducibility of Nano3P-seq in quantifying RNA abundances, we examined the performance of Nano3P-seq in synthetic RNA mixes (sequins)¹⁸ that had been spiked into samples in independent flow cells. Our results showed that Nano3P-seq provided accurate estimates of RNA abundances both at the per-gene (Pearson's $R = 0.93$, slope = 0.93) (Fig. 1d) and per-isoform (Pearson's $R = 0.89$, slope = 0.92) (Extended Data Fig. 1h) level. These quantifications were highly reproducible across biological replicates, both at the per-gene (Pearson's $R = 0.98$) and per-isoform (Pearson's $R = 0.97$) (Extended Data Fig. 1i) level. We should note that previous works using Illumina RNA-seq reported a global correlation of 0.89 and 0.86 between observed and expected sequin counts at per-gene and per-transcript level, respectively¹⁸.

Nano3P-seq recapitulates transcriptomic switch that occurs during MZT

Next, we employed Nano3P-seq to examine the RNA dynamics that occur during the maternal-to-zygotic transition (MZT) at single-molecule resolution (Fig. 2a). To this end, we isolated total RNA from zebrafish embryos at 2, 4, and 6 hours post-fertilization (h.p.f.) in biological duplicates, ribodepleted the samples, and sequenced them using the Nano3P-seq protocol (Extended Data Fig. 2a, b; Methods). Quantification of the mRNA abundances in three independent biological replicates showed that per-gene abundances (reads per million (RPM)) obtained using Nano3P-seq were highly reproducible (Fig. 2b and Extended Data Fig. 2c, d).

A comparative analysis of mRNA population dynamics across time points showed that Nano3P-seq recapitulated the transcriptomic switch that occurs during the MZT^{7,8}, with a decay of mRNA genes previously reported to have a 'maternal decay mode' (Fig. 2c), in agreement with previous studies^{19,20}. Notably, in addition to polyadenylated RNAs, Nano3P-seq captured a wide variety of RNA biotypes without polyA tails that are also present in early embryo stages. We observed a significant increase in the abundance of lincRNAs and snoRNAs during the MZT, as well as a sharp increase in miRNA expression at 4 h.p.f., followed by a decrease at 6 h.p.f. (Fig. 2d). Analysis of individual miRNA gene expression patterns during the MZT revealed that the sharp increase observed at 4 h.p.f. was primarily caused by increased expression of genes belonging to the miR-430 family, which is known to be essential

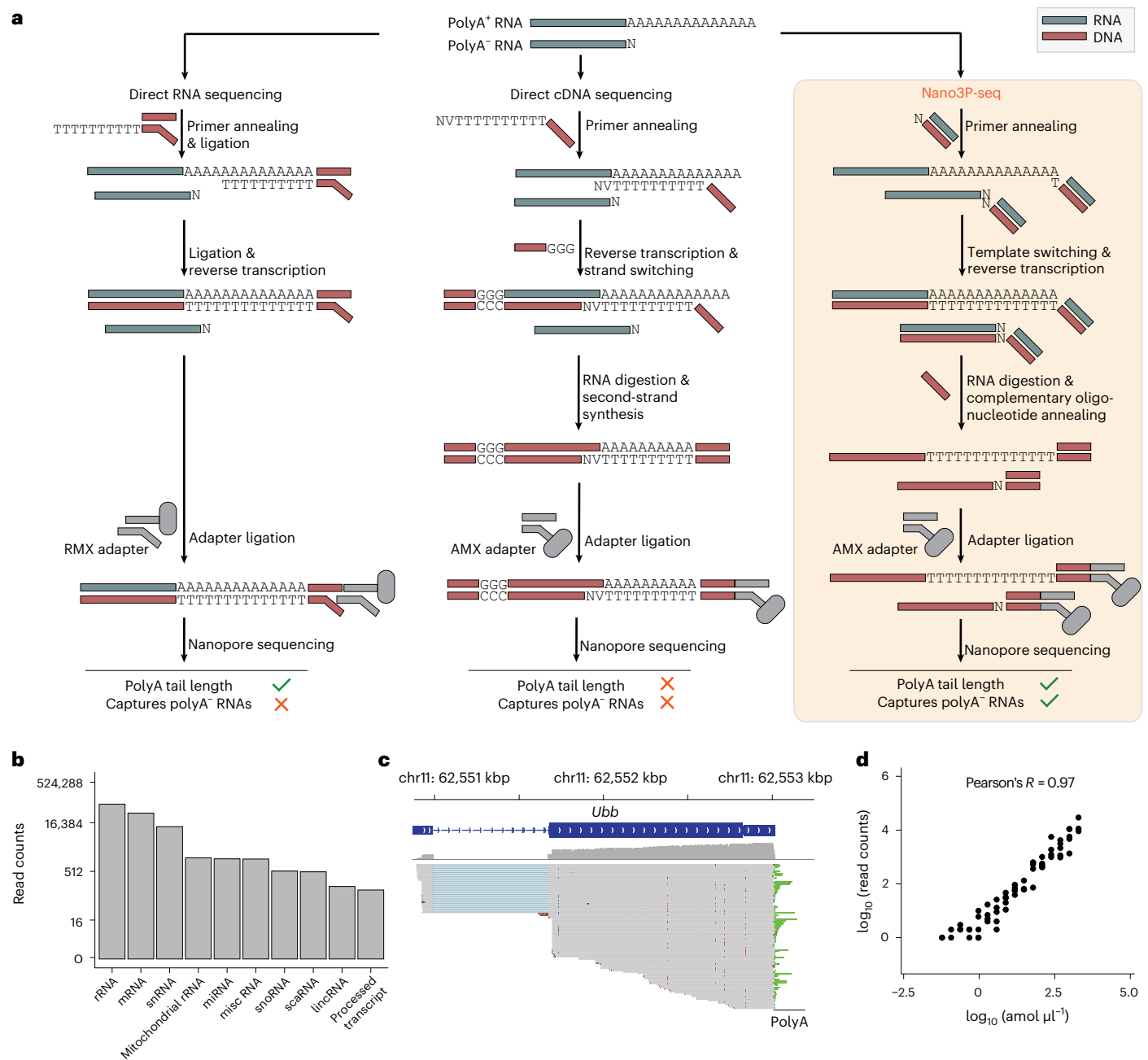


Fig. 1 | Nano3P-seq captures polyadenylated and non-polyadenylated RNAs, while retaining polyA tail length information. a, Schematic overview comparing three different library preparation methods for studying the transcriptome using nanopore sequencing. RMX, RNA adapter mix (provided with the SQK-RNA002 dRNA-seq library preparation kit); AMX, adapter mix (provided with the SQK-DCS109 dcDNA-seq library preparation kit). **b**, Nano3P-seq captures a wide range of RNA biotypes in a mouse brain nuclear/

mitochondrial RNA sample. **c**, Integrative Genome Viewer (IGV) snapshot of reads generated with Nano3P-seq, mapped to the *Ubb* gene, illustrating the diversity of polyA tail lengths captured across different reads. The polyA tail region is shown in green. kbp, kilobase pairs; scaRNA, small Cajal body-specific RNA. **d**, Scatter plot of log transformed concentrations ($\text{amol } \mu\text{l}^{-1}$) and read counts of sequin genes (Pearson's $R = 0.93$, slope = 0.93). Each dot represents a sequin. See also Extended Data Fig. 1h,i.

in mediating the decay of a group of maternal mRNAs during the MZT (mirR-430-dependent decay)²¹ (Extended Data Fig. 2e). By contrast, much fewer non-coding RNA populations were globally captured (relative to coding RNA populations) when dRNA-seq was applied to the same samples (Fig. 2e).

We noted, however, that mitochondrial rRNAs were not enriched in Nano3P-seq datasets relative to dRNA-seq datasets (Fig. 2f). Indeed, analysis of zebrafish mitochondrial rRNA reads revealed that many 16S mitochondrial rRNA reads contained a polyA tail, which explains the lack of enrichment of mitochondrial rRNAs in Nano3P-seq datasets

relative to dRNA-seq datasets (Extended Data Fig. 2f). In agreement with this observation, we found that polyA-tailed 16S mitochondrial rRNAs were present not only in zebrafish (Extended Data Fig. 2g), but also in mouse (Extended Data Fig. 2h,i), suggesting that this feature is conserved across species and not a sequencing artifact, in agreement with previous reports^{22,23}. The presence of polyA tails in mouse 16S mitochondrial rRNAs was further validated using an orthogonal method (polyA tail length assay coupled to Sanger sequencing; Methods), which confirmed that 16S mitochondrial rRNA is polyadenylated (Extended Data Fig. 2j).

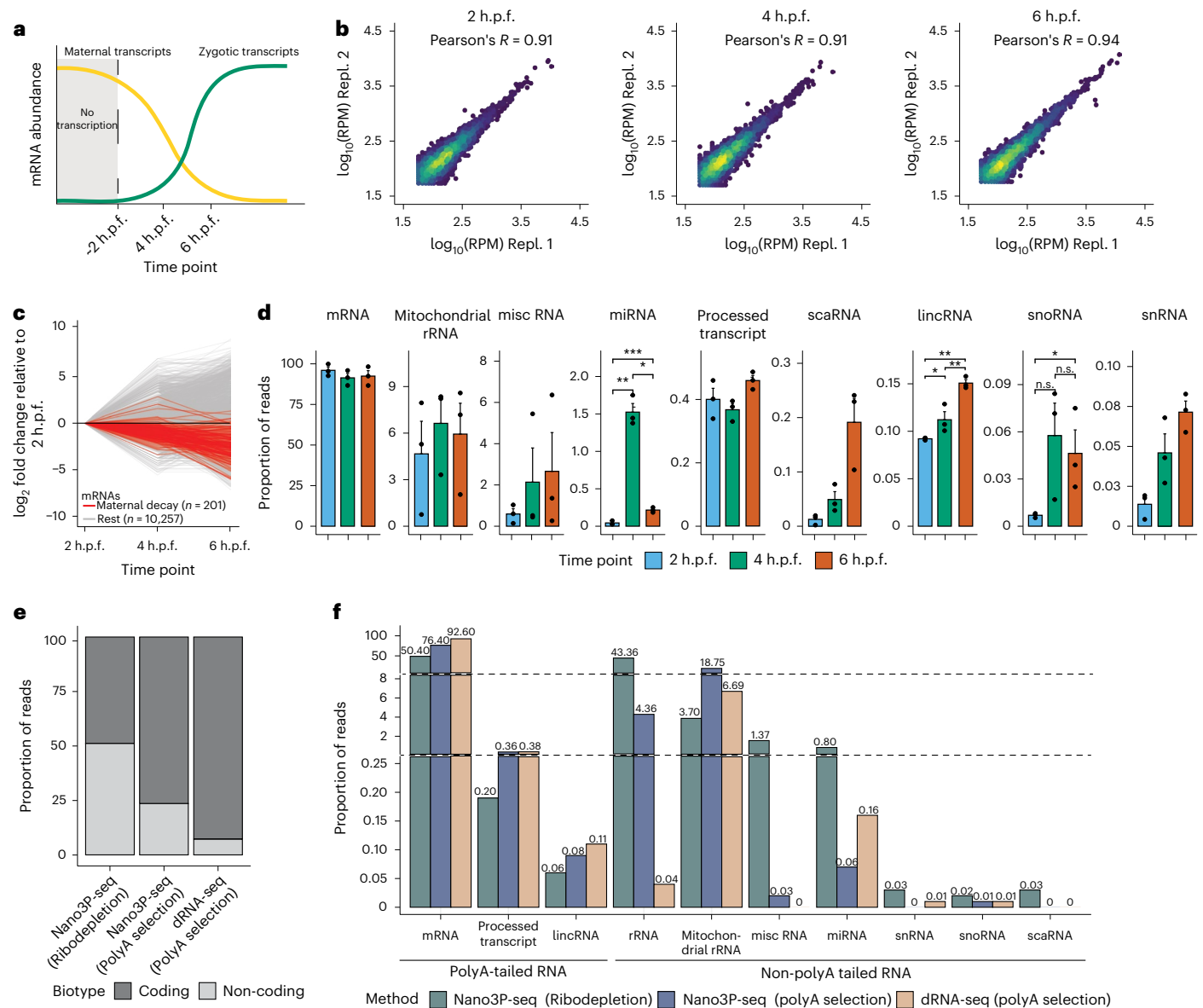


Fig. 2 | Nano3P-seq captures a wide diversity of coding and non-coding RNAs and their expression dynamics during the maternal-to-zygotic transition (MZT).

a, Schematic overview of the transcriptional change that occurs during the MZT in zebrafish. **b**, Scatter plots depicting the correlation of mRNA log transformed RPM between biological replicates at three different time points during the MZT. **c**, Changes in mRNA abundance during the MZT ($t = 2, 4,$ and 6 h.p.f.), relative to 2 h.p.f. Genes previously reported to have a 'maternal decay mode' are depicted in red. **d**, Bar plots depicting the abundance of different RNA biotypes captured by Nano3P-seq during the MZT (2, 4, and 6 h.p.f., shown

in blue, green, and red, respectively). Statistical analyses were performed using the Kruskal–Wallis test. $n = 3$ biological replicates, data are presented as mean \pm s.e.m. $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, $****P \leq 0.0001$. n.s., not significant ($P > 0.05$). **e**, Relative proportion of coding and non-coding RNAs captured using dRNA-seq (on polyA-selected samples), Nano3P-seq (on polyA-selected samples), and Nano3P-seq (on ribodepleted samples). **f**, Percentage of reads mapping to distinct biotypes captured using Nano3P-seq (on ribodepleted samples) (green), Nano3P-seq (on polyA-selected samples) (blue), and dRNA-seq (on polyA-selected samples) (light brown).

PolyA tail lengths can be accurately estimated using Nano3P-seq

We then examined whether Nano3P-seq accurately estimated polyA tail lengths. We note that algorithms for detecting polyA tails in native RNA nanopore sequencing reads are well established and benchmarked^{12,24–26}, but their applicability to cDNA reads, such as those from Nano3P-seq, remains unclear. To this end, we first examined whether the tailfindR polyA tail prediction software²⁵ would capture the presence or absence of polyA tails on synthetic RNAs that were either polyadenylated or non-polyadenylated and had been sequenced using Nano3P-seq. We found that tailfindR can identify both polyadenylated and non-polyadenylated Nano3P-seq reads (Fig. 3a and Extended Figure 3a). Then, we assessed the accuracy of polyA tail length

predictions of tailfindR in Nano3P-seq datasets that included a battery of synthetic RNAs (sequins)¹⁸ or synthetic cDNA sequences with known polyA tail lengths (Fig. 3b). Our results showed that polyA tail length estimations of sequins in Nano3P-seq data were highly reproducible across replicates ($R = 0.993$; Extended Data Fig. 3b), with an accuracy similar to that observed when performing polyA tail length estimations in sequins that had been sequenced using dRNA-seq (Fig. 3c and Extended Data Fig. 3c). Moreover, the variance of tail length estimates across reads belonging to the same transcript was smaller in Nano3P-seq datasets than in dRNA-seq datasets (Extended Data Fig. 3d,e). Similar results were obtained when we assessed the performance of the algorithm in a set of synthetic cDNA sequences that

spanned a broader range of polyA tail lengths (0, 15, 30, 60, 90, and 120 nucleotides) (Fig. 3b,c and Extended Data Fig. 3f).

Finally, we performed a comparative analysis of mRNA polyA tail lengths from human, mouse, yeast, and zebrafish. We observed that mouse brain mRNAs, had the longest mRNA tails among the four species, with a median polyA tail length of 90 nucleotides, whereas the shortest polyA tail lengths were observed in yeast, with a median polyA tail length of 23.5 nucleotides (Fig. 3d), in agreement with previous studies¹⁰.

Charting polyA tail length dynamics in vivo using Nano3P-seq

We explored whether Nano3P-seq could be used to investigate polyA tail length dynamics in vivo. To this end, we first examined the ability of Nano3P-seq to identify which RNA biotypes were polyadenylated in mouse brain total RNA samples that had been previously enriched in nuclear/mitochondrial content to increase the proportion of non-coding RNAs. We found that polyA tails were mainly predicted on mRNAs, but also in lincRNAs and processed transcripts, which are also known to be polyadenylated^{27,28} (Fig. 3e).

We next analyzed the polyA tail length dynamics across developmental stages of zebrafish mRNAs during the MZT ($t = 2, 4,$ and 6 h.p.f.). PolyA tail length estimates were highly reproducible across independent biological replicates sequenced in independent flow cells for all three time points studied ($R = 0.85–0.95$) (Fig. 3f). We observed an overall increase in the mean mRNA polyA tail length during the MZT (Fig. 3g and Extended Data Fig. 3g), in agreement with previous reports¹⁰. All mRNAs examined were found to be polyadenylated, with the exception of histone mRNAs, which had a median polyA tail length of zero (Extended Data Fig. 3h), in agreement with previous works reporting their non-polyadenylated status²⁹. These findings show that Nano3P-seq can capture RNA molecules with structured 3' ends, such as those found in histones³⁰. Finally, we note that per-gene polyA tail length estimates obtained by Nano3P-seq during the MZT showed a good correlation with those obtained using orthogonal methods such as PAL-seq¹⁰ (Pearson's $R = 0.71–0.85$) (Extended Data Fig. 3i,j).

Next, we examined the correlation between polyA tail length dynamics and mRNA decay. To this end, mRNA transcripts were binned depending on their decay mode (maternal decay, zygotic activation-dependent decay, miR-430-dependent decay, and no decay), as previously described²¹. We observed that the three groups of mRNAs that are known to decay (maternal, zygotic, and miR-430) showed a significant decrease in mRNA abundance (Fig. 3h), as expected. However, the patterns of polyA tail length dynamics strongly varied depending on the decay mode of the transcript (Fig. 3i). Specifically, we observed that transcripts that decayed in an miR-430-dependent manner showed a significant decrease in polyA tail length during the MZT, in agreement with previous studies^{19,21}. By contrast, for mRNAs with the zygotic genome activation-dependent decay mode, this shortening only occurred after 4 h.p.f., and maternal mRNAs did not present a decrease in polyA tail length, but instead showed a consistent increase in tail length throughout the MZT. These observations are consistent

with the reanalysis of PAL-seq data (Extended Data Fig. 3k,l). Overall, these results show that not all decay modes are associated with a reduction in transcript polyA tail lengths and demonstrate the applicability of Nano3P-seq to identify polyadenylated RNA populations, study their RNA abundance, and estimate their polyA tail length dynamics, at both the global level and the level of individual transcripts, and thus provide mechanistic insights into different gene regulatory programs.

Nano3P-seq captures isoform-specific polyA tail length changes during the MZT

A major feature that distinguishes nanopore sequencing from NGS is its ability to produce long reads, allowing RNA polyadenylation dynamics to be studied at the isoform level. Therefore, we wondered whether Nano3P-seq could identify differentially polyadenylated transcript isoforms during the MZT.

To this end, we first examined whether polyA tail lengths significantly diverged across time points at the per-isoform level. We note that only reads mapping to genes encoding for at least two annotated isoforms, and with mapping coverage greater than ten reads per isoform, were maintained for further analyses. Our analyses revealed that 55.3% ($\pm 8.62\%$) of analyzed transcripts varied significantly in polyA tail length ($P < 0.05$) during the MZT. Notably, we observed that analyses at the per-gene level often revealed a different picture compared with analyses at per-isoform level. For example, in *khdrbs1a* and *syncrip*, per-isoform analysis revealed opposite tail length dynamics among isoforms during the MZT, with one isoform decreasing and another isoform increasing in polyA tail length as the MZT progressed (Fig. 4a,b and Extended Data Fig. 4a,b).

Next, we compared isoform-specific polyA tail lengths across isoforms encoded by the same gene and found that 17.3% ($\pm 6.7\%$) of analyzed genes presented significant differences in their polyA tail lengths across isoforms (Fig. 4c and Extended Data Fig. 4c). Altogether, these results show that polyA tail length dynamics are not only dependent on the gene and embryogenesis stage, but are also specific to individual transcript isoforms. Moreover, these findings demonstrate that Nano3P-seq can provide transcriptome-wide measurements of the polyadenylation status of diverse biological samples with both single-read and single-isoform resolution.

Detection of isoform-specific RNA modifications using Nano3P-seq

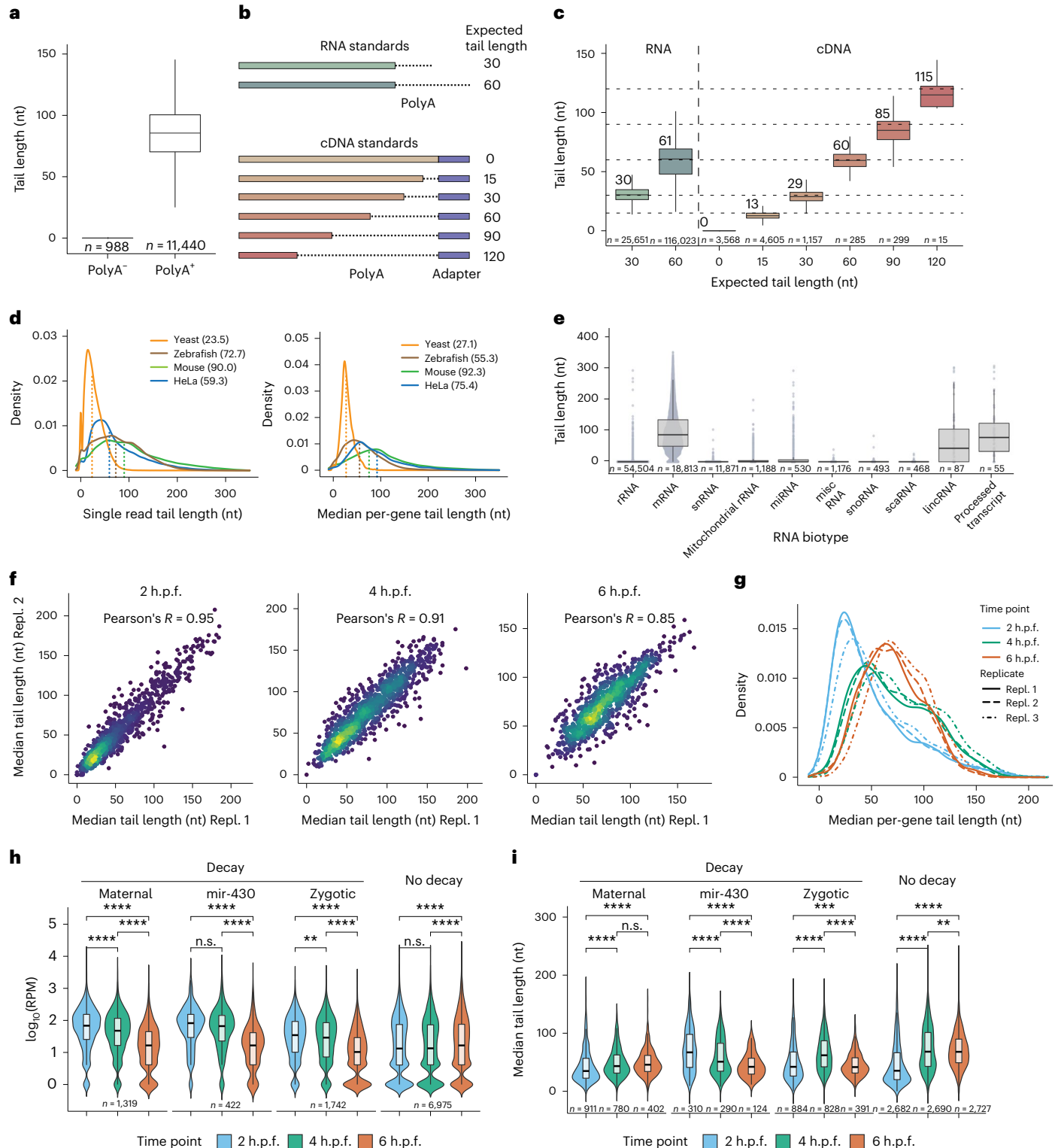
RNA molecules are decorated with chemical modifications, which are essential for the stability, maturation, fate, and function of the RNA^{31–34}. Some modifications occur in base positions that are involved in Watson–Crick (WC) base pairing, causing a disruption during reverse transcription. Consequently, these modifications can be seen as increased 'errors' and drop-off rates in RNA-seq datasets^{35–38}. One example is the hypermodified base m¹acp³Ψ, which is present in the eukaryotic small subunit (SSU) rRNA³⁹ crucial for the final processing steps of precursor rRNA (pre-rRNA) into mature SSU rRNA^{40,41}.

Fig. 3 | Nano3P-seq can be used to accurately estimate polyA tail lengths in individual molecules. **a**, PolyA tail length estimates of non-polyadenylated (curlcake 1) and polyadenylated (curlcake 2) synthetic RNAs sequenced with Nano3P-seq. See also Extended Data Fig. 1a–c. nt, nucleotides. **b**, Schematic overview of the standards used to assess the tail length estimation accuracy of Nano3P-seq. **c**, Box plots depicting tail length estimations of RNA and cDNA standards sequenced with Nano3P-seq. Values on box plots indicate the median polyA tail length estimation for each standard. **d**, PolyA tail length distribution of yeast, zebrafish, and mouse mRNAs represented as single-transcript values (left) and per-gene medians (right). **e**, PolyA tail length estimates across different RNA biotypes from mouse brain total RNA enriched in nuclear/mitochondrial RNA. Each dot represents a read. **f**, Replicability of median per-gene polyA tail length estimations of zebrafish embryonic mRNAs between two biological replicates for three different time points (2, 4, and 6 h.p.f.). **g**, Median per-gene

polyA tail length distribution of zebrafish embryonic mRNAs across zebrafish developmental stages (2, 4, and 6 hpf, shown in blue, green, and red, respectively) in three biological replicates (shown as full lines, dashed lines, and dotted/dashed lines, respectively). **h**, Comparative analysis of mRNA abundances (shown as \log_{10} (RPM) counts) of zebrafish mRNAs binned according to their annotated decay mode (maternal decay, zygotic activation-dependent decay, miR-430-dependent decay, and no decay) during early embryogenesis ($t = 2, 4,$ and 6 h.p.f.). **i**, Median per-gene polyA tail length estimations of zebrafish mRNAs binned according to their decay mode (maternal, miR-430, zygotic, and no decay) at 2, 4, and 6 h.p.f. For Fig. 3h,i; statistical analyses were performed using the Kruskal–Wallis test. **c,e,h,i**, The number of observations included in the analysis is shown below each box and violin plot. Box plot limits are defined by lower (bottom) and upper (top) quartiles. The bar indicates the median, and whiskers indicate $\pm 1.5 \times$ interquartile range.

Therefore, we examined whether Nano3P-seq could capture RNA modification differences across precursor and mature rRNA molecules from distinct maturation stages. To this end, reads mapping to SSU rRNAs were assigned to either ‘precursor’ or ‘processed’ isoforms (on the basis of the location of the 3’ end of the read) and analyzed the sequencing error patterns in the two populations (Fig. 4d). We observed that the mismatch frequency (mismatches from the reverse transcriptase) at the m¹acp³Ψ-modified site was very high in mature rRNAs but not present in pre-rRNAs, suggesting that this

modification is only present in mature rRNA populations. The presence of the hypermodification was also accompanied by a drop-off in sequencing coverage at the m¹acp³Ψ-modified site, in agreement with its ability to disrupt the Watson–Crick base pairing. Analysis of the ‘error’ signatures along the SSU transcripts showed that this position was the only one to change between precursor and processed rRNA molecules (Fig. 4d). These results were also orthogonally confirmed in dRNA-seq datasets, although the difference between pre-rRNA and mature rRNA error patterns were less evident than in



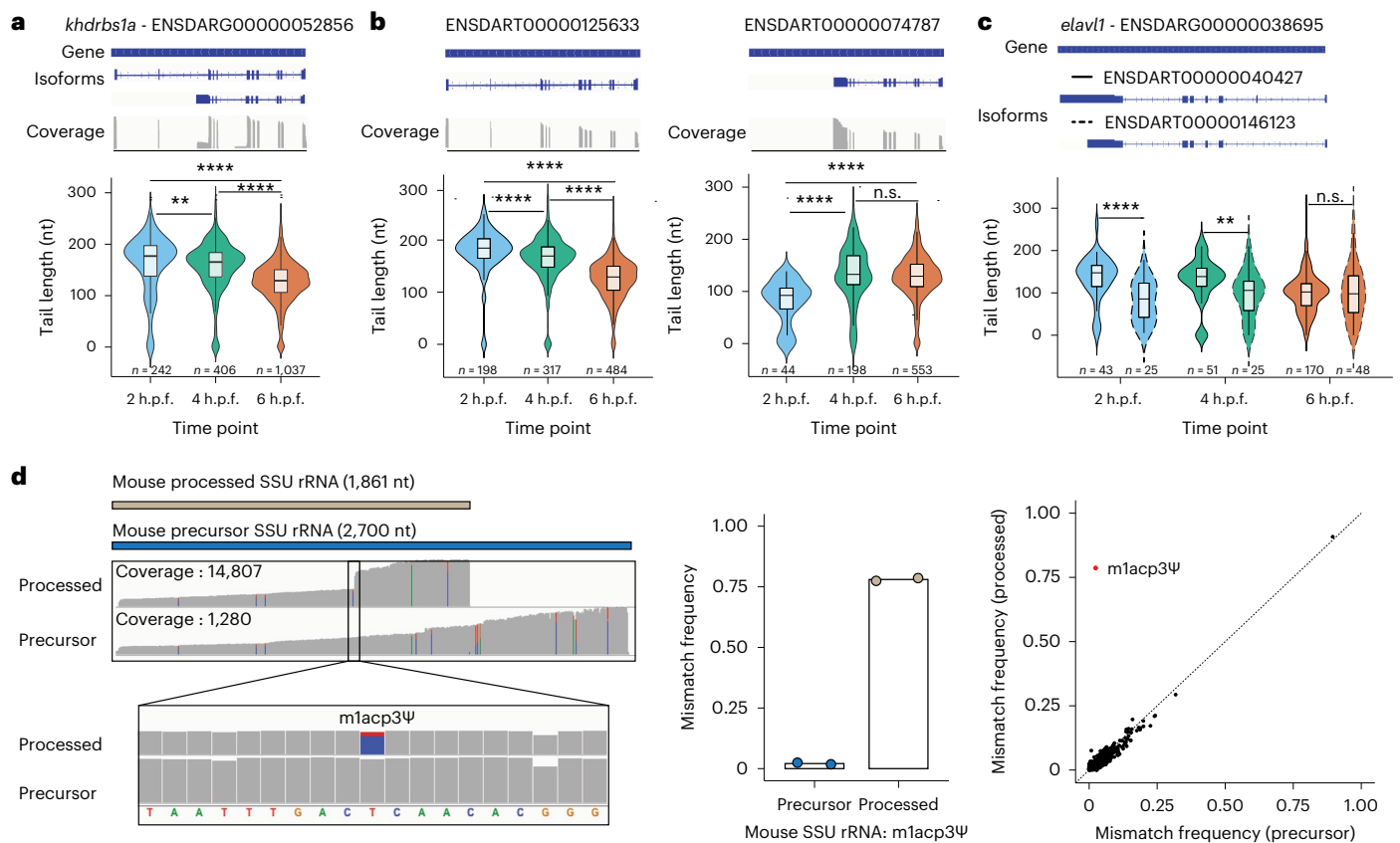


Fig. 4 | Isoform-specific polyA tail and modification dynamics can be captured using Nano3P-seq. **a**, Comparison of polyA tail length distributions of reads mapping to *khdrbs1a*, illustrated at the per-gene level, measured at three time points during the zebrafish MZT. Annotations of the gene and two main isoforms are shown at the top of the panels, along with an IGV coverage track of the reads mapping to the gene. **b, c**, Comparison of polyA tail length distributions of reads mapping to two distinct isoforms (full and dashed outline) of *elavl1* measured at three time points during the zebrafish MZT. Annotations of the gene and two main isoforms are shown at the top of the panels. **a–c**, Only isoforms with more than ten reads are shown. The number of reads included in the analysis is shown below each violin plot. *P* values have been computed using the Kruskal–Wallis test and corrected for multiple testing using the Benjamini–Hochberg method. Box plot limits are defined by lower (bottom) and upper (top) quartiles.

The bar indicates the median, and whiskers indicate $\pm 1.5 \times$ interquartile range. **d**, IGV coverage tracks of reads mapping to mouse processed small subunit rRNA (top track) and precursor SSU rRNA (bottom track), including a magnified image at the position known to be modified with m¹acp³Ψ (left). Reads mapping to SSU rRNAs were assigned to either ‘precursor’ or ‘processed’ isoforms on the basis of the overlap between 3′ end of the read and annotated end of the isoforms. Only reads with 3′ ends within ± 10 nucleotides of the annotated end of an isoform were kept. Positions with a mismatch frequency lower than 0.1 are shown in gray. Middle, the mismatch frequency values in mouse precursor and processed SSU rRNA at the position known to be modified with m¹acp³Ψ ($n = 2$ biological replicates) are shown. Right, the per-site mismatch frequencies observed in reads mapping to mouse precursor SSU rRNA and mouse processed SSU rRNA are compared, showing that the only outlier is m¹acp³Ψ.

Nano3P-seq datasets, likely owing to the presence of m¹Ψ modification (precursor of m¹acp³Ψ) at this site in the pre-rRNAs, which also causes increased ‘errors’ in dRNA-seq datasets, but not in Nano3P-seq datasets (Extended Data Fig. 4d). Finally, we noted that the same pattern of m¹acp³Ψ modification was observed in yeast SSU rRNA (Extended Data Fig. 4e), in agreement with previous studies⁴¹. This analysis could not be performed in the zebrafish Nano3P-seq datasets owing to the absence of de novo transcription of zygotic rRNAs in early embryo stages⁴². Altogether, our results demonstrate that Nano3P-seq can identify isoform-specific and/or maturation-dependent RNA modification in the form of altered mismatch frequencies and/or reverse transcription drop-offs.

Nano3P-seq identifies variations in tail composition during the MZT

Recent works using TAIL-seq have reported that a number of terminal modifications in polyA tails, such as polyuridine stretches, play a role in mRNA decay^{11,20,43,44}. However, these methods cannot be used to detect tail modifications among the vast majority of tail nucleotides, as Illumina sequencing quality strongly deteriorates in homopolymeric stretches and with increased read length. By contrast,

Pacific BioSciences (PacBio)-based approaches such as FLAM-seq can sequence through the entire tail, however, they cannot unambiguously identify 3′ terminal modifications⁴⁵.

Therefore, we explored whether Nano3P-seq could accurately identify nucleotide composition variations (either internal or terminal) within polyA tails. To this end, we designed synthetic cDNA molecules with polyA tails ending with A (30A), U (1U), UUU (3U), UUUUU (5U), CCCCC (5C), and GGGGG (5G), as well as a polyA tail that contained several internal G nucleotides at fixed positions (IntG) (Fig. 5a; Methods). Synthetic molecules were sequenced using Nano3P-seq, and for each read, the ‘tail’ was defined as the set of nucleotides found between the nanopore adapter and the last nucleotide mapped to the cDNA standard reference (Fig. 5b and Extended Data Fig. 5a). Nucleotide composition analysis of the last 20 nucleotides of each synthetic tail revealed that Nano3P-seq accurately estimated the non-A base content in the tails (Extended Data Fig. 5b) and accurately identified the position in which these non-A bases were found (Fig. 5c).

Next, we analyzed the mRNA tail composition of zebrafish embryos from different developmental stages (2, 4, and 6 h.p.f.). Analysis of tail base abundance revealed that G was the most common non-A base in zebrafish mRNAs and that there was a significant decrease in non-A

bases with progression of the MZT (Extended Data Figure 6a,b). As a control, we examined the composition of bases in synthetic molecules ('sequins'), which are expected to contain an homogeneous polyA tail, and found that the frequency of non-A bases in biological sequences was significantly higher (6–15-fold) than in synthetic 'sequins' molecules (Extended Data Figure 6a), suggesting that the majority of non-A bases observed in Nano3P-seq biological datasets are not base-calling artifacts.

We then binned the tails on the basis of the composition of the last 30 nucleotides (All-A, contains only A; Int-G, contains internal Gs; Int-U, contains internal Us; Int-C, contains internal Cs; Term-G, contains terminal Gs; Term-U, contains terminal Us; Term-C, contains terminal Cs) (Fig. 5d and Extended Data Figure 6c). Our analysis showed that non-A internal bases were more abundant than non-A terminal bases, with internal Gs (Int-G) being the most abundant type of non-A bases in polyA tails (Fig. 5d). Moreover, we found that the abundance of transcripts with non-A base tails (internal or terminal) typically decreased with progression of the MZT, except for transcripts with terminal Us, for which the observed decrease was not statistically significant.

Finally, we examined the relationship between tail length and the presence of non-A bases in the tails. Although the median tail length increases during the MZT (Figs 5e and 3g), we found that this increase did not occur for all transcripts, as it depended on tail composition. Specifically, terminally uridylated tails (Term-U) were short (median lengths of 43, 50, and 32 nucleotides at 2, 4, 6 h.p.f., respectively) regardless of the time point examined (Fig. 4e,f, see also Table S1). An equivalent analysis in mouse Nano3P-seq datasets revealed that Term-U tails were also significantly shorter than the other tail types examined (Extended Data Figure 6d and Supplementary Table 1); by contrast, this trend was not observed in yeast Nano3P-seq datasets (Extended Data Figure 6e, and Supplementary Table 1). Our results are in agreement with previous studies in mouse and human cell lines showing that G and U bases are common in polyA tails and that tails ending with U bases are more frequent in shorter polyA tails^{11,46}.

Assessment of tail length bias caused by polyA selection

Because the vast majority of cellular RNA is composed of rRNA, transcriptomic studies typically remove a significant portion of rRNA molecules to sequence a wider diversity of RNA biotypes. This removal can be achieved by (i) ribodepletion of the sample using biotinylated oligonucleotides that are complementary to rRNAs or (ii) selective enrichment of polyA⁺ transcripts using oligo(dT) beads. Although these two approaches are often used interchangeably, their effects on the transcriptome composition are not equal. Nano3P-seq allows us to compare the effects of these two approaches on both the transcriptome composition and polyA tail length distribution. In terms of its effects on transcriptome composition, we found that ribodepletion captures a larger variety of RNA biotypes compared with polyA selection, including several non-polyA-tailed RNA biotypes, as expected (Fig. 2f). However, we did not observe a significant difference in the distribution of mRNA polyA tail lengths between the two methods (Extended Data Figure 7a–c), suggesting that, at least in zebrafish MZT transcriptomes, oligo(dT) enrichments do not significantly bias the polyA⁺ mRNA populations by preferentially enriching for those with longer polyA tails.

Comparison of Nano3P-seq to orthogonal methods

We then performed a comparative analysis of zebrafish polyA tail lengths, read lengths, and per-read quality in libraries sequenced using either Nano3P-seq or dRNA-seq. We found that Nano3P-seq captured RNA molecules regardless of their tail ends, resulting in the capture of diverse RNA biotypes (Fig. 1b and Extended Data Fig. 1c–g) including deadenylated mRNA molecules (Extended Data Figs 3h and 7) and molecules terminating with non-A bases in their 3' ends (Fig. 5 and Extended Data Figs 5 and 6). By contrast, dRNA-seq only captured longer polyadenylated transcripts, as this method relies on the presence of polyA tail lengths greater than 10 nucleotides. Indeed, when comparing the distribution of per-read polyA tail length estimations of mRNAs, we observed that Nano3P-seq captured some mRNAs with predicted tail lengths of zero, whereas dRNA-seq only captured reads with longer tails (Extended Data Figure 7d,e). Moreover, we found that Nano3P-seq produced reads with significantly longer lengths than dRNA-seq (Extended Data Figure 7f), whereas the per-read qualities and 3' end biases were relatively similar between the two methods (Extended Data Figure 7g–i), in agreement with previous works¹². In terms of sequencing output, the yields of Nano3P-seq runs were similar—or slightly better—to those observed in dRNA-seq runs, producing ~100,000–200,000 reads in Flongle devices and ~500,000–2,000,000 in MinION devices, depending on the RNA input type and quality of the flowcell (Table S2).

In addition to comparing Nano3P-seq to dRNA-seq in matched zebrafish samples, we also compared Nano3P-seq polyA tail length estimates to publicly available datasets generated using orthogonal methods, including PAL-seq¹⁰, TAIL-seq¹¹, PAT-seq⁴⁷, and FLAM-seq⁴⁵ (Extended Data Figs 3i and 8). The highest correlation was observed between Nano3P-seq and PAL-seq on zebrafish samples (Pearson's $R = 0.71$ – 0.85 ; Extended Data Fig. 3i and Supplementary Table 3). Correlations were more modest when comparing Nano3P-seq to FLAM-seq (Pearson's $R = 0.47$) and TAIL-seq (Pearson's $R = 0.19$) on HeLa cell lines (Extended Data Fig. 8a and Supplementary Table 3), as well as when comparing Nano3P-seq to PAT-seq and PAL-seq on *Saccharomyces cerevisiae* samples (Pearson's $R = 0.43$; Extended Data Fig. 8b and Supplementary Table 3) most likely due to the intrinsic differences between the samples, given that the correlations were also modest when comparing the estimates of polyA tail lengths across orthogonal methods ($R = 0.1$ – 0.42 ; Supplementary Table 3).

Discussion

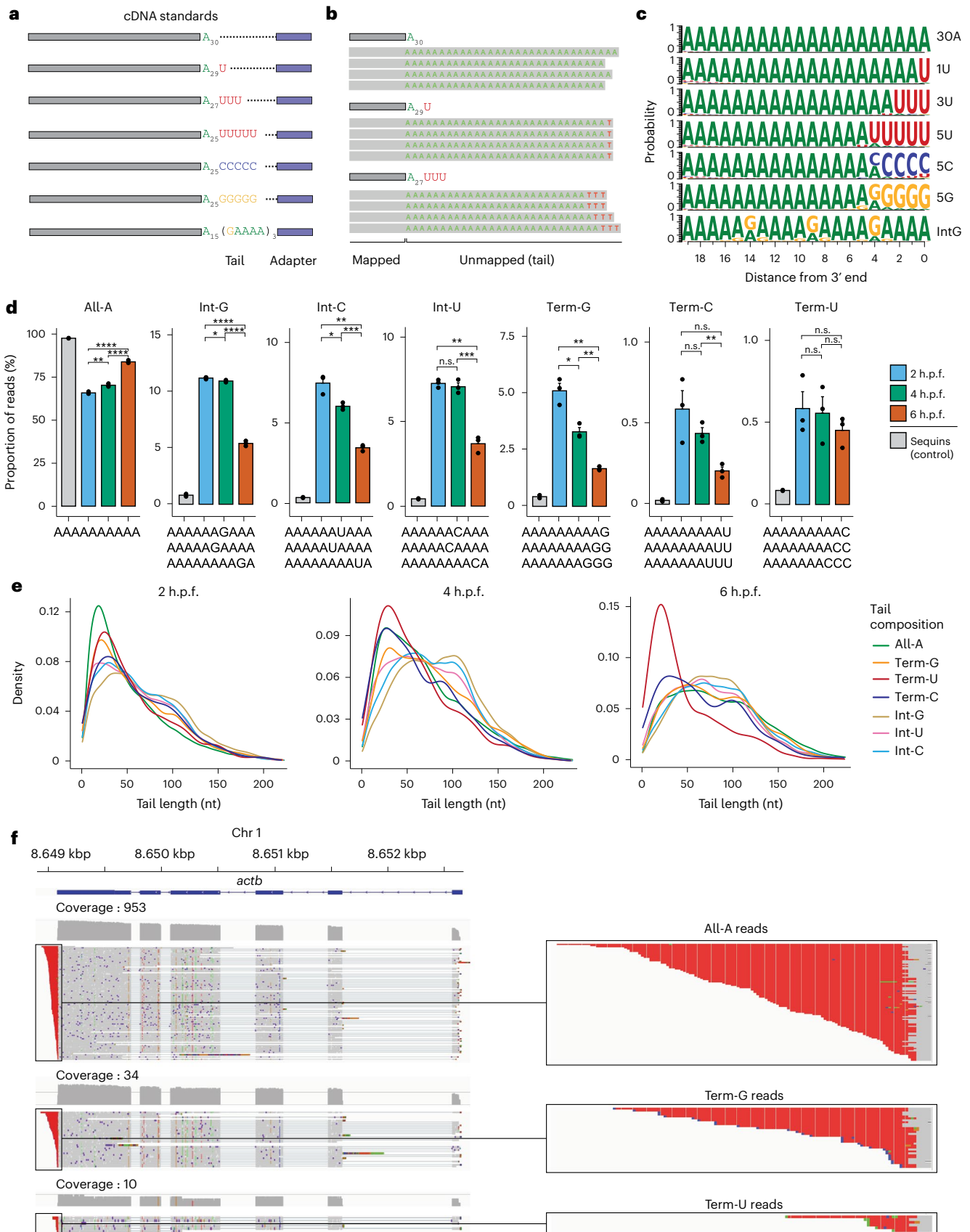
In the past few years, a variety of NGS-based high-throughput methods have been developed to characterize the 3' ends of RNA molecules at a transcriptome-wide scale, including methods to reveal polyA tail sites (for example, 3P-seq⁶, PAS-seq⁴⁸, and PAT-seq⁴⁷) and to estimate polyA tail lengths (for example, PAL-seq¹⁰, TAIL-seq¹¹, and mTAIL-seq⁴⁹). However, a major limitation of NGS-based methods is their inability to assign a given polyA tail length to a specific transcript isoform, which causes a loss of isoform-specific tail length information. In addition, NGS-based methods cannot measure tail lengths greater than the read length, thus biasing our view of polyA tail dynamics to those transcripts that display shorter tail lengths.

Fig. 5 | Analysis of tail composition using Nano3P-seq. **a**, Schematic overview of the standards used to assess the ability of Nano3P-seq to accurately quantify the base content of polyA tails. **b**, IGV snapshots of nucleotide composition in cDNA standard tails sequenced using Nano3P-seq. Gray regions indicate the mapped part of the reads, whereas colored letters indicate soft-clipped bases (unmapped), which are the base-called tails, after trimming the adapter. **c**, Probability of base composition (A, green; G, orange; C, blue; U, red) per position in the last 20 nucleotides of the cDNA standard tails. See also Supplementary Note 1. **d**, Percentage of reads belonging to groups classified on the basis of their polyA tail base composition. Some sequence examples belonging to different groups are illustrated below the bar plots. Samples in this analysis are embryonic

mRNAs across zebrafish developmental stages (2, 4, and 6 h.p.f., shown in blue, green, and red, respectively) in three biological replicates and a control that includes sequin R1 and R2 groups of RNAs (gray). Statistical comparison of means was performed using the Kruskal–Wallis test. $n = 3$ biological replicates, data are presented as mean \pm s.e.m. **e**, PolyA tail length estimation distributions of mRNA reads belonging to groups classified on the basis of their polyA tail base composition across zebrafish development stages (2, 4, and 6 h.p.f.). **f**, Left, IGV snapshots of reads mapping to zebrafish *actb* mRNA. Right, zoomed images of individual reads with different terminal bases (top, all-A reads; middle, Term-G reads; bottom, Term-U reads) are shown.

More recently, methods for estimating polyA tail lengths using PacBio long-read sequencing technologies have been developed, such as FLAM-seq⁴⁵ and PALso-seq⁴⁶. In contrast to NGS, these methods can

capture isoform–tail relationships; however, they are still affected by PCR amplification and ligation biases, in addition to producing relatively modest outputs in terms of the number of reads^{50–52}. In this



regard, several works have concluded that nanopore sequencers do not suffer from GC-content biases, as these are mainly introduced during the PCR amplification step^{53–55}.

Nanopore dRNA-seq has been proposed as an alternative long-read sequencing technology for studying polyA tail lengths^{12,13}; however, the standard dRNA-seq approach cannot capture deadenylated RNAs, molecules with non-canonical tailings (for example, polyuridine), or molecules with polyA tails shorter than 10 nucleotides, thus biasing the view of the transcriptome toward polyadenylated molecules. Customized dRNA-seq methods involving in vitro poly(G/I)-tailing have been developed to overcome some of these limitations, but a lack of bioinformatic tools to distinguish polyI and polyA signals limits their applicability to study polyA tail length differences across transcripts in these datasets^{14,15}. In addition, dRNA-seq requires 500 ng RNA as input, whereas Nano3P-seq requires as little as 50 ng, thus decreasing the required input material by tenfold. Nano3P-seq addresses the current limitations by offering a simple and robust solution for studying the coding and non-coding transcriptome simultaneously regardless of the presence or absence of polyA tails or 3' tail composition, without PCR or ligation biases, and with single-read and single-isoform resolution. Moreover, the use of thermostable group II intron reverse transcriptase (TGIRT) in the Nano3P-seq protocol not only maximizes the production of full-length cDNAs, but also ensures the inclusion of RNA molecules that are highly structured and/or modified, which would often not be captured (or their representation would be significantly biased) using standard viral reverse transcriptases^{56,57}.

Nano3P-seq also provides quantitative measurements of RNA abundances (Fig. 1d) and captures diverse RNA biotypes regardless of their tail end composition (Fig. 2d). We have shown that Nano3P-seq can be applied to diverse species with a wide range of polyA tail lengths (Fig. 3d,e) and can be used to study the dynamics of polyadenylation (Fig. 3f,g,i). Specifically, we have demonstrated that Nano3P-seq provides per-read-resolution transcriptome-wide maps of RNA abundance and polyadenylation dynamics during the zebrafish MZT.

Overall, our work demonstrates that Nano3P-seq can simultaneously capture both non-polyA-tailed and polyA-tailed transcriptomes, making it possible to accurately quantify the RNA abundances and polyA tail lengths at per-read and per-isoform levels, while minimizing the amount of biases introduced during library preparation. These features set Nano3P-seq as a simple, potent, and robust method that can provide mechanistic insights into the regulation of RNA molecules and improve our understanding of mRNA tailing processes and post-transcriptional control.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-022-01714-w>.

References

- Licatalosi, D. D. & Darnell, R. B. RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.* **11**, 75–87 (2010).
- Berkovits, B. D. & Mayr, C. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363–367 (2015).
- Dreyfus, M. & Régnier, P. The poly(A) tail of mRNAs. *Cell* **111**, 611–613 (2002).
- Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11**, 2755–2766 (1997).
- Tian, B., Hu, J., Zhang, H. & Lutz, C. S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **33**, 201–212 (2005).
- Ulitsky, I. et al. Extensive alternative polyadenylation during zebrafish development. *Genome Res.* **22**, 2054–2066 (2012).
- Yartseva, V. & Giraldez, A. J. The maternal-to-zygotic transition during vertebrate development: a model for reprogramming. *Curr. Top. Dev. Biol.* **113**, 191–232 (2015).
- Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* **30**, 581–613 (2014).
- Walser, C. B. & Lipshitz, H. D. Transcript clearance during the maternal-to-zygotic transition. *Curr. Opin. Genet. Dev.* **21**, 431–443 (2011).
- Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**, 66–71 (2014).
- Chang, H., Lim, J., Ha, M. & Kim, V. N. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol. Cell* **53**, 1044–1052 (2014).
- Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
- Tudek, A. et al. Global view on the metabolism of RNA poly(A) tails in yeast *Saccharomyces cerevisiae*. *Nat. Commun.* **12**, 4951 (2021).
- Vo, J. M. et al. Synthesis of modified nucleotide polymers by the poly(U) polymerase Cid1: application to direct RNA sequencing on nanopores. *RNA* **27**, 1497–1511 (2021).
- Drexler, H. L., Choquet, K. & Churchman, L. S. Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* **77**, 985–998 (2020).
- Begik, O. et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.* **39**, 1278–1291 (2021).
- Levy, B. W., Johnson, C. B. & McCarthy, B. J. Diversity of sequences in total and polyadenylated nuclear RNA from *Drosophila* cells. *Nucleic Acids Res.* **3**, 1777–1790 (1976).
- Hardwick, S. A. et al. Spliced synthetic genes as internal controls in RNA sequencing experiments. *Nat. Methods* **13**, 792–798 (2016).
- Giraldez, A. J. et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
- Chang, H. et al. Terminal uridylyltransferases execute programmed clearance of maternal transcriptome in vertebrate embryos. *Mol. Cell* **70**, 72–82.e7 (2018).
- Vejnar, C. E. et al. Genome wide analysis of 3' UTR sequence elements and proteins regulating mRNA stability during maternal-to-zygotic transition in zebrafish. *Genome Res.* **29**, 1100–1114 (2019).
- Bratic, A. et al. Mitochondrial polyadenylation is a one-step process required for mRNA integrity and tRNA maturation. *PLoS Genet.* **12**, e1006028 (2016).
- Baserga, S. J. et al. Polyadenylation of a human mitochondrial ribosomal RNA transcript detected by molecular cloning. *Gene* **35**, 305–312 (1985).
- Kim, D. et al. The architecture of SARS-CoV-2 transcriptome. *Cell* **181**, 914–921 (2020).
- Krause, M. et al. tailfinder: alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA* **25**, 1229–1241 (2019).
- Loman, N. J., Quick, J. & Simpson, J. T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* **12**, 733–735 (2015).
- Carninci, P. et al. The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).

28. Guttman, M. et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
29. Marzluff, W. F., Wagner, E. J. & Duronio, R. J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9**, 843–854 (2008).
30. Battle, D. J. & Doudna, J. A. The stem-loop binding protein forms a highly stable and specific complex with the 3' stem-loop of histone mRNAs. *RNA* **7**, 123–132 (2001).
31. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
32. Begik, O. et al. Integrative analyses of the RNA modification machinery reveal tissue- and cancer-specific signatures. *Genome Biol.* **21**, 97 (2020).
33. Boo, S. H. & Kim, Y. K. The emerging role of RNA modifications in the regulation of mRNA stability. *Exp. Mol. Med.* **52**, 400–408 (2020).
34. Zhao, B. S. et al. m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542**, 475–478 (2017).
35. Ryvkin, P. et al. HAMR: high-throughput annotation of modified ribonucleotides. *RNA* **19**, 1684–1692 (2013).
36. Khoddami, V. et al. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc. Natl Acad. Sci. USA* **116**, 6784–6789 (2019).
37. Motorin, Y., Muller, S., Behm-Ansmant, I. & Branlant, C. Identification of modified residues in RNAs by reverse transcription-based methods. in *Methods in Enzymology* Vol. 425 (ed. Tainer, J. A.) 21–53 (Academic Press, 2007).
38. Werner, S. et al. Machine learning of reverse transcription signatures of variegated polymerases allows mapping and discrimination of methylated purines in limited transcriptomes. *Nucleic Acids Res.* **48**, 3734–3746 (2020).
39. Babaian, A. et al. Loss of m1acp3Ψ ribosomal RNA modification is a major feature of cancer. *Cell Rep.* **31**, 107611 (2020).
40. Brand, R. C., Klootwijk, J., Planta, R. J. & Maden, B. E. Biosynthesis of a hypermodified nucleotide in *Saccharomyces carlsbergensis* 17S and HeLa-cell 18S ribosomal ribonucleic acid. *Biochem. J.* **169**, 71–77 (1978).
41. Liang, X.-H., Liu, Q. & Fournier, M. J. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* **15**, 1716–1728 (2009).
42. Locati, M. D. et al. Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. *RNA* **23**, 1188–1199 (2017).
43. Lim, J. et al. Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* **159**, 1365–1376 (2014).
44. Morgan, M. et al. mRNA 3' uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. *Nature* **548**, 347–351 (2017).
45. Legnini, I., Alles, J., Karaiskos, N., Ayoub, S. & Rajewsky, N. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nat. Methods* **16**, 879–886 (2019).
46. Liu, Y., Nie, H., Liu, H. & Lu, F. Poly(A) inclusive RNA isoform sequencing (PALso-seq) reveals wide-spread non-adenosine residues within RNA poly(A) tails. *Nat. Commun.* **10**, 5292 (2019).
47. Harrison, P. F. et al. PAT-seq: a method to study the integration of 3'-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA* **21**, 1502–1510 (2015).
48. Shepard, P. J. et al. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17**, 761–772 (2011).
49. Lim, J., Lee, M., Son, A., Chang, H. & Kim, V. N. mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev.* **30**, 1671–1682 (2016).
50. Oikonomopoulos, S. et al. Methodologies for transcript profiling using long-read technologies. *Front. Genet.* **11**, 606 (2020).
51. Ramsköld, D. et al. Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
52. Au, K. F. et al. Characterization of the human ESC transcriptome by hybrid sequencing. *Proc. Natl Acad. Sci. USA* **110**, E4821–E4830 (2013).
53. Goldstein, S., Beka, L., Graf, J. & Klassen, J. L. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. *BMC Genomics* **20**, 23 (2019).
54. Sevim, V. et al. Shotgun metagenome data of a defined mock community using Oxford Nanopore, PacBio and Illumina technologies. *Sci. Data* **6**, 285 (2019).
55. Delahaye, C. & Nicolas, J. Sequencing DNA with nanopores: troubles and biases. *PLoS One* **16**, e0257521 (2021).
56. Xu, H., Yao, J., Wu, D. C. & Lambowitz, A. M. Improved TGIRT-seq methods for comprehensive transcriptome profiling with decreased adapter dimer formation and bias correction. *Sci. Rep.* **9**, 7953 (2019).
57. Mohr, S. et al. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958–970 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

¹Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain. ²Garvan Institute of Medical Research, Sydney, New South Wales, Australia. ³Faculty of Medicine, University of New South Wales, Sydney, New South Wales, Australia. ⁴Universitat Pompeu Fabra, Barcelona, Spain. ⁵Department of Genetics, Yale University, New Haven, CT, USA. ⁶Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway. ⁷Sars International Centre for Marine Molecular Biology, University of Bergen, Bergen, Norway. ⁸Department of Genetics and Genome Sciences, Institute for Systems Genomics, University of Connecticut Health Center, Farmington, CT, USA. ⁹School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales, Australia. ✉e-mail: eva.novoa@crg.eu

Methods

In vitro transcription of RNAs

The synthetic 'curlcake' sequences⁵⁸ (Curlcake 1, 2,244 bp and Curlcake 2, 2,459 bp) were in vitro transcribed using Ampliscribe T7-Flash Transcription Kit (Lucigen-ASF3507). Curlcake 2 was polyadenylated using *Escherichia coli* polyA Polymerase (NEB-M0276S). polyA-tailed RNAs were purified using RNAClean XP beads. The quality of the in vitro transcribed (IVT) products as well as the addition of polyA tail to the synthetic constructs was assessed using Agilent 4200 TapeStation (Extended Data Fig. 1a). Concentration was determined using Qubit Fluorometric Quantitation. The purity of the IVT products was measured with NanoDrop 2000 Spectrophotometer.

Poly(A) tail length assay and Sanger sequencing

Poly(A) Tail-Length Assay Kit (Thermo Fisher, 764551KT) was used according to the manufacturer's instructions. PCR products corresponding to the tail- and gene-specific primer combinations of mouse 16S rRNA (universal forward primer: 5'-GGTCGGTTCTATCTATTTACGATTCTC-3', gene-specific reverse primer: 5'-TTCTCTAGGTTAGAGGGGTGACGTA TAT-3') and human *ACTB* (assay control, primer composition not disclosed by manufacturer) were loaded on a 2.5% agarose gel (Lonza, 50010) and stained with GelRed (Merck, SCT123). Product sizes determined using the GeneRuler 50-base-pair DNA ladder (Thermo Scientific, SM0371). Subsequently, tail- and gene-specific PCR products of mouse 16S rRNA were purified by gel-elution (Cytiva Life Sciences, 28903470) and sent for Sanger sequencing with the shared forward primer (5'-GGTCGGTTCTATCTATTTACGATTCTC-3'). Resulting chromatograms were analyzed using SnapGene (v.6.0.2). After confirming alignment to the reference sequence, the unclipped chromatograms were used to visualize 3' ends.

Yeast culturing and total RNA extraction

Saccharomyces cerevisiae (strain BY4741) was grown at 30 °C in standard YPD medium (1% yeast extract, 2% Bacto Peptone and 2% dextrose). Cells were then quickly transferred into 50-ml pre-chilled falcon tubes, and centrifuged for 5 min at 3,000g in a 4 °C pre-chilled centrifuge. Supernatant was discarded, and cells were flash frozen. Flash frozen pellets were resuspended in 700 µl TRIzol (Life Technologies) with 350 µl acid washed and autoclaved glass beads (425–600 µm, Sigma G8772). The cells were disrupted using a vortex on top speed for 7 cycles of 15 s (the samples were chilled on ice for 30 s between cycles). Afterwards, the samples were incubated at room temperature for 5 min and 200 µl chloroform was added. After briefly vortexing the suspension, the samples were incubated for 5 min at room temperature. Then they were centrifuged at 14,000g for 15 min at 4 °C and the upper aqueous phase was transferred to a new tube. RNA was precipitated with 2× volume molecular grade absolute ethanol and 0.1× volume sodium acetate (3 M, pH 5.2). The samples were then incubated for 1 h at –20 °C and centrifuged at 14,000g for 15 min at 4 °C. The pellet was then washed with 70% ethanol and resuspended with nuclease-free water after air drying for 5 min on the benchtop. Purity of the total RNA was measured with the NanoDrop 2000 Spectrophotometer. Total RNA was then treated with Turbo DNase (Thermo, #AM2238) (2 µl enzyme for 50 µl reaction of 200 ng µl⁻¹ RNA) at 37 °C for 15 min, with a subsequent RNAClean XP bead cleanup.

Mice breeding

Experiments were performed with male mice aged between 8 and 10 weeks. All mice were euthanized using CO₂ and tissues were snap frozen in liquid nitrogen. Animals were kept on a 12:12 h light:dark cycle and provided with water and food ad libitum.

RNA isolation from mouse brain

To isolate nuclear/mitochondrial-enriched RNA from the mouse (*Mus musculus*) brain, we followed previously published protocols (<https://doi.org/10.17504/protocols.io.3fkgjkw>) with minor changes. A quarter

of a C57BL6/J mouse brain was used for this protocol, and all samples and reagents were kept on ice during the protocol. Brain tissue was mined with a razor blade into smaller pieces. Cold Nuclei EZ Lysis Buffer (0.01 M Tris-Cl, pH 7.5, 0.06 M KCl, 0.001 M EDTA, 0.5% NP40, 1× Protease Inhibitor (complete Protease Inhibitor Cocktail Tablets, #11697498001, Roche)) was added to the tissue in 1.5-ml eppendorf tube. The sample was homogenized using a 1-ml dounce (stroking –10–20 times), and the homogenate was transferred into a 2-ml eppendorf tube. One milliliter of cold Nuclei EZ Lysis Buffer was added and mixed, followed by 4 min incubation on ice. During the incubation, the sample was gently mixed a couple of times using a pipette. Homogenate was filtered using a 70-µm strainer mesh, and the flowthrough was collected in a polystyrene round-bottom FACS tube and subsequently transferred into a new 2-ml tube. The sample was centrifuged at 500g for 5 min at 4 °C and the supernatant was removed. The nuclei/mitochondria enriched sample was resuspended in another 1.5 ml EZ Lysis buffer and incubated for 5 min on ice. The sample was centrifuged at 500g for 5 min at 4 °C and the supernatant was discarded (cytoplasm). Five-hundred microliters of nuclei wash and resuspension buffer (1 M phosphate-buffered saline, 1% bovine serum albumin, 5 µl SUPERase In RNase Inhibitor (Thermo Fisher, AM2694)) was added to the sample and incubated for 5 min without resuspending to allow buffer interchange. After incubation, 1 ml nuclei wash and resuspension buffer was added and the sample was resuspended. The sample was centrifuged at 500g for 5 min at 4 °C. The supernatant was removed and only ~50 µl was left. Using 1.4 ml nuclei wash and resuspension buffer, the sample was resuspended and transferred to a 1.5-ml eppendorf tube. The last washing step was repeated and the pellet was resuspended in 300 µl nuclei wash and resuspension buffer. RNA was extracted using TRIzol (Life Technologies) following the manufacturer's protocol.

Zebrafish breeding

Wild-type zebrafish (*Danio rerio*) embryos were obtained through natural mating of the TU-AB strain of mixed ages (5–18 months). Mating pairs were randomly chosen from a pool of 60 males and 60 females allocated for each day of the month. Embryos and adult fish were maintained at 28 °C.

Zebrafish total RNA extraction and polyA selection

For RNA samples, 25 embryos per developmental stage and per replicate were collected and flash frozen in liquid nitrogen. Frozen embryos were thawed and lysed in 1 ml TRIzol (Life Technologies) and total RNA was extracted using the manufacturer's protocol. Total RNA concentration was calculated by nanodrop.

For polyA-selected RNA samples, polyadenylated RNAs were isolated with oligo(dT) magnetic beads (New England Biolabs, S1419S) according to the manufacturer's protocol and eluted in 30 µl before nanodrop quantification.

Zebrafish total RNA ribodepletion

Ribodepletion was performed on zebrafish total RNA using riboPOOL oligos (siTOOLS, 055) following the manufacturer's protocol. In brief, 5 µg total RNA in 14 µl was mixed with 1 µl resuspended riboPOOL oligos, 5 µl hybridization buffer and 0.5 µl SUPERase•In RNase Inhibitor (Thermo Fisher, AM2694). The mix was incubated for 10 min at 68 °C, followed by a slow cool down (3 °C min⁻¹) to 37 °C for hybridization. In the meantime, Dynabeads MyOne Streptavidin C1 (Thermo Fisher, 65001) beads were resuspended by carefully vortexing at medium speed. Eighty microliters of Dynabeads bead resuspension (10 mg ml⁻¹) was transferred into a tube, which then was placed on a magnetic rack. After aspirating the supernatant, 100 µl of bead resuspension buffer (0.1 M NaOH, 0.05 M NaCl) was added to the sample and beads were resuspended in this buffer by agitating the tube. Sample was placed on a magnetic rack and the supernatant was aspirated. This step was performed twice. Beads were then resuspended in 100 µl bead wash buffer (0.1 M NaCl)

and placed on magnet to aspirate the supernatant. Beads were then resuspended in a 160 μ l depletion buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl). This suspension was then divided into two tubes of 80 μ l, which will be used consecutively. Twenty microliters of hybridized riboPOOL and total RNA was briefly centrifuged to spin down droplets and it was pipetted into the tube containing 80 μ l of beads in depletion buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl). The tube containing the mix was agitated to resuspend the solution well. Then the mix was incubated at 37 °C for 15 min, followed by a 50 °C incubation for 5 min. Immediately before use, the second tube containing 80 μ l of beads was placed on a magnetic rack and the supernatant was aspirated. After the incubation at 50 °C, the first depletion reaction was placed on a magnet and the supernatant was transferred into the tube containing the other set of beads. The mix was incubated again at 37 °C for 15 min, followed by a 50 °C incubation for 5 min. After briefly spinning down the droplets, the mix was placed on a magnet for 2 min. The supernatant was transferred into a different tube and cleaned up using RNA Clean & Concentrator-5 (Zymo, R1013).

HeLa cell line culture and polyA RNA selection

HeLa cell lines (obtained from ATCC, *Mycoplasma*-tested in-house) were cultured in DMEM high glucose (Thermo Fisher, 41965) with 10% fetal bovine serum supplement (Thermo Fisher, #10270106). Pellets were obtained from 6 million cells and 1 ml TRIzol (Life Technologies) was added to each pellet. Sample was incubated at room temperature for a few minutes after adding 200 μ l chloroform and vortexing briefly. After the incubation, the solution was centrifuged at 4 °C with 14,000g. Aqueous phase from the previous step was transferred to another eppendorf tube and equal amount of absolute ethanol (Merck, #1009835000) was added to the solution. This suspension was then transferred to an RNeasy Mini Spin Column from RNeasy Mini Kit (Qiagen, #74104), and total RNA was isolated following the manufacturer's protocol. Then, polyadenylated RNAs were isolated with Dynabeads Oligo(dT) 25 (Thermo Fisher, #61005) according to the manufacturer's protocol and eluted in 20 μ l. Concentration and quality of RNA was evaluated using a Nanodrop 2000 Spectrophotometer and an Agilent 4200 TapeStation, respectively.

Nano3P-Seq library preparation

The protocol is based on the direct cDNA Sequencing ONT protocol (DCB_9091_v109_revC_04Feb2019), with several modifications to be able to perform TGIRT template switching. Before starting the library preparation, 1 μ l 10 μ M R_RNA (oligo: 5' rGrArArGrArUrArGrArGrCrGrArCrArGrGrCrArArGrUrGrArUrCrGrGrArArG/3SpC3/3') and 1 μ l 10 μ M D_DNA (5' /5Phos/CTTCCGATCACTTGCCTGTCGCTCTATCTTCN 3') were mixed with 1 μ l 0.1 M Tris pH 7.5, 1 μ l 0.5 M NaCl, 0.5 μ l RNase Inhibitor Murine (NEB, M0314S) and 5.5 μ l RNase-free water. The mix was incubated at 94 °C for 1 min and the temperature was ramped down to 25 °C (-0.1 °C s^{-1}) in order to pre-anneal the oligos. Then, 50 ng RNA was mixed with 2 μ l pre-annealed R_RNA + D_DNA oligo, 1 μ l 100 mM dithiothreitol, 4 μ l 5 \times TGIRT Buffer (2.25 M NaCl, 25 mM MgCl₂, 100 mM Tris-HCl, pH 7.5), 1 μ l RNase Inhibitor Murine (NEB, M0314S), 1 μ l TGIRT (InGex) and nuclease-free water up to 19 μ l. We should note that if 50 ng are used as input, only 1 μ l TGIRT is needed, whereas if 100 ng is used as input, 2 μ l TGIRT enzyme is needed. The reverse transcription mix was initially incubated at room temperature for 30 min before adding 1 μ l 10 mM dNTP mix. Then the mix was incubated at 60 °C for 60 min and inactivated by heating at 75 °C for 15 min before moving to ice. RNase Cocktail (Thermo Scientific, AM2286) was added to the mix to digest the RNA, and the mix was incubated at 37 °C for 10 min. The reaction was then cleaned up using 0.8 \times AMPure XP Beads (Agencourt, A63881). To be able to ligate the sequencing adapters to the first cDNA strand, 1 μ l 10 μ M CompA_DNA (5' GAAGATAGAGCGACAGGCAAGTATCG-GAAGA 3') was annealed to the 15 μ l cDNA in a tube with 2.25 μ l 0.1 M Tris pH 7.5, 2.25 μ l 0.5 M NaCl and 2 μ l nuclease-free water. The mix was

incubated at 94 °C for 1 min and the temperature was ramped down to 25 °C (-0.1 °C s^{-1}) to anneal the complementary to the first-strand cDNA. Then, 22.5 μ l first-strand cDNA was mixed with 5 μ l Adapter Mix (AMX), 22.5 μ l RNase-free water and 50 μ l Blunt/TA Ligase Mix (NEB, M0367S) and incubated in room temperature for 10 min. The reaction was cleaned up using 0.8 \times AMPure XP beads, using WSB (Washing buffer) buffer for washing. The sample was then eluted in elution buffer and mixed with sequencing buffer and loading beads before loading onto a primed R9.4.1 flowcell. Libraries were run on either Flongle or MinION flow cells with MinKNOW acquisition software version v.3.5.5. A detailed step-by-step Nano3P-seq protocol is provided as a Supplementary Protocol.

Annealing-based dcDNA-seq library preparation with TGIRT

Some adjustments were made to the original Direct cDNA-Sequencing ONT protocol (SQK-DCS109), to be able to use TGIRT (InGex) as reverse transcription enzyme for nanopore sequencing, as this enzyme does not produce CCC overhang, which is typically exploited by the dcDNA-seq library preparation protocol (Fig. 1a). In brief, 1 μ l 100 μ M reverse transcription primer VNP (5' /5Phos/ACTTGCTGTCGCTC-TATCTTCTTTTTTTTTTTTTTTTTTTTTVN 3') and 1 μ l 100 μ M of in-house designed complementary oligo (CompA: 5' GAAGATAGAGCGACAG-GCAAGTA 3') were mixed with 1 μ l 0.2 M Tris pH 7.5, 1 μ l 1 M NaCl and 16 μ l RNase-free water. The mix was incubated at 94 °C for 1 min and the temperature was ramped down to 25 °C (-0.1 °C s^{-1}) to pre-anneal the oligonucleotides. Then, 50 ng polyA-tailed RNA was mixed with 1 μ l pre-annealed VNP + CompA, 1 μ l 100 mM dithiothreitol, 4 μ l 5 \times TGIRT Buffer (2.25 M NaCl, 25 mM MgCl₂, 100 mM Tris-HCl, pH 7.5), 1 μ l RNase Inhibitor Murine (NEB, M0314S), 1 μ l TGIRT and nuclease-free water up to 19 μ l. The reverse transcription mix was initially incubated at room temperature for 30 min before adding 1 μ l 10 mM dNTP mix. Then the mix was incubated at 60 °C for 60 min and inactivated by heating at 75 °C for 15 min before moving onto ice. Furthermore, RNase Cocktail (Thermo Scientific, AM2286) was added to the mix to digest the RNA and the mix was incubated at 37 °C for 10 min. Then the reaction was cleaned up using 0.8 \times AMPure XP Beads (Agencourt, A63881). To be able to ligate the sequencing adapters the first strand, 1 μ l 10 μ M CompA was again annealed to the 15 μ l cDNA in a tube with 2.25 μ l 0.1 M Tris pH 7.5, 2.25 μ l 0.5 M NaCl and 2 μ l nuclease-free water. The mix was incubated at 94 °C for 1 min and the temperature was ramped down to 25 °C (-0.1 °C s^{-1}) to anneal the complementary to the first-strand cDNA. Furthermore, 22.5 μ l first-strand cDNA was mixed with 5 μ l Adapter Mix (AMX), 22.5 μ l RNase-free water and 50 μ l Blunt/TA Ligase Mix (NEB, M0367S) and incubated at room temperature for 10 min. The reaction was cleaned up using 0.8 \times AMPure XP beads, using WSB Buffer for washing. The sample was then eluted in elution buffer and mixed with sequencing buffer and loading beads before loading onto a primed R9.4.1 flowcell and run on a MinION sequencer with MinKNOW acquisition software version v.3.5.5.

Synthetic cDNA standards

A total of 12 synthetic cDNA standards were synthesized as ultramers by IDT (Integrated DNA Technologies) to assess the tail length estimation and tail composition quantification accuracy of Nano3P-seq.

Synthetic cDNA standards designed to assess accuracy in tail length estimation:

cDNA_pA_standard_0:/5Phos/CTTCCGATCACTTGCCTGTCGCTC-TATCTTCGTAATAGAAATAGACTAGCTCCACTTTTAAAGAATTATTATG-CAATTAATACATGGGTGACCAAAAGAGCGGGCGGATACACCGTCAC-CACAAGCAGAATAAAAGGTAAACCTGAAATGTTTTAACATAAAAT-GAAAAATGCTTGTGCAACCCTATATAGAA

cDNA_pA_standard_15:/5Phos/CTTCCGATCACTTGCCTGTCGCTC-TATCTTCTTTTTTTTTTTTTTTGTAATAGAAATAGACTAGCTCCACTTT-TAAGAATTATTATGCAATTAATACATGGGTGACCAAAAGAGCGGGCG-GATACACCGCTCACCACAAGCAGAATAAAAGGTAAACCTGAAATT-GTTTTAACATAAAATGAAAAATGCTTGTGTTG

sequences used to map the runs mentioned above are included in the Nano3P-seq GitHub repository.

Estimation of polyA tail lengths

For dRNA-seq reads, polyA tail length estimation was performed using NanoTail, a module from Master of Pores⁶³, a nextflow workflow for the analysis of direct RNA datasets, which uses internally Nanopolish v0.11.1²⁶. In NanoTail, all reads stored in the fastq files are first indexed with 'nanopolish index' using default parameters, and the function 'nanopolish polyA' is used to perform polyA tail length estimations. We should note that neither Nanopolish nor TailfindR estimate polyA tail lengths from the base-called sequence (owing to increased deletion frequencies observed in homopolymeric regions, which is an inherent problem of nanopore sequencing⁶⁴); rather, they estimate polyA tail lengths by comparing the relative duration of the current signal corresponding to the polyA tail region to the total duration of the sequenced read.

For Nano3P-seq reads, polyA tail length estimation was performed using the Nano3P-seq version of tailfindR (<https://github.com/adnaniazitailfindr/tree/nano3p-seq>). All code used to estimate polyA tail lengths and post-process Nano3P-seq data can be found at https://github.com/novoalab/Nano3P_Seq.

Analysis of tail composition

Base-called reads mapped to the zebrafish genome and cDNA standards for tail composition quantification were first trimmed using the Porechop tool (<https://github.com/rrwick/Porechop>) with the following parameters '--extra_end_trim 0 --end_threshold 50', to remove the adapter sequences. Because Porechop sometimes failed at removal of the adapter sequences, only reads containing more than 80% A bases in their tail composition were kept for downstream analyses, thus ensuring that untrimmed reads are not included in downstream analyses (Extended Data Figure 9). Finally, an in-house Python script was used to extract the tail FASTA sequences of the tail regions from trimmed reads, which has been made available in GitHub (https://github.com/novoalab/Nano3P_Seq/blob/master/python_scripts/soft_clipped_content.py).

Animal ethics

Fish lines were maintained according to the International Association for Assessment and Accreditation of Laboratory Animal Care research guidelines, and protocols were approved by the Yale University Institutional Animal Care and Use Committee (IACUC). Mice maintenance was approved by the Garvan/St Vincent's Hospital Animal Ethics Committee, in accordance with the guidelines of the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (Project No. 16/14 and 16/26). All animals were entered into the study in a randomized order and operators were blinded to genotype and treatments.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Base-called FAST5 reads from Nano3P-seq and dRNA-seq libraries are publicly available in the European Nucleotide Archive, under accession code [PRJEB53494](https://www.ebi.ac.uk/ena/record/PRJEB53494). PAL-Seq polyA tail length estimates used in this work were obtained from the Gene Expression Omnibus (GEO) with the accession code [GSE52809](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52809) (ref. ¹⁰). PAT-seq polyA tail length estimates used in this work were obtained from GEO with the accession code [GSE53461](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53461) (ref. ⁴⁷). FLAM-seq polyA tail length estimates used in this work were obtained from GEO with the accession code [GSE126465](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126465) (ref. ⁴⁵). TAIL-seq polyA tail length estimates used in this work were obtained from GEO with the accession code [GSE51299](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51299) (ref. ⁴⁵). All sequencing runs included in this work are listed in Supplementary Table 2. Source data are provided with this paper.

Code availability

All scripts and code used in this work have been made available on GitHub (https://github.com/novoalab/Nano3P_Seq). The code for analyzing Nano3P-seq polyA tail lengths using tailfindR is available on GitHub (<https://github.com/adnaniazitailfindr/tree/nano3p-seq>) and is included as Supplementary Software.

References

- Liu, H. et al. Accurate detection of m6A RNA modifications in native RNA sequences. *Nat. Commun.* **10**, 4079 (2019).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Thomas, N. K. et al. Direct nanopore sequencing of individual full length tRNA strands. *ACS Nano* **15**, 16642–16653 (2021).
- Cozzuto, L. et al. MasterOfPores: a workflow for the analysis of Oxford Nanopore direct RNA sequencing datasets. *Front. Genet.* **11**, 211 (2020).
- Huang, Y.-T., Liu, P.-Y. & Shih, P.-W. Homopolish: a method for the removal of systematic errors in nanopore sequencing by homologous polishing. *Genome Biol.* **22**, 95 (2021).

Acknowledgements

We thank all the members of the Novoa lab for their valuable insights and discussion. O.B. is supported by a UNSW International PhD fellowship and Australian Government Research Training Program Scholarship. G.D. is part of the ROPES ITN, which received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement number 956810. A.D.-T. is supported by an FPI-SO fellowship (PRE2019-088498). This work was supported by the Australian Research Council (DP180103571 to E.M.N.), the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) (PGC2018-098152-A-100 to E.M.N.), the European Research Council (ERC-StG-2021 No 101042103 to E.M.N.) and the NIH (R35GM122580 and RO1HD00035 to A.J.G.). We acknowledge the support of the MEIC to the EMBL partnership, Centro de Excelencia Severo Ochoa and CERCA Programme/Generalitat de Catalunya. We thank T. Mercer (Garvan Institute, Australia) for providing us with the sequins used as spike-ins in Nano3P-seq zebrafish runs. We thank the Sdelci lab (CRG, Spain) for providing us with HeLa cell lines pellets used in this work.

Author contributions

O.B. performed the majority of wet laboratory experiments. O.B. analyzed the data, together with E.M.N. G.D. performed quantitative PCR experiments and G/I-based polyA tail length experiments to orthogonally validate Nano3P-seq predictions. H.L. contributed code for the analysis of polyA tail lengths and nucleotide content. A.D.-T. processed and analyzed the direct RNA sequencing zebrafish data. A.M.N. and E.V. adapted tailfindR code to accommodate the adapters that are used in Nano3P-seq libraries. C.K., A.J.G. and J.-D.B. provided the zebrafish total and polyA-selected RNA samples used in this study. E.M.N. conceived the project. E.M.N. supervised the work, with the assistance of J.S.M. O.B. and E.M.N. wrote the paper, with contributions from all authors.

Competing interests

E.M.N. has received travel and accommodation expenses to speak at Oxford Nanopore Technologies conferences. E.M.N. is a member of the Scientific Advisory Board of IMMAGINA Biotech. All authors declare that the research was conducted in the absence of any

commercial or financial relationships that could be construed as a potential conflict of interest.

Additional information

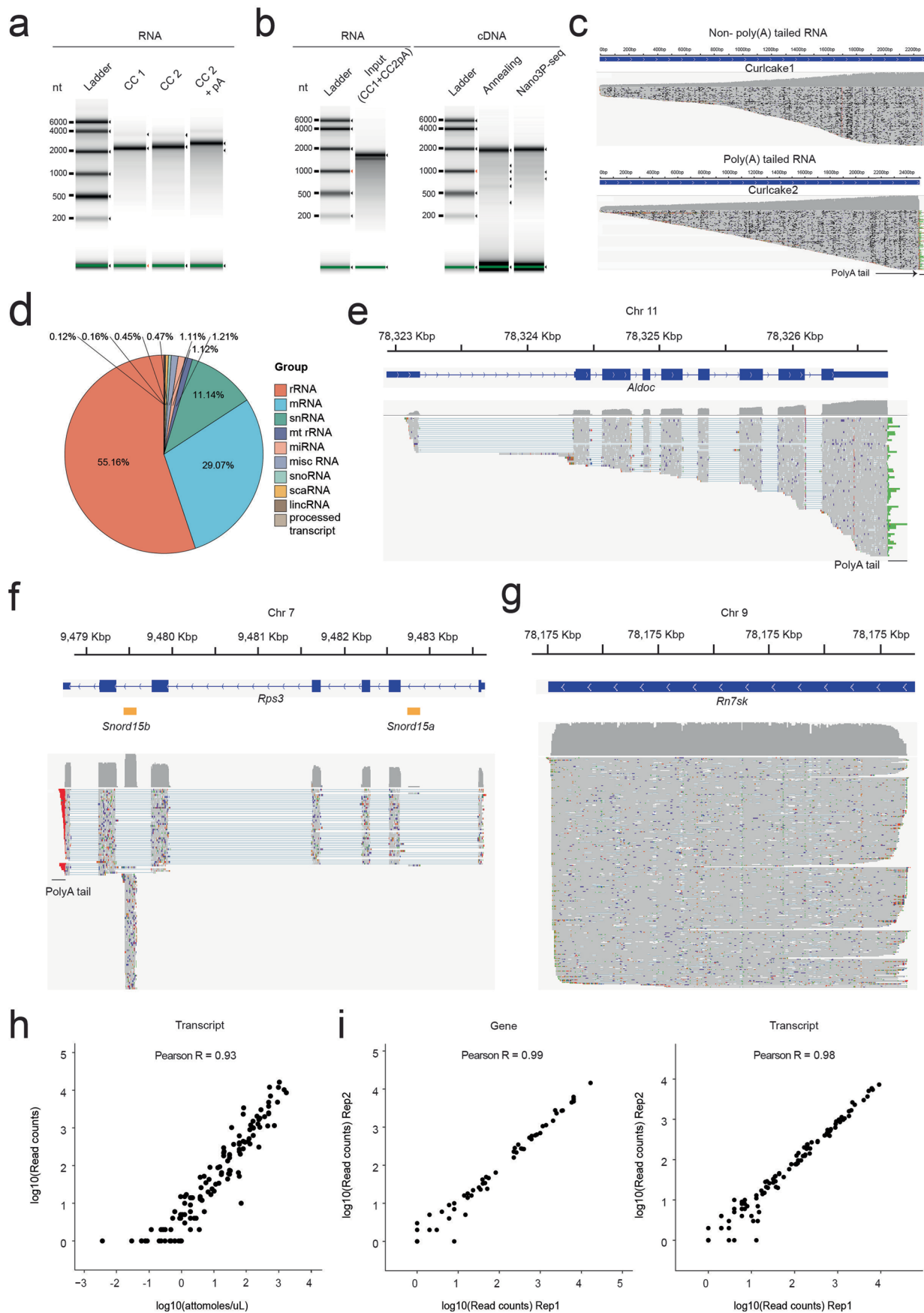
Extended data is available for this paper at <https://doi.org/10.1038/s41592-022-01714-w>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-022-01714-w>.

Correspondence and requests for materials should be addressed to Eva Maria Novoa.

Peer review information *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Lei Tang, in collaboration with the *Nature Methods* team. Peer reviewer reports are available.

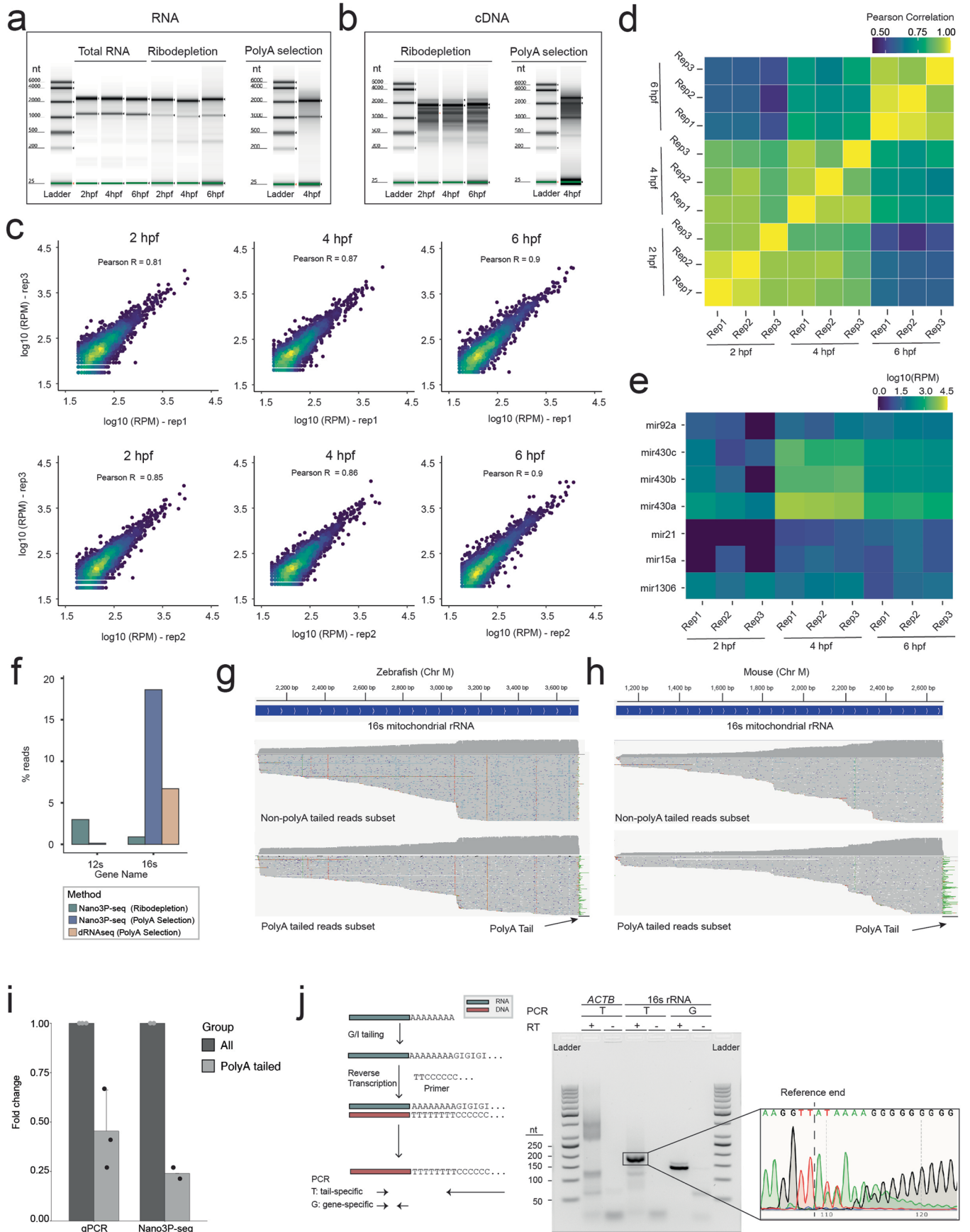
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Nano3P-seq captures non-polyA-tailed and polyA-tailed RNAs. (a) Tape-station profiles of synthetic RNAs ('curlcakes') after being *in-vitro* transcribed and polyA tailed (pA). Similar profiles were consistently obtained in independent experiments. (b) Tape-station profiles of the input RNA (curlcake mix) for reverse-transcription and cDNA produced after annealing-based or template-switching based (Nano3P-seq) reverse-transcription. (c) IGV snapshots of synthetic RNAs (Curlcake1 and Curlcake2, see *Methods*) illustrating that Nano3P-seq captures both non-polyadenylated (above) and polyadenylated (below) RNAs. The PolyA tail region is shown in green. (d) Pie chart showing the abundance of different RNA types in Nano3P-seq of mouse nuclear/mitochondria

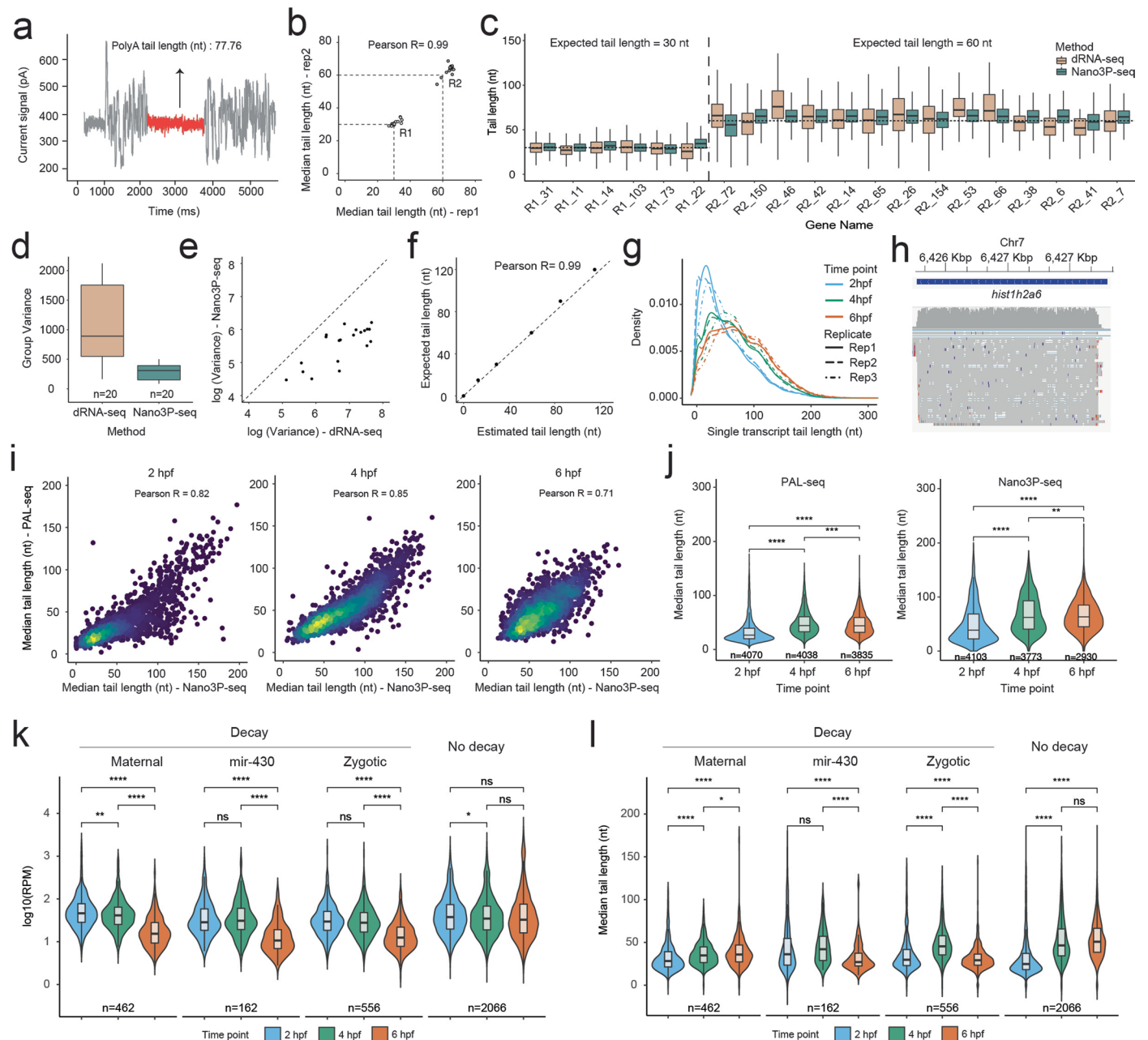
enriched RNA. (e) IGV snapshot of reads mapping to *Aldoc* gene with polyA tail shown in green. (f) IGV snapshot of reads mapping to *Rps3* and *Snord15b* genes. PolyA tail can be seen in green on the reads mapping to *Rps3* mRNA, while it can't be seen in *Snord15b* snoRNA. (g) IGV snapshot of reads mapping to *Rn7sk* miscRNA, which are not expected to contain polyA tails. (h) Scatter plot of the log transformed concentrations (Attomoles/uL) and read counts of sequin transcripts (Pearson R: 0.89, Slope: 0.92). Each dot represents a sequin transcript. (i) Scatter plot of the replicability of the log (read counts) of synthetic sequins using Nano3P-seq, both at per-gene level (left panel, Pearson's R: 0.99) as well as per-transcript level (right panel, Pearson's R: 0.98).



Extended Data Fig. 2 | See next page for caption.

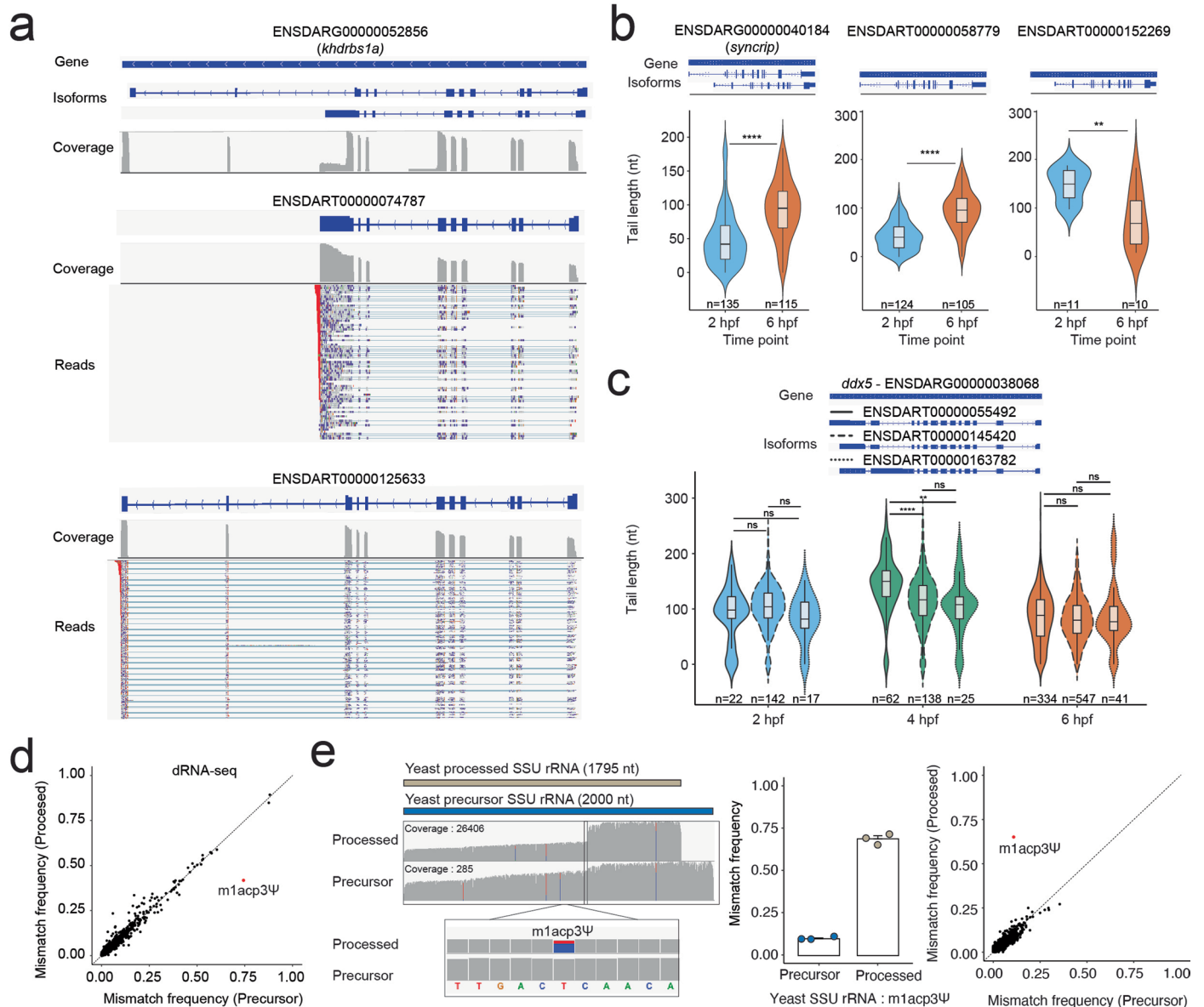
Extended Data Fig. 2 | Analysis of abundances and polyA tails in mitochondrial rRNAs. **(a)** TapeStation profiles of total, ribodepleted and polyA-selected RNAs from zebrafish embryos collected in three different time points during the MZT. Similar TapeStation profiles were obtained across other biological replicates ($n = 3$). **(b)** TapeStation profiles of the reverse-transcription products of ribodepleted (left) and polyA-selected (right) samples from zebrafish embryos collected in three different time points during the MZT. Similar TapeStation profiles were obtained across other biological replicates ($n = 3$). **(c)** Scatterplots depicting the correlation of mRNA RPM (Read per million) levels biological replicates in three different time points during the MZT. **(d)** Pearson correlation matrix illustrating the similarity between biological replicates and different time points during the MZT. **(e)** Heatmap of $\log_{10}(\text{RPM})$ values of micro-RNAs in three different time points during the MZT in three biological replicates. **(f)** Percentage of reads mapping to 12 s and 16 s mitochondrial rRNAs in two different methods: Nano3P-seq of ribodepleted and polyA-selected samples and dRNA-seq of polyA-selected samples from zebrafish embryos at 4 hours post-fertilization. **(g)** IGV snapshot of reads mapping to zebrafish 16 s mitochondrial rRNA, where reads have been grouped as non-

polyA tailed and polyA tailed based on their predicted polyA tail length. PolyA tail region is shown with an arrow and colored green. **(h)** IGV snapshot of reads mapping to mouse 16 s mitochondrial rRNA, where reads have been grouped as non-polyA tailed and polyA tailed based on their predicted polyA tail length. PolyA tail region is shown with an arrow and colored green. **(i)** Fold change of the polyA tailed 16 s mitochondrial rRNA amount to the total 16 s mitochondrial rRNA amount measured by qPCR and Nano3P-seq ($n = 3$ technical replicates for qPCR and $n = 2$ biological replicates for Nano3P-seq. Data are presented as mean values \pm standard deviation). **(j)** Outline of PolyA Tail-Length Assay Kit (Thermo, #764551KT) (left panel). Agarose gel electrophoresis image of the PCR products of the PolyA Tail-Length Assay Kit illustrating bands in the tail-specific PCR of *ACTB* control and both tail-specific and gene-specific PCR of mouse 16 s mitochondrial rRNA (middle panel). Sanger sequencing result of the PCR product extracted from the agarose gel showing the presence of polyA tail after the reference end indicated by a dashed line (right panel). Also refer to Supplementary Note 2. PolyA Tail Length Assay experiments were performed in 2 independent replicates.



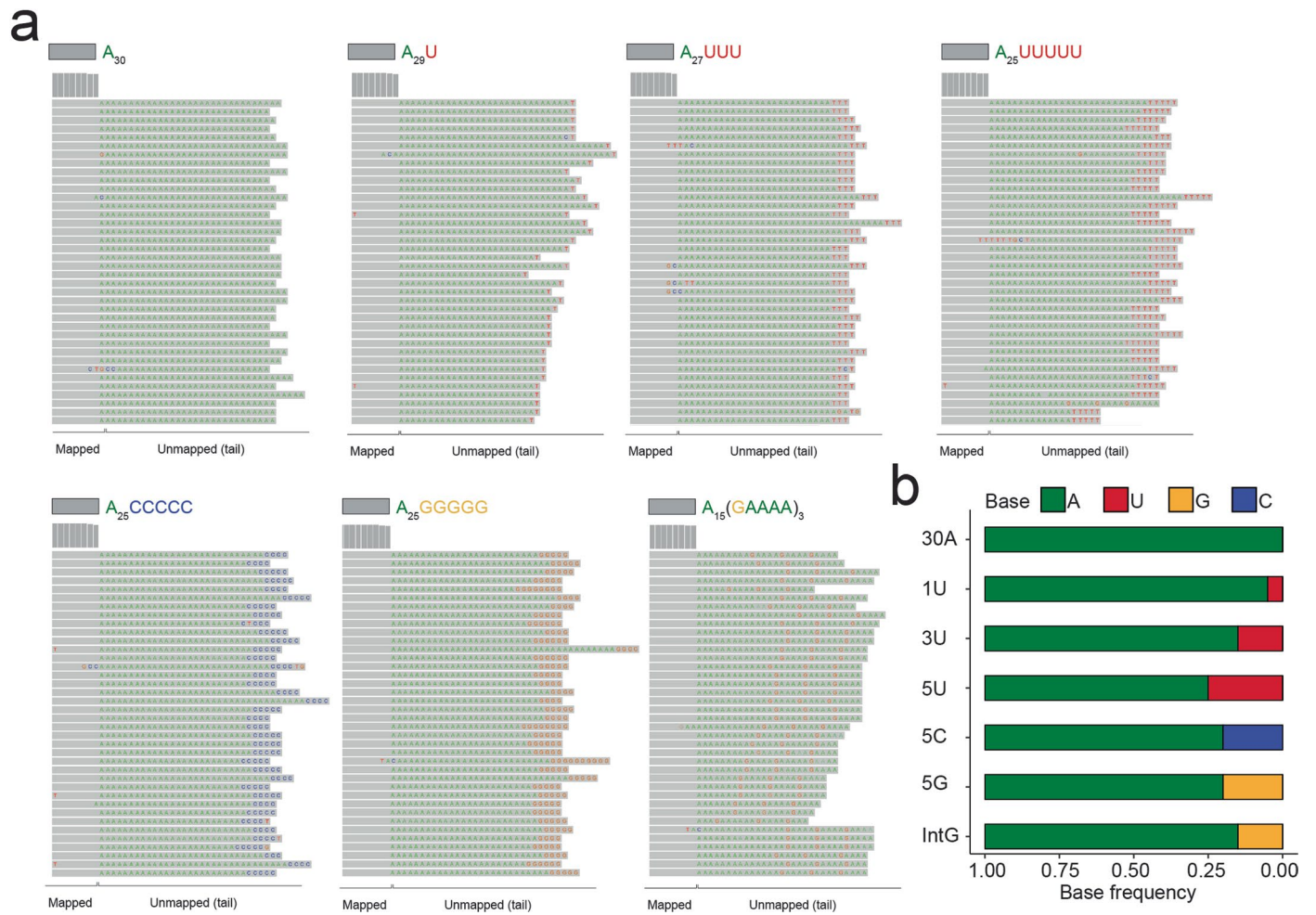
Extended Data Fig. 3 | Analysis of polyA tail lengths using Nano3P-seq. (a) Current intensity (pA) plot of a synthetic polyA+ read, obtained using Nano3P-seq. The homopolymeric poly(T) region is highlighted in orange. **(b)** Replicability of median per-gene polyA tail length estimation in sequins captured with Nano3P-seq. The polyA tail length of synthetic sequins is 30 nt (R1 sequins) or 60 nt (R2 sequins). **(c)** Overall comparison of polyA tail length estimation of R1 and R2 sequins which contain 30 nt and 60 nt polyA tail lengths, respectively, was obtained using dRNA-seq (orange) and Nano3P-seq (green). Genes with coverage greater than 30 reads are included. Data are presented as mean values \pm standard error (Please see Table S4 for exact median tail length estimations and read counts for each sequin gene). **(d)** Per-gene variance of polyA tail length estimations of sequins obtained using dRNA-seq (orange) and Nano3P-seq (green). The number of reads included in the analysis is shown below each boxplot. **(e)** Scatter plot showing the comparison between per-gene variance of polyA tail length estimations obtained using dRNA-seq and Nano3P-seq. **(f)** Correlation between expected tail length (nt) and estimated tail length (nt) of cDNA standards. **(g)** Distribution of polyA tail lengths in mRNAs across zebrafish developmental stages (2, 4 and 6 hpf, shown in blue, green and red respectively)

in three biological replicates (shown as full lines, dashed lines, and dotted/dashed lines respectively). **(h)** IGV snapshot of reads mapping to *hist1h2a6* mRNA, which lacks polyA tails. **(i)** Scatterplots of median per-gene polyA tail length estimations using Nano3P-seq and PAL-seq from zebrafish mRNAs at 2 hpf (left), 4 hpf (middle) and 6 hpf (right). Each dot represents the median polyA tail length of a given gene. **(j)** Violin plots depicting the distribution of median per-gene polyA tail length estimations during the zebrafish MZT, estimated using PAL-Seq (left) or Nano3P-seq (right). **(k)** Comparative analysis of the abundance (shown as \log_2 RPM) of zebrafish mRNAs that have been binned according to their previously annotated decay mode (maternal decay, zygotic activation-dependent decay, miR-430-dependent decay and no decay) during MZT using PAL-seq data. **(l)** Median per-gene polyA tail length estimations of the 4 groups of zebrafish mRNAs during MZT using PAL-seq data. For figures S3j–l; the number of genes included in the analysis is shown below each violinplot. Boxplot limits are defined by lower (bottom) and upper (top) quartile, while the bar indicates the median and whiskers indicate $\pm 1.5X$ inter-quartile range. Statistical analyses were performed using Kruskal-Wallis test. ($p > 0.05$: ns, $p \leq 0.05$ *, $p \leq 0.01$ ***, $p \leq 0.001$ ****).



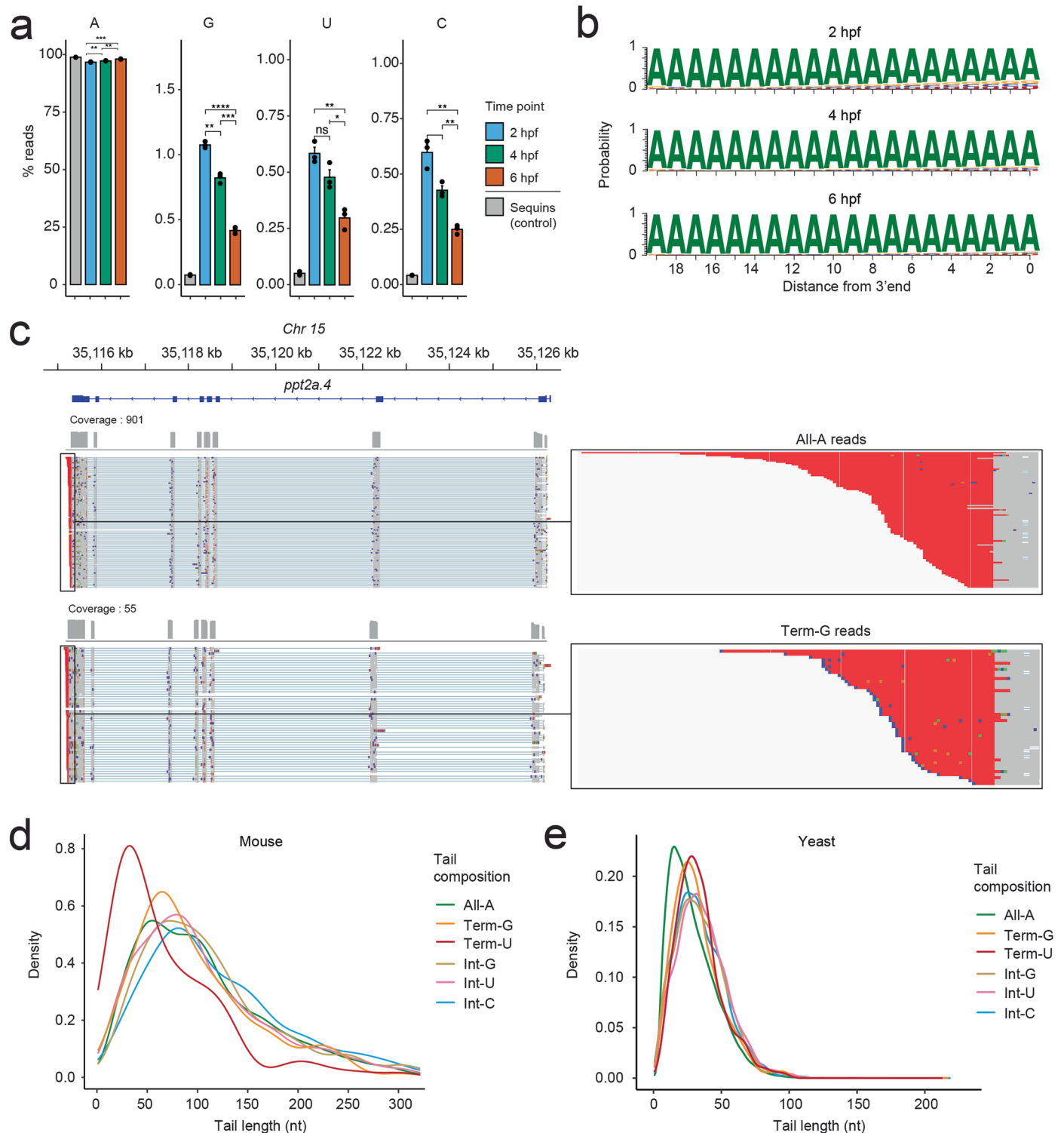
Extended Data Fig. 4 | Analysis of isoform-specific polyA tail and modification dynamics using Nano3P-seq. (a) IGV coverage tracks of reads mapping to *khdrbs1a* gene from zebrafish embryos at 2 hpf obtained by Nano3P-seq. Annotations of the gene and two main isoforms are shown at the top of the panels. Individual reads mapping to each isoform is illustrated below each coverage track. PolyA tails of individual reads are shown in red. **(b)** Comparison of polyA tail length distributions of reads mapping to *syncrip*, illustrated at the per-gene (left panel) and per-isoform (middle and right panel) level measured at two time points during the zebrafish MZT. Annotations of the gene and two main isoforms are shown at the top of the panels. **(c)** Comparison of polyA tail length distributions of reads mapping to three distinct isoforms (full, dashed and dotted outline) of *ddx5* measured at the three time points during the zebrafish MZT. Annotations of the gene and three main isoforms are shown at the top of the panels. Only isoforms with more than 10 reads are shown. The number of reads included in the analysis is shown below each violinplot. P-values have been computed using the Kruskal–Wallis test and corrected for multiple testing

using the Benjamini–Hochberg procedure ($p > 0.05$:ns, $p \leq 0.05$ *, $p \leq 0.01$ ***, $p \leq 0.001$:****). **(d)** Comparison of the per-site mismatch frequencies observed in reads mapping to yeast precursor SSU rRNA and to yeast processed SSU rRNA sequenced by dRNA-seq, showing that the unique outlier is *m¹acp³Ψ*. **(e)** IGV coverage tracks of reads mapping to yeast processed small subunit (SSU) rRNA (upper track) and precursor SSU rRNA (lower track), including a magnified image at the position known to be modified with *m¹acp³Ψ* (left panel). Positions with a mismatch frequency lower than 0.1 are shown in gray. Mismatch frequency values in yeast precursor and processed SSU rRNA at the position known to be modified with *m¹acp³Ψ* (middle panel) ($n = 3$ biological replicates, data are presented as mean values \pm standard error). Comparison of the per-site mismatch frequencies observed in reads mapping to yeast precursor SSU rRNA and to yeast processed SSU rRNA, showing that the unique outlier is *m¹acp³Ψ* (right panel). Boxplot limits are defined by the lower (bottom) and upper (top) quartile. The bar indicates the median, and whiskers indicate ± 1.5 X interquartile range.



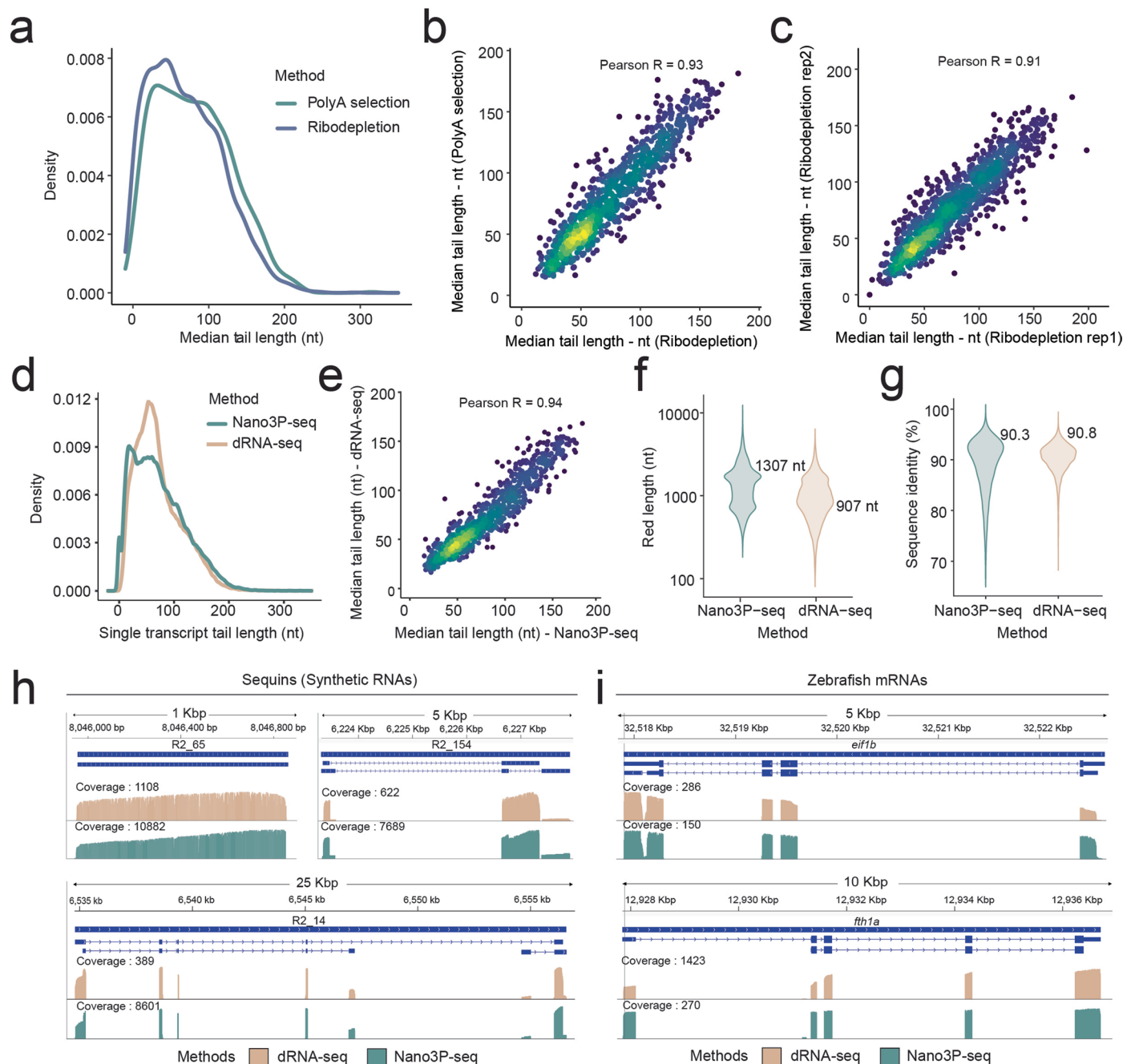
Extended Data Fig. 5 | Analysis of tail composition using Nano3P-seq in vitro samples. (a) IGV snapshots of nucleotide composition in cDNA standard tails sequenced using Nano3P-seq. Gray regions indicate the mapped part of the reads, whereas colored letters indicate soft-clipped bases (unmapped) which are

the base-called tails, after trimming the adapter. **(b)** PolyA tail base frequency distribution (A: green, G: orange, C: blue, U: red) of cDNA standards sequenced with Nano3P-seq.



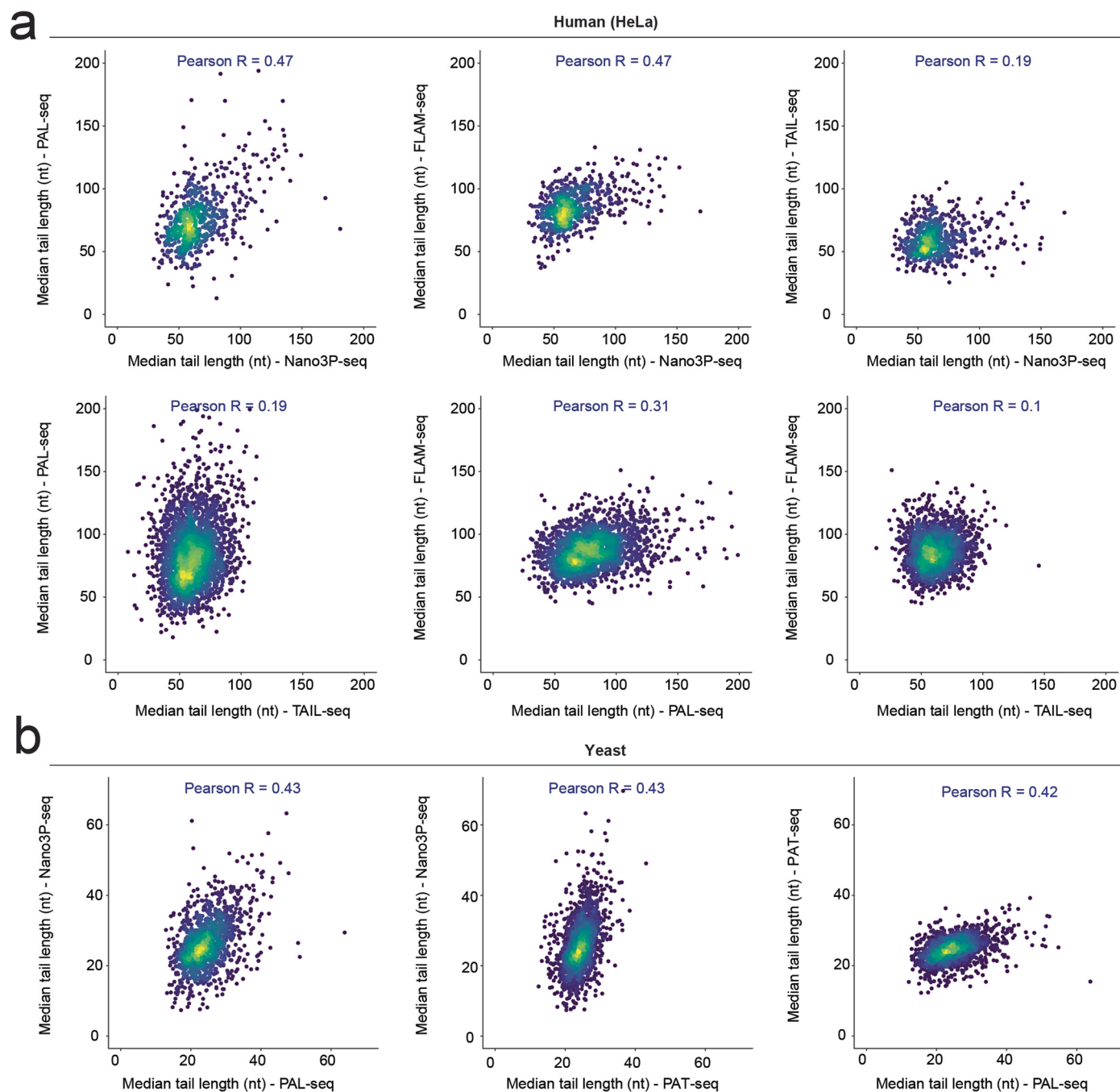
Extended Data Fig. 6 | Analysis of tail composition using Nano3P-seq in vivo samples. (a) Overall nucleotide composition in mRNA tails at 3 time points during the zebrafish MZT (2, 4 and 6 hpf, shown in blue, green and red respectively) and control that includes sequin R1 and R2 groups of RNAs (gray). P-values have been computed using Kruskal-Wallis test. ($n = 3$ biological replicates, data are presented as mean values \pm standard error). (b) Probability of base composition (A: green, G: orange, C: blue and U: red) per position in the last 20 nucleotide of the mRNA tails at 3 time points during the zebrafish MZT

(2, 4 and 6 hpf). (c) IGV snapshots of reads mapping to zebrafish *ppt2a.4* mRNA (left panel). In the right panel, zoomed images of 3' ends of individual reads with different terminal bases (all reads: top, Term-G reads: bottom) are shown. (d) PolyA tail length estimation distributions of mRNA reads belonging to groups classified based on their polyA tail base composition in mouse. (e) PolyA tail length estimation distributions of mRNA reads belonging to groups classified based on their polyA tail base composition in yeast. ($p > 0.05$: ns, $p \leq 0.05$: *, $p \leq 0.01$: **, $p \leq 0.001$: ***, $p \leq 0.0001$: ****).



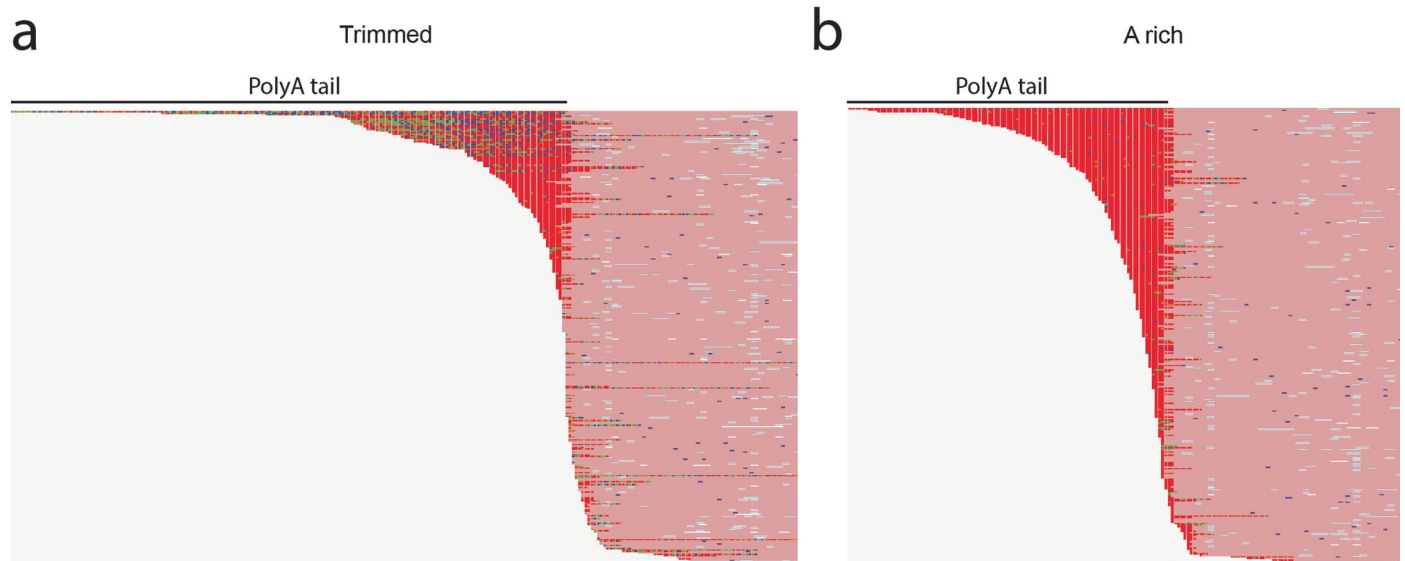
Extended Data Fig. 7 | Comparison of polyA tail length estimations and run stats between dRNA-seq and Nano3P-seq. (a) Distribution of median per-gene polyA tail length estimations from 4 hpf zebrafish embryos, isolated using either polyA selection (green) or ribodepletion (blue). (b) Comparison of median per-gene polyA tail length estimations between polyA-selected and ribodepleted zebrafish mRNAs isolated at 4 hpf. Each dot represents a gene. (c) Comparison of median per-gene polyA tail length estimations of mRNAs in zebrafish ribodepleted samples (replicate 1 and 2) isolated at 4 hpf. Each dot represents a gene. (d) Distribution of polyA tail length estimations in mRNAs from 4 hpf zebrafish embryos isolated using polyA selection and sequenced with dRNA-seq (orange) or Nano3P-seq (green). (e) Comparison of median per-gene polyA tail length estimations of polyA-selected mRNAs isolated at 4 hpf with dRNA-seq or Nano3P-seq. Each dot represents a gene. (f) Read length distribution

of mapped reads from 4 hpf zebrafish embryos isolated using polyA selection and sequenced with Nano3P-seq (green) and dRNA-seq (orange), with median lengths of 1307 nt and 907 nt, respectively. (g) Sequence identity (%) of the reads from 4 hpf zebrafish embryos isolated using polyA selection and sequenced with either Nano3P-seq (green) or dRNA-seq (orange) with median values of 90.3 and 90.8, respectively. The reads were mapped in both cases to *D. rerio* GRCz11 reference. (h) IGV coverage tracks of reads mapping to R2_65, R2_154, and R2_14 from synthetic 'sequin' RNAs, obtained using either dRNA-seq (orange) and Nano3P-seq (green). Annotations of the gene and isoforms are shown at the top of each panel. (i) IGV coverage tracks of reads mapping to genes *eif3b* and *fth1a* from zebrafish embryos at 4 hpf, obtained using either dRNA-seq (orange) or Nano3P-seq (green). Annotations of the gene and isoforms are shown at the top of each panel.



Extended Data Fig. 8 | PolyA tail length estimation comparisons between different methods. (a) Scatterplots of median per-gene polyA tail length estimations from HeLa mRNAs using four different methods: Nano3P-seq, PAL-seq (data from Subtelny et al.,¹⁰), FLAM-seq (data from Legnini et al.,⁴⁵), and TAIL-

seq (data from Lim et al.,⁴³). **(b)** Scatterplots of median per-gene polyA tail length estimations from *S. cerevisiae* mRNAs using three different methods: Nano3P-seq, PAL-seq (data from Subtelny et al.,¹⁰), and PAT-seq (data from Harrison et al.,⁴⁷). Each dot represents the median polyA tail length of a given gene.



Extended Data Fig. 9 | Comparison of read ends mapping to *actb1* gene before and after filtering by adenine base (A) enrichment. (a) Reads that are trimmed with porechop. **(b)** Same reads after removing incorrectly trimmed ones (filtered based on their A content). The labeled part indicates the polyA tail, which is dominantly colored in red.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- | | |
|-----------------|--|
| Data collection | Nanopore data was acquired using Minknow software v.3.5.5 . |
| Data analysis | Nanopore data was basecalled using Guppy version 4.0. TailfindR tool (Nano3P-seq mode) was used for tail length estimation. Minimap2 and BWA tools were used for mapping the basecalled nanopore reads. Code used in this study is made available in the following link : https://github.com/novoalab/Nano3P_Seq . TailfindR nano3pseq mode can be accessed using this link : https://github.com/adnaniazi/tailfindr/tree/nano3p-seq . |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Base-called FAST5 reads from Nano3P-seq and dRNA-seq libraries are publicly available in ENA, under accession code PRJEB53494. PAL-Seq polyA tail length

estimates used in this work were obtained from GEO with the accession code GSE52809. PAT-seq polyA tail length estimates used in this work were obtained from GEO with the accession code GSE53461. FLAM-seq polyA tail length estimates used in this work were obtained from GEO with the accession code GSE126465. TAIL-seq polyA tail length estimates used in this work were obtained from GEO with the accession code GSE51299.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size was chosen as n=3 for most experiments, allowing to perform statistical tests to assess significance in the observed differences between groups. Sample size in this case corresponds to biological replicates

Data exclusions

No data was excluded from our analyses.

Replication

In-vitro constructs (RNA and cDNA) were sequenced once. *S. cerevisiae* data has three biological replicates. *Mus musculus* data has two biological replicates. *Danio rerio* has three replicates for three time-points. HeLa (Human cell line) has two biological replicates.

Randomization

Randomization is not relevant to our study.

Blinding

Blinding was not relevant to our study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	HeLa-CCL-2 was purchased from ATCC.
Authentication	None of the cell lines were authenticated.
Mycoplasma contamination	Cell lines were tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified samples were used in this study.

Animals and other research organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	Mus musculus and Danio rerio were used in this study. Experiments were performed with male mice aged between 8 and 10 weeks. All mice were euthanized using CO2 and tissues were snap frozen in liquid nitrogen. Animals were kept on a 12:12h light:dark cycle and provided with water and food ad libitum. Wild-type zebrafish (Danio rerio) embryos were obtained through natural mating of the TU-AB strain of mixed ages (5–18 months). Mating pairs were randomly chosen from a pool of 60 males and 60 females allocated for each day of the month. Embryos and adult fish were maintained at 28°C.
Wild animals	No wild animals were used in this study.
Reporting on sex	Male mouse were used in this study.
Field-collected samples	No field collected samples were used in this study.
Ethics oversight	Fish lines were maintained according to the International Association for Assessment and Accreditation of Laboratory Animal Care research guidelines, and protocols were approved by the Yale University Institutional Animal Care and Use Committee (IACUC). Mice maintenance was approved by the Garvan/St Vincent's Hospital Animal Ethics Committee, in accordance with the guidelines of the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (Project No. 16/14 and 16/26).

Note that full information on the approval of the study protocol must also be provided in the manuscript.