**SUPPLEMENTARY INFORMATION**

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

Oguzhan Begik1,2,3, Huanle Liu1, Anna Delgado-Tejedor1,5, Cassandra Kontur6, Antonio J. Giraldez6, Jean-Denis Beaudoin7, John S. Mattick2,4 and Eva Maria Novoa1,5\*

*1 Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain*

*2 Garvan Institute of Medical Research, 384 Victoria St, Darlinghurst, 2010, NSW, Australia.*

*3 Faculty of Medicine, UNSW Sydney, Sydney NSW 2052, Australia.*

*4 School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Sydney NSW 2052, Australia.*

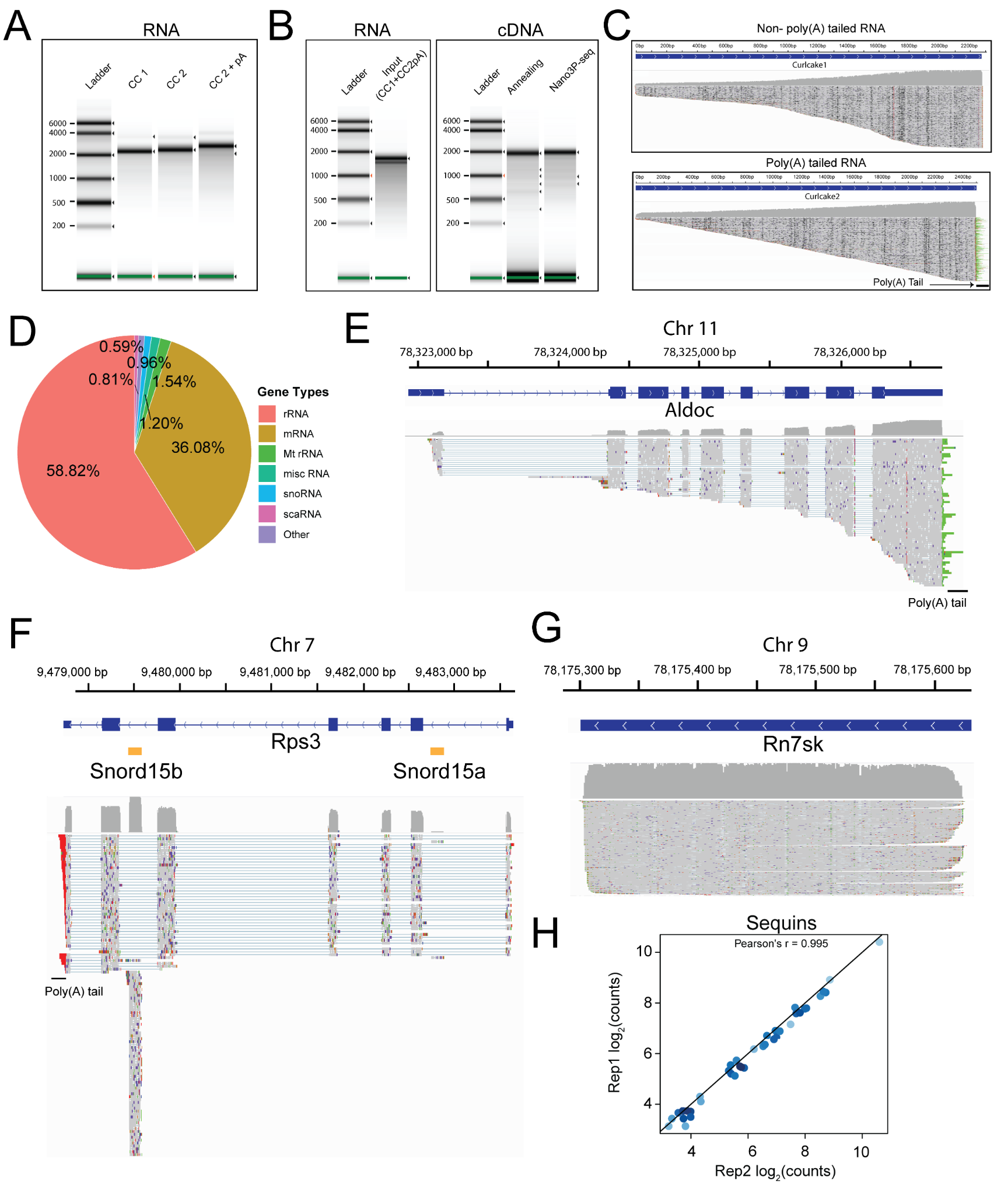
*5 Universitat Pompeu Fabra, Barcelona, Spain.*

*6 Department of Genetics, Yale University, USA*

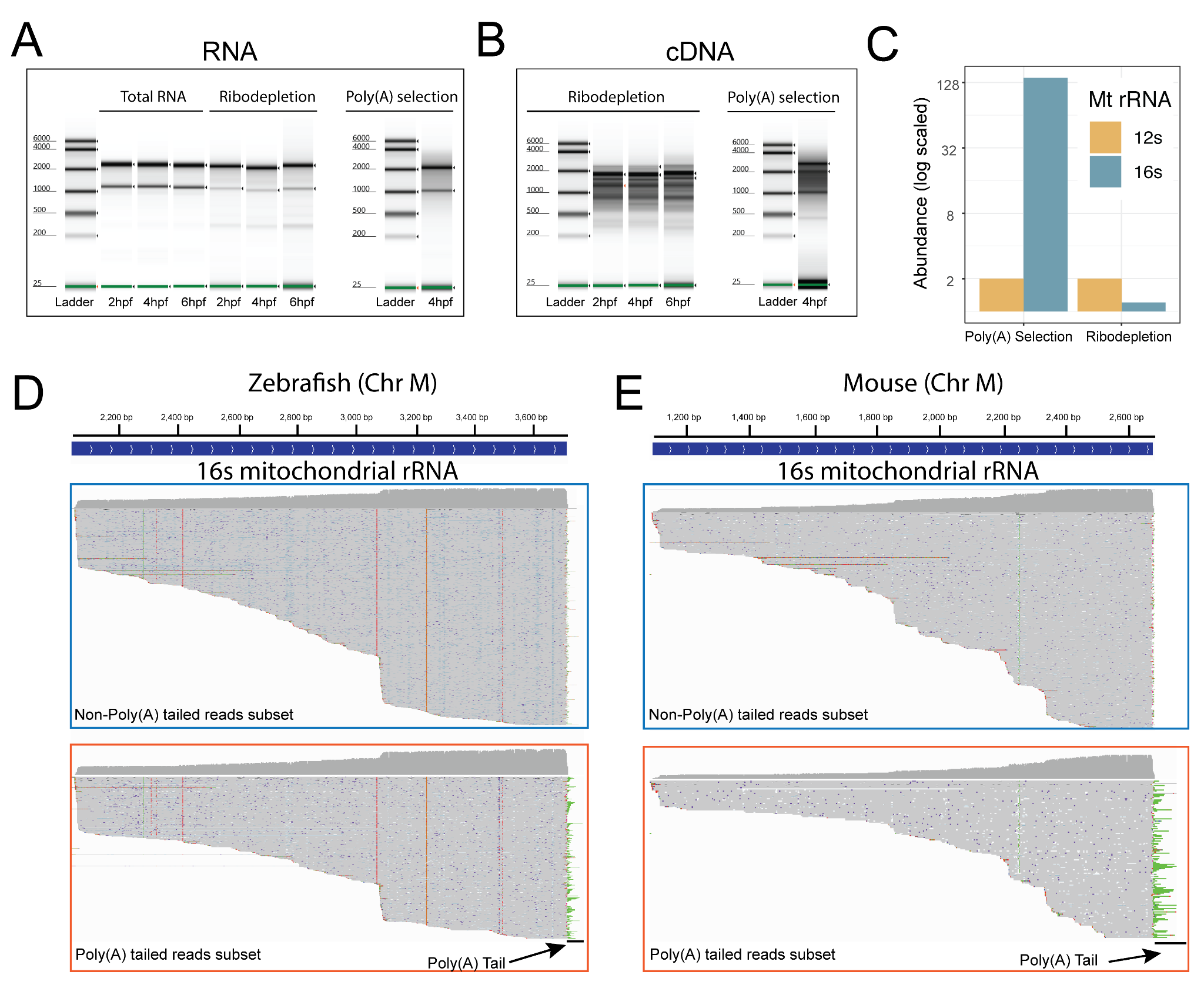
*7 Department of Genetics and Genome Sciences, Institute for Systems Genomics, University of Connecticut Health Center, Farmington, CT, USA.*

\* Correspondence to: Eva Maria Novoa (eva.novoa@crg.eu)

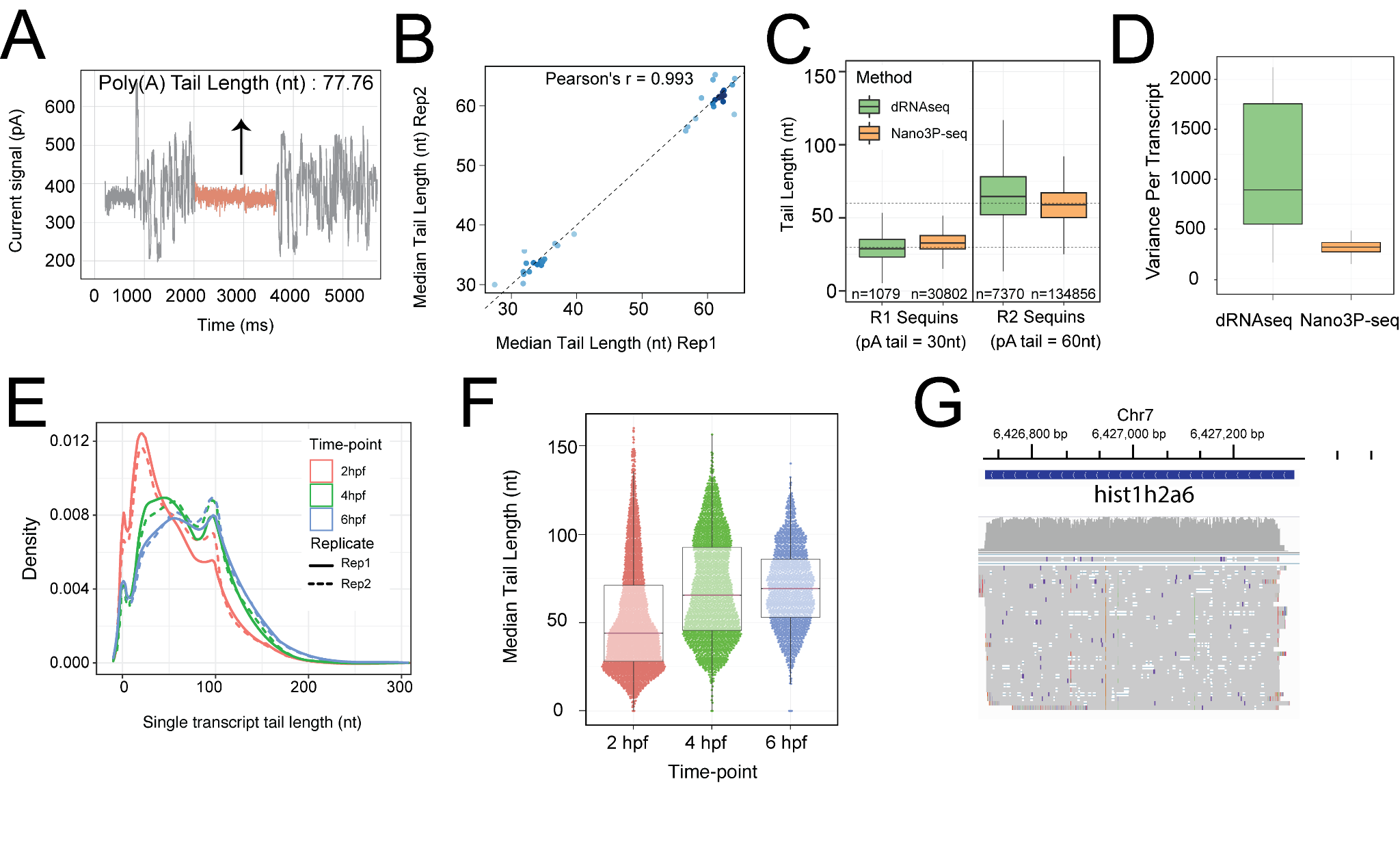
**Figure S1. Nano3P-seq captures non-poly(A)-tailed and poly(A)-tailed RNAs (A)** Tapestation profiles of synthetic RNAs (‘curlcakes’) after being *in-vitro* transcribed and poly(A) tailed (pA). **(B)** Tapestation profiles of the input RNA (curlcake mix) for reverse-transcription and cDNA produced after annealing based or template-switching based (Nano3P-seq) reverse-transcription **(C)** IGV snapshots of synthetic RNAs illustrating that Nano3P-seq captures both non-polyadenylated (left) and polyadenylated (right) RNAs. In addition, a diversity of poly(A) tail lengths are also captured by Nano3P-seq, which are shown in green (right panel).**(D)** Pie chart showing the abundance of different RNA types in Nano3P-seq of nuclear/mitochondria enriched RNA. **(E)** IGV snapshot of reads mapping to Aldoc gene with poly(A) tail shown in green. **(F)** IGV snapshot of reads mapping to Rps3 and Snord15b genes. Poly(A) 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 poly(A) tails. **(H)** Correlation ofper-gene counts observed in synthetic sequins that were used as spike-ins in the sequencing runs.

****

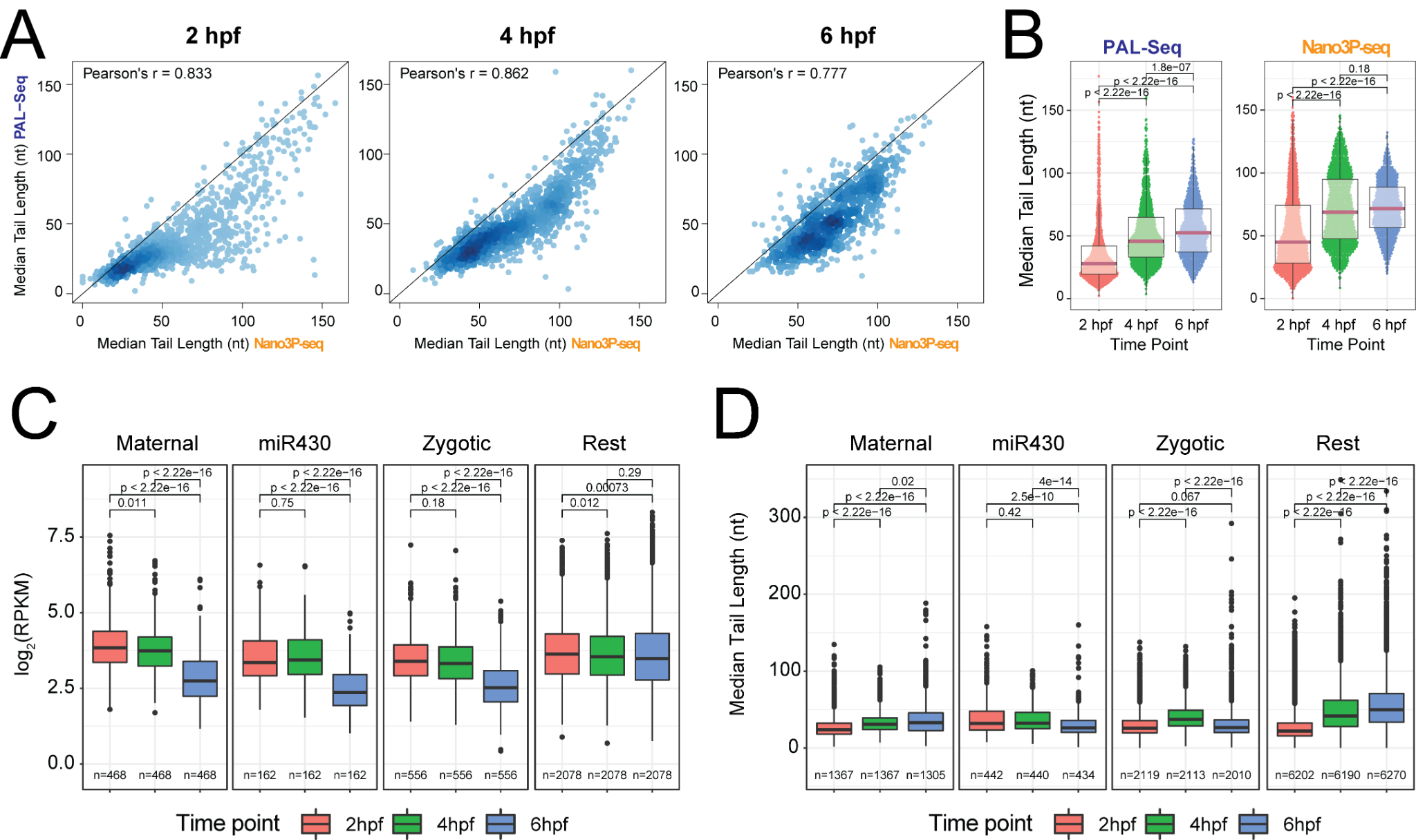
**Figure S2. Analysis of abundances and poly(A) tails in mitochondrial rRNAs. (A)** Tapestation profiles of RNAs from zebrafish embryos at different developmental time points (2,4,6 hours post-fertilization). Profiles include total RNA, ribodepleted RNA and poly(A)+ selected RNA. **(B)** Tapestation profiles of the reverse-transcription products of ribodepleted (left) and poly(A)+ selected (right) samples from zebrafish embryos collected at different developmental time points (2,4,6 hours post-fertilization). **(C)** mRNA abundances (log scaled) of 12s and 16s mitochondrial rRNAs in poly(A)+ selected (left) and ribodepleted (right) samples. **(D)** IGV snapshot of reads mapping to zebrafish 16s mitochondrial rRNA, where reads have been grouped as non-poly(A) tailed and poly(A) tailed based on their predicted poly(A) tail length. Poly(A) tail region is shown with an arrow. **(E)** IGV snapshot of reads mapping to mouse 16s mitochondrial rRNA, where reads have been grouped as non-poly(A) tailed and poly(A) tailed based on their predicted poly(A) tail length. Poly(A) tail region is shown with an arrow.

****

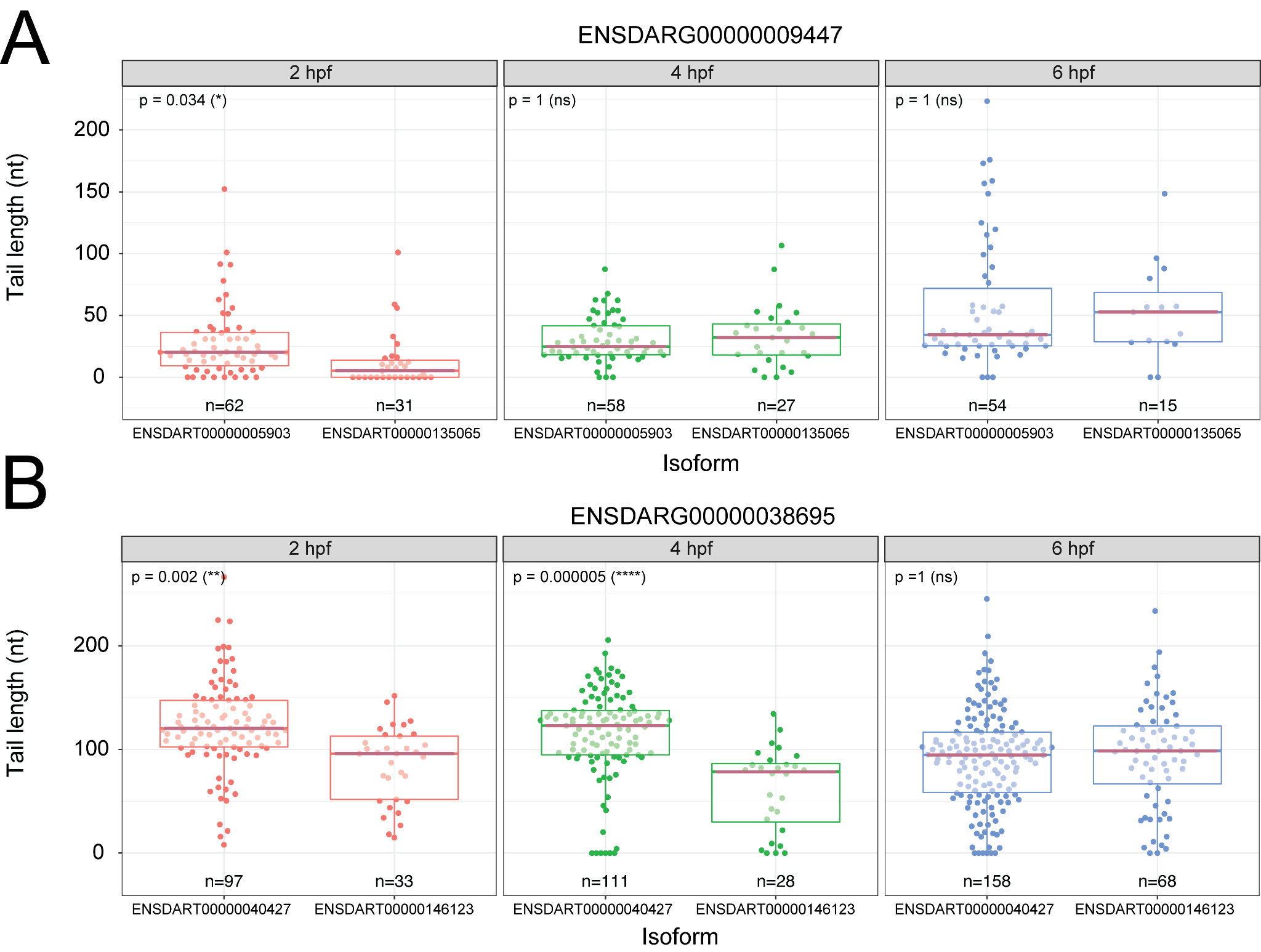
**Figure S3. Analysis of poly(A) tail lengths using Nano3P-seq.** **(A)** Current intensity plot of a synthetic poly(A)+ read, obtained using Nano3P-seq. The homopolymeric poly(T) region is highlighted in orange. **(B)** Replicability ofmedian per-genepoly(A) tail length estimation in sequins captured with Nano3P-seq. The poly(A) tail length of synthetic sequins is 30nt (R1\_sequins) or 60nt (R2\_sequins). **(C)** Overall comparison of poly(A) tail length estimation of R1 and R2 sequins which contain 30 nt and 60 nt poly(A) tail lengths, respectively, obtained using dRNAseq (green) and Nano3P-seq (orange). **(D)** Per-transcript variance of poly(A) tail length estimations of sequins obtained using dRNAseq (green) and Nano3P-seq (orange). **(E)** Distribution of poly(A) tail lengths in mRNAs across zebrafish developmental stages (2, 4 and 6hpf, shown in red, green and blue respectively) in two biological replicates (shown as full lines and dashed lines, respectively). **(F)** Median per-genepoly(A) tail length distribution of mRNAs during the zebrafish MZT (t=2, 4 and 6hpf). **(G)** IGV snapshot of reads mapping to hist1h2a6 mRNA, which do not contain poly(A) tails.

****

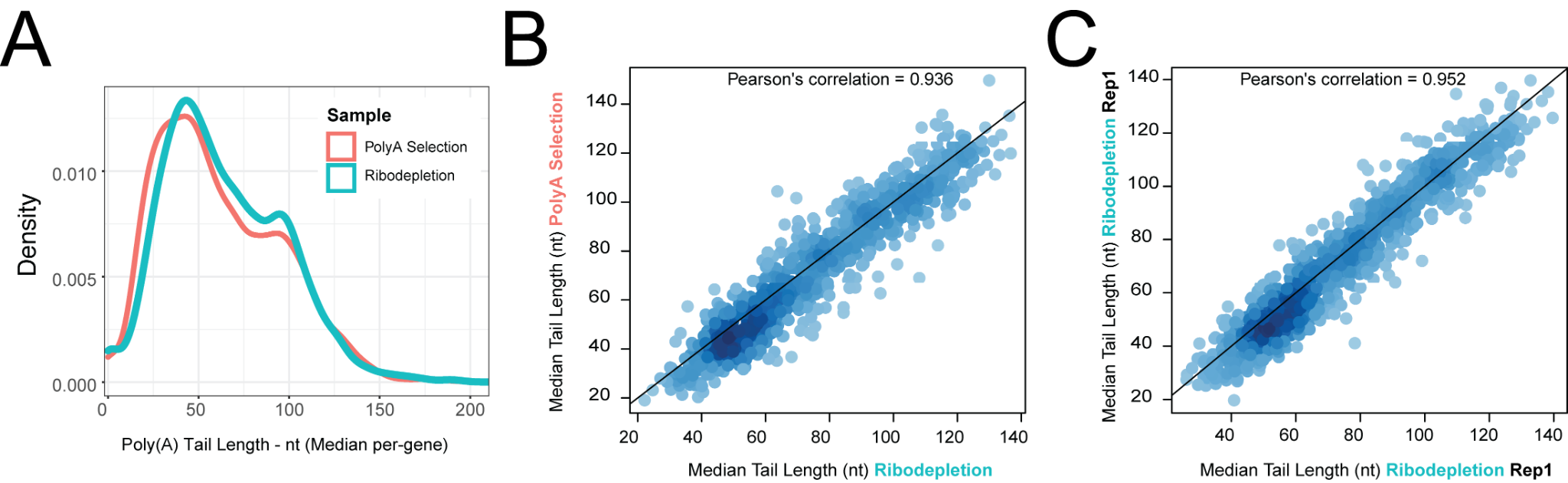
**Figure S4. Comparison of poly(A) tail length estimations using PAL-Seq and Nano3P-seq. (A)** Scatterplots of per-gene poly(A) tail length estimations using Nano3P-seq and PAL-seq from zebrafish mRNAs at 2 hpf (left), 4 hpf (middle) and 6 hpf (right). Each dot represents the median poly(A) tail length of a given gene. **(B)** Boxplots depicting the distribution of poly(A) tail lengths during the zebrafish MZT, estimated using PAL-Seq (left) or Nano3P-seq (right). Statistical comparison of means was made using Kruskal Wallis test. **(C)** Comparative analysis of the mRNA abundance (shown as log2 RPKM) for the 4 groups of zebrafish mRNAs (maternal, miR-430, zygotic and rest) during embryogenesis (t= 2, 4 and 6 hpf) using PAL-seq data. Statistical comparison of means was performed using Kruskal-Wallis test. **(D)** Median tail length estimations of the 4 groups of zebrafish mRNAs (maternal, miR-430, zygotic and rest) at 2, 4 and 6 hpf using PAL-seq data. Statistical analyses were performed using Kruskal-Wallis test.

****

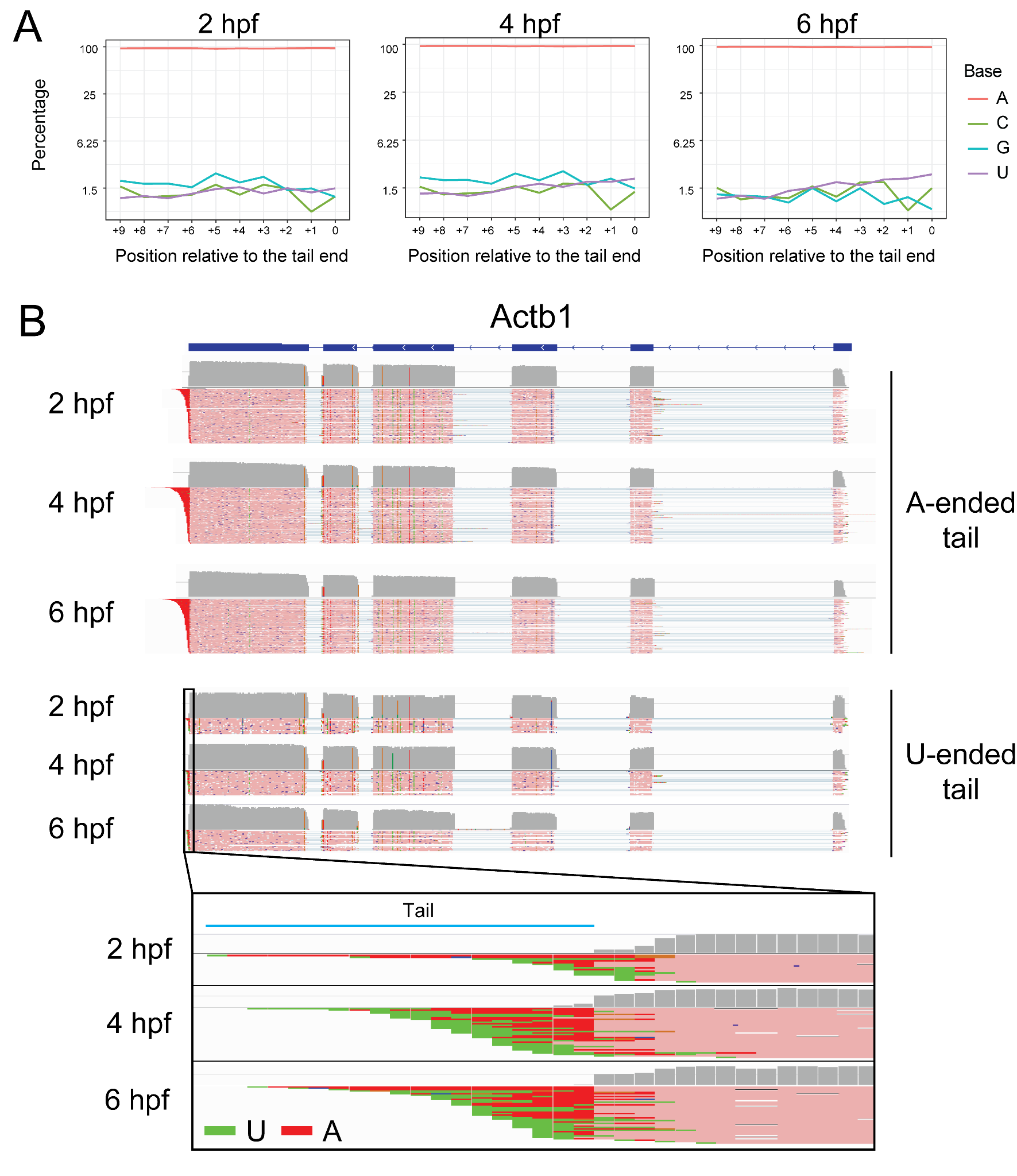
**Figure S5. Analysis of Isoform-specific poly(A) tail dynamics using Nano3P-seq.** Examples of genes with differentially polyadenylated isoforms between 2 and 6 hpf.

****

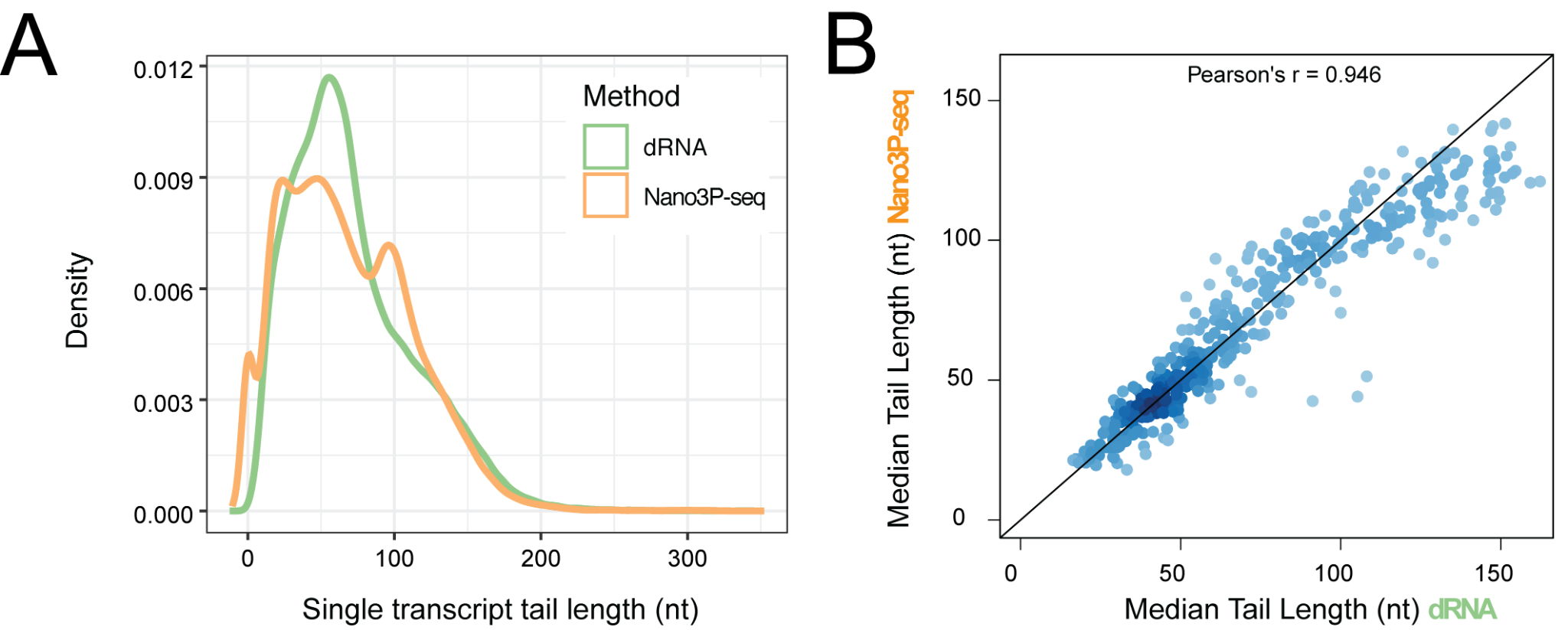
**Figure S6. Comparison of poly(A) tail length estimations using poly(A)-selected and ribodepleted samples. (A)** Distribution of per-genemRNApoly(A) tail lengths from 4 hpf zebrafish embryos, isolated using either poly(A) selection (red) or ribo-depletion (cyan). **(B)** Comparison of median per-genepoly(A) tail length estimation between poly(A) selected and ribo-depleted zebrafish mRNAs isolated at 4 hpf. Each dot represents a gene. **(C)** Comparison of median per-genepoly(A) tail length estimation of mRNAs in zebrafish ribo-depleted samples (replicate 1 and 2) isolated at 4 hpf. Each dot represents a gene.

****

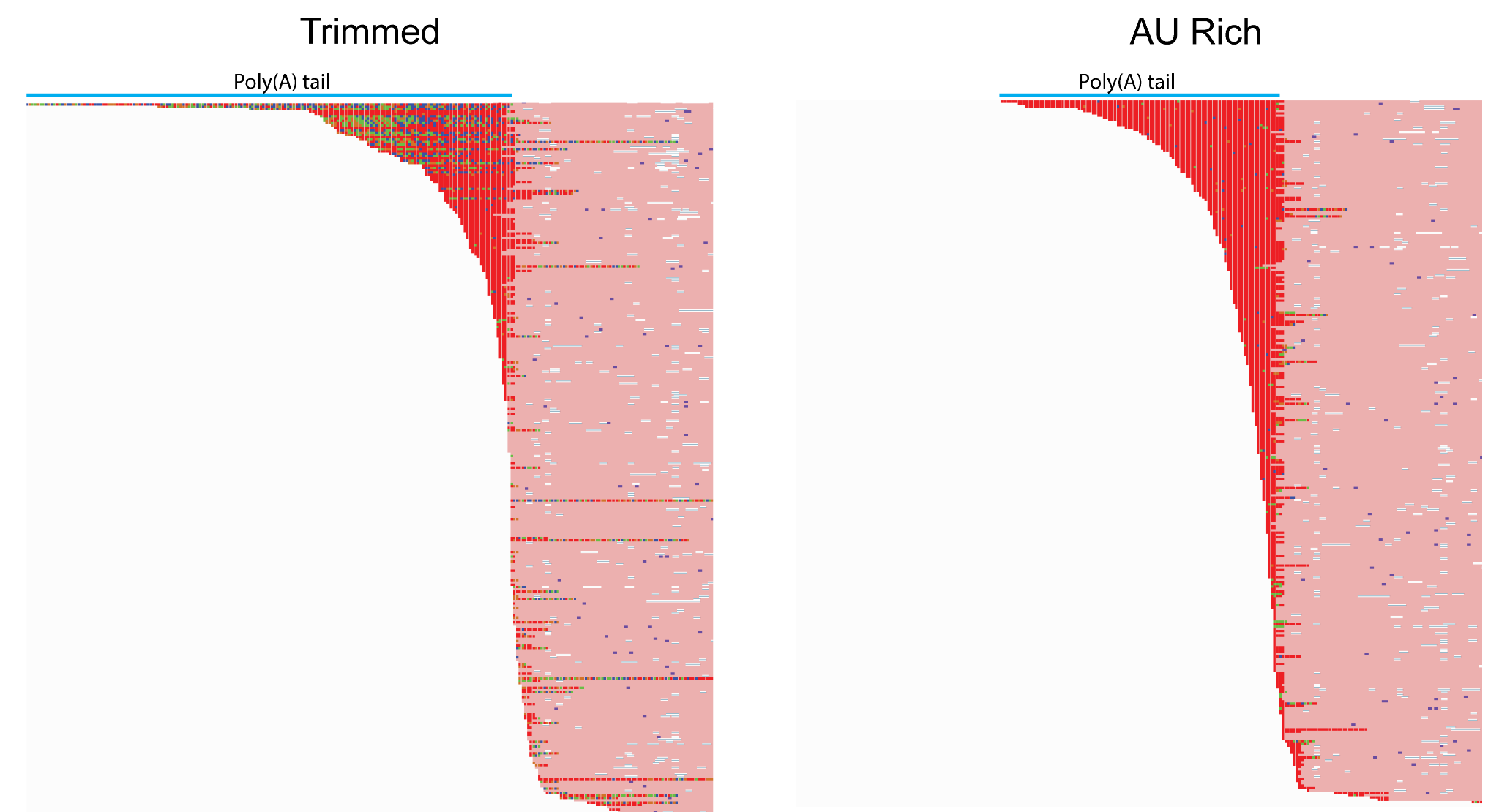
**Figure S7. Analysis of tail composition using Nano3P-seq. (A)** Tail composition of the last 10 nucleotides of the reads enriched for A and U bases. The last nucleotide is labeled as 0. The Y axis is log scaled. **(B)** IGV snapshots of the reads mapping to *actb1* gene in zebrafish samples (t=2, 4 and 6hpf). Reads ending with A nucleotide are grouped as “A-ended tail”, whereas reads ending with U nucleotide are grouped as “U-ended tail”.

****

**Figure S8. Comparison of poly(A) tail length estimations between dRNAseq and Nano3P-seq. (A)** Distribution of per-genemRNApoly(A) tail lengths from 4 hpf zebrafish embryos isolated using poly(A) selection and sequenced with dRNAseq (green) or Nano3P-seq (orange). **(B)** Comparison of median per-genepoly(A) tail length estimation of poly(A)-selected mRNAs isolated at 4 hpf with dRNAseq (green) or Nano3P-seq (orange). Each dot represents an mRNA.

****

**Figure S9. Comparison of read ends mapping to *actb1* gene before and after filtering by A/U enrichment.** Left panel contains the reads that are trimmed with porechop. Right panel shows the same reads after removing incorrectly trimmed ones (filtered based on their A/U content). Labeled part indicates the poly(A) tail.

****