Appendix - Table of contents

3′ RACE

Generic RNA-Seq RNA adapters and DNA oligonucleotides are listed here:

https://support.illumina.com/downloads/illumina-customer-sequence-letter.html

Mitochondrial RNA-specific primers are listed below.

RNA preparation and cDNA synthesis. Total RNA (6 µg) was treated with 20 units of CIAP (Calf Intestinal Alkaline Phosphatase) for 30 minutes at 37 °C. After phenol/chloroform extraction and ethanol precipitation, 5 µg of the CIAP-treated RNA was mixed with 100 pmol of RA3 monophosphorylated RNA adapter, denatured for 2 minutes at 70 °C and placed on ice for 2 minutes. RNA was combined with 40 units of RNase OUT (Invitrogen) and 15 units of T4 RNA ligase 1 (NEB) in 50 µL of 1X buffer T4 RNA ligase with 1 mM ATP, and incubated at 16 ºC overnight. After phenol/chloroform extraction and ethanol precipitation, ½ of RNA was used for the reverse transcription reaction with 40 pmol of RTP primer using SuperScript III Reverse Transcriptase per manufacturer's protocol in 20 µl reaction. Upon completion of cDNA synthesis, 1 unit of ribonuclease H and was added and incubated for further 20 minutes at 37°C.

Library construction. First PCR reaction was performed in a 100 µL of HF buffer, dNTPs (0.2 mM of each), index primer (0.6 µM), primers C654 to C680 (0.02 µM of each), RNase A (Qiagen, 1 µg) and Phusion Hot Start II High Fidelity DNA polymerase (ThermoScientific, 2 units) using the following amplification program:

Initial denaturation: 1 minute at 98°C

98°C - 15 seconds

50°C - 30 seconds

72°C for 15 seconds

5 cycles

98°C - 15 seconds

60°C - 30 seconds

72°C - 15 seconds

5 cycles

Final elongation: 72°C for 5 min

PCR products were purified using the DNA Clean and Concentrator 5 kit (Zymo) and eluted in 20 µL of elution buffer.

A second PCR reaction was set-up in a 50 µL final volume with 2 µL of purified PCR product, HF buffer, 0.2 mM dNTPs, PR1 primer (0.6 µM), index primer (0.6 µM) and Phusion Hot Start II High Fidelity DNA polymerase (ThermoScientific, 1 units) using the following amplification program:

Initial denaturation: 1 minute at 98°C

98°C - 15 seconds

60°C - 30 seconds

72°C - 15 seconds

16 cycles

Final elongation: 72°C for 5 min

PCR products were purified using the DNA Clean and Concentrator 5 kit (Zymo) and eluted in 8 µL of elution buffer.

Library purification. PCR products were separated on a 20-cm long, 7.5% non-denaturing acrylamide gel (29:1 acrylamide/Bis-acrylamide, 1xTBE buffer) along with 5 µg of Quick-Load 100 bp DNA Ladder (NEB). The gel was stained with SYBR Green I Nucleic Acid Gel Stain and the areas of interest (130 bp to 1 kbp) were cut out under Safe Imager 2, and DNA was extracted into 1 mL of 0.5 M ammonium acetate, 0.1% SDS, 1 mM EDTA overnight at room temperature on a Nutator. Supernatant was collected and concentrated 5-6 times with 2-butanol, and purified DNA was precipitated with ethanol. DNA pellets were resuspended in 50 µL of water and re-purified using the DNA Clean and Concentrator 5 kit (Zymo), and eluted in 20 µL of elution buffer. Purified library was sequenced either on HiSeq 2500 platform (single-end, 100 bp) or MiSeq platform (paired-end, 300 bp).

CLAP-Seq data processing

The single-end stranded RNA CLAP-Seq raw reads were barcode-separated and adaptor sequences (5'-TGGAATTCTCGGGTGCCAAGG-3') were removed by using Trim-galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with the default setting. To eliminate nuclear contamination, reads that aligned to the nuclear genome (http://tritrypdb.org, release 5.0) by the short read aligner BWA (version 0.7.5) (Li & Durbin, 2009) were eliminated. The filtered reads then were aligned to mitochondrial maxicircle sequences by BWA. The coverage of CLAP-Seq reads on maxicircle were generated by using genomeCoverageBed tool in BEDTools Suite (Quinlan & Hall, 2010).

Composite plot for CLAP-Seq relative to transcript end

Instead of aligning reads onto maxicircle, the trimmed and filtered reads described above were aligned to 20 maxicircle individual transcripts with a 100nt extension downstream of transcript end on the maxicircle. The read coverage on each transcript was calculated using

genomeCoverageBed tool in BEDTools Suites. Then a custom Python script was developed to calculate the average read coverage per base pair when aligned all the transcripts by their transcript end sites from -1000nt to +100nt.

Tail length distribution in KPAF3 CLAP-Seq data

The trimmed and filtered reads described above were aligned to 20 maxicircle individual transcript with a 100nt extension downstream of transcript end on the maxicircle. For each read, if there exists a 3′ fragment that is not aligned to the transcript based on the CIGAR string in the alignment SAM file, a custom Python script was used to extract the 3′ most unaligned sequences. These sequences were considered as tail sequences bound or protected by KPAF3. The length distribution of all such tail sequences was summarized into a frequency distribution plot by a custom Python script.

Whole transcriptome 3′ mRNA-Seq data processing

The adaptor sequences 5'-TGGAATTCTCGGGTGCCAAGG-3' were removed from single-end stranded 300 nt mRNA-Seq raw reads by using Trim-galore with default setting and trimmed reads shorter than 50 nt were also removed. The trimmed reads were aligned to maxicircle and 8 mature sequences of fully edited transcripts with a 100 nt extension after their transcript end from maxicircle sequences by BWA. To extract the tail sequences and locate the exact transcript end site, we selected the alignment which yields a shorter tail sequence from multiple alignments. We then tried to assign the tail type for each 3' tail sequences. If the tail sequences contain >90% U, we define this tail as a U tail. If after removing the longest U stretches in the 5' end, the rest of the tail sequences contain >=90% A, then we define this tail as an A tail. If there is no overhang 3' unaligned sequences, then this transcript belongs to "unmodified" type. And the rest of tail sequences are considered as "other" tail type. For the "other" tail sequences we adopted another round of no-U alignment strategy described in previous literature (Simpson et al., 2016), we removed all U from reads and aligned it to the U-removed 8 fully edited transcripts by Blat(Kent, 2002). If the tail length is shorter based on the no-U alignment strategy compared with U containing alignment, the 3′ end location and tail sequences were adjusted based on No-U alignment by custom Python script.

Gene-specific 3' mRNA-Seq data processing

We first merged the paired end RNA-Seq data with PEAR (Zhang et al., 2014). Only the merged pair-end reads were carried into the next step. The adaptor sequences 5'-TGGAATTCTCGGGTGCCAAGG-3' were removed from merged single read by using Trim-galore with default setting and trimmed reads shorter than 50 nt were also removed. The rest of procedures were identical to whole transcriptome 3′ mRNA-Seq data Processing as described above.

KPAF3 binding motif analysis

Peak calling on KPAF3 CLAP-Seq was performed on forward and reverse strand of maxicircle separately by MACS (Zhang et al., 2008). Then we only selected the peaks that were detected in both high and low RNase UV+ sample. Then we extended the peak region on both ends to 100nt,

and performed motif detection using MEME algorithm on the stranded mode, with the nucleotide frequency in maxicircle as the background nucleotide frequency(Bailey & Elkan, 1994). We also queried the motif against edited and pre-edited maxicircle transcripts using FIMO with a P value cutoff at 0.001(Grant et al., 2011).

Per base nucleotide composition plot

FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) analysis was performed on KPAF3 trimmed fastq reads.

References

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings. International Conference on Intelligent Systems for Molecular Biology 2: 28-36
Grant CE, Bailey TL, Noble WS (2011) FIMO: scanning for occurrences of a given motif. Bioinformatics 27: 1017-8
Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res 12: 656-64
Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-60
Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-2
Simpson RM, Bruno AE, Bard JE, Buck MJ, Read LK (2016) High-throughput sequencing of partially edited trypanosome mRNAs reveals barriers to editing progression and evidence for alternative editing. Rna 22: 677-95
Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30: 614-20
Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9: R137

Appendix Table S1. DNA Constructs and T7 Transcription

Name	Sequence, 5′-3′	Description
B876	GGTAATACGACTCACTATAGGGAACCCTTTGTTTTGGTTAAAGAAACATCGTTTAGAAGAGATTTTAGAATAAGATATGTTTTTAATATTTTTTTTA	T7 transcription RPS12 pre-cursor sense
B877	ATAATCAAAAACTTAGAAAATTCCTACCAAACATAAATGAACCTGATATAAACCCAAACATTATAAAAAATAAAAAAAATATTAAAAACATATCTTATTC	T7 transcription RPS12 pre-cursor antisense
A717	TGGAAGCTTAACGCGTAAAGGAATTGGTGGAGGAAC	Sense, KPAF3 RNAi
A718	AGGTCTAGAACTTTCCAGCCTTGAGCAAAAGGGTAC	Antisense, KPAF3 RNAi
B85	CGCAAGCTTATGAAACTACTGCGACGAGGGA	Sense, KPAF3 MHTAP
B86	GCGTCTAGAATGTGCACGATCATCTGCACCCA	Antisense, KPAF3 MHTAP
B72	GCACATATGATGAAACTACTGCGACGAGGG	Sense, KPAF3 in pET28
B73	GGCAAGCTTTTAATGTGCACGATCATCTGCAC	Antisense, KPAF3 in pET28

3′ RACE primers
P: pre-edited, E: edited.

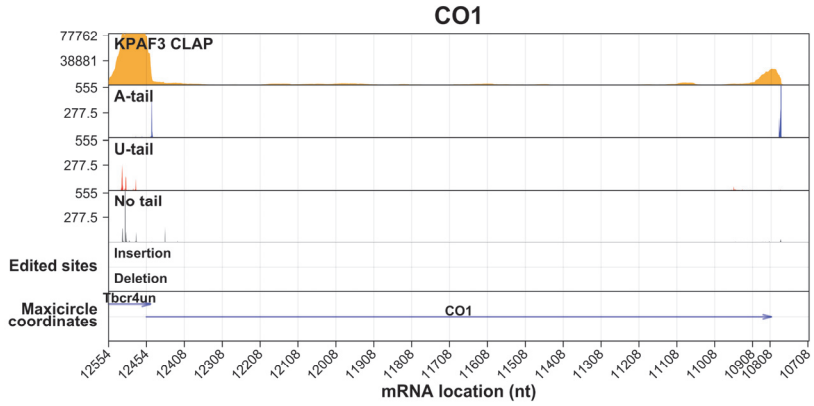
Name	Sequence, 5′-3′	Description
C652	GTTCAGAGTTCTACAGTCCGACGATCCTACTTATTACCATGATTGATTGTTC	3RACE 12S
C653	GTTCAGAGTTCTACAGTCCGACGATCGGATTATAAATTGAAAGTGGTAATATC	3RACE 9S
C654	GTTCAGAGTTCTACAGTCCGACGATCATTTTTGCCAACGCATTC	3RACE ND8p
C655	GTTCAGAGTTCTACAGTCCGACGATCGTAGGGAAGAAACATCGAGG	3RACE ND9p
C656	GTTCAGAGTTCTACAGTCCGACGATCCATATCTGGCATTTTAATTTGAC	3RACE MURF5
C657	GTTCAGAGTTCTACAGTCCGACGATCACCCCAAAGGGATTTGAG	3RACE ND7p
C658	GTTCAGAGTTCTACAGTCCGACGATCAAGGAATACAATTTGCAGAGG	3RACE CO3p
C659	GTTCAGAGTTCTACAGTCCGACGATCTTTATAGTATATGAATGAGTGG	3RACE CYB
C660	GTTCAGAGTTCTACAGTCCGACGATCATGGAATTGGGAATTGCC	3RACE A6p

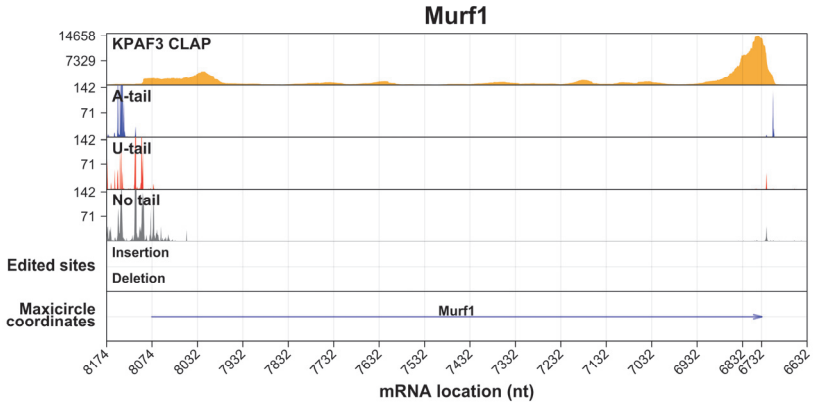
C661	GTTCAGAGTTCTACAGTCCGACGATCTTTCTTATCAATTTCTTGG	3RACE MURF1
C662	GTTCAGAGTTCTACAGTCCGACGATCGGGTTTAGGGACAGAGGG	3RACE CR3p
C663	GTTCAGAGTTCTACAGTCCGACGATCCTTATACCACGTGTTATATGC	3RACE ND1
C664	GTTCAGAGTTCTACAGTCCGACGATCGTAGAGAACCTGGTAGGTG	3RACE CO2
C665	GTTCAGAGTTCTACAGTCCGACGATCTATTTTGATATATTCGGATC	3RACE MURF2p
C666	GTTCAGAGTTCTACAGTCCGACGATCATGAGTACCAGTTTGTATGG	3RACE CO1
C667	GTTCAGAGTTCTACAGTCCGACGATCGGTTTTAGTTGGGGAGAAAG	3RACE CR4p
C668	GTTCAGAGTTCTACAGTCCGACGATCTATTCATTTGTATGGTTGG	3RACE ND4
C669	GTTCAGAGTTCTACAGTCCGACGATCATGGGAGATGGGTTTTGG	3RACE ND3p
C670	GTTCAGAGTTCTACAGTCCGACGATCGAACCGACGGAGAGCTT	3RACE RPS12p
C671	GTTCAGAGTTCTACAGTCCGACGATCGATTGCATAATGTTGTTTTGG	3RACE ND5
C672	GTTCAGAGTTCTACAGTCCGACGATCTATTTATTGGCGCCCAAG	3RACE ND8e
C673	GTTCAGAGTTCTACAGTCCGACGATCCCATTATTTGGTTTGTTTGTATTG	3RACE ND9e
C674	GTTCAGAGTTCTACAGTCCGACGATCGGATGTTTGTTTGCGTGG	3RACE ND7e
C675	GTTCAGAGTTCTACAGTCCGACGATCTTGTGTATGGATACACGTTTTG	3RACE CO3e
C676	GTTCAGAGTTCTACAGTCCGACGATCTGATTTTGCAGTTGATAATGG	3RACE A6e
C677	GTTCAGAGTTCTACAGTCCGACGATCAATATGGGTTTATTGTTGTGTTTA	3RACE CR3e
C678	GTTCAGAGTTCTACAGTCCGACGATCTGTTTTTTGTTTTGGGTGG	3RACE CR4e
C679	GTTCAGAGTTCTACAGTCCGACGATCTGTATGTATAGGATTTGTGTGG	3RACE ND3e
C680	GTTCAGAGTTCTACAGTCCGACGATCTTTTGTTTGGATGTTGCG	3RACE RPS12e
B863	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA	forward primer for library construction
B864	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	reverse primer for library construction

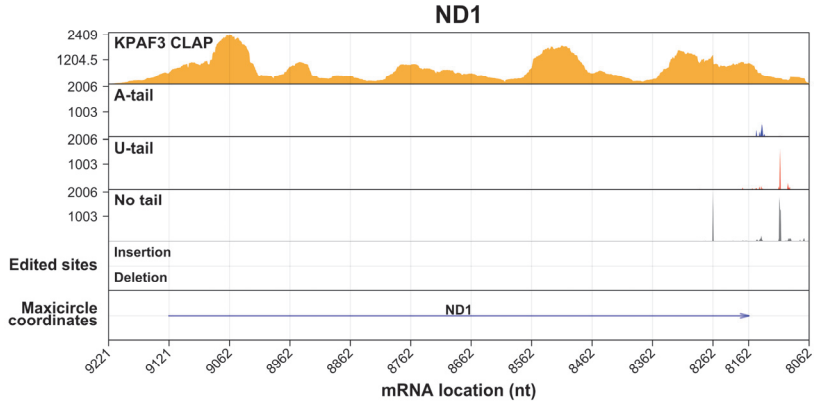
Appendix Figure S1. KPAF3 binding correlates with downstream adenylation events.

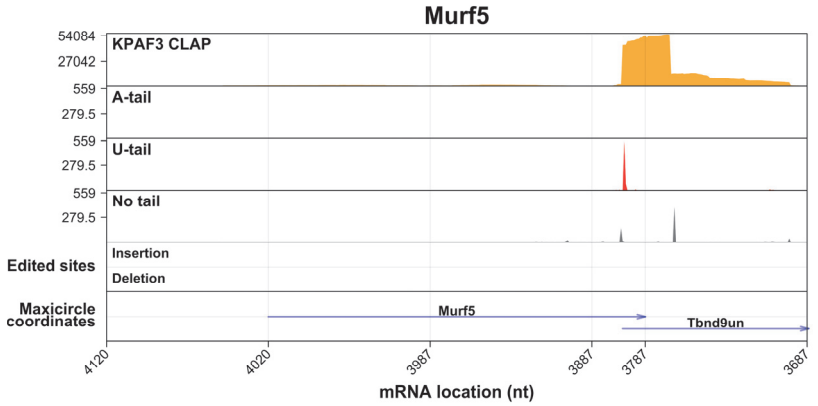
Adenylated, uridylated and unmodified 3′ ends of unedited mitochondrial mRNAs were mapped along with KPAF3 CLAP-Seq reads onto maxicircle DNA and fully-edited mRNA sequences where applicable. Gene positions are shown by arrows.

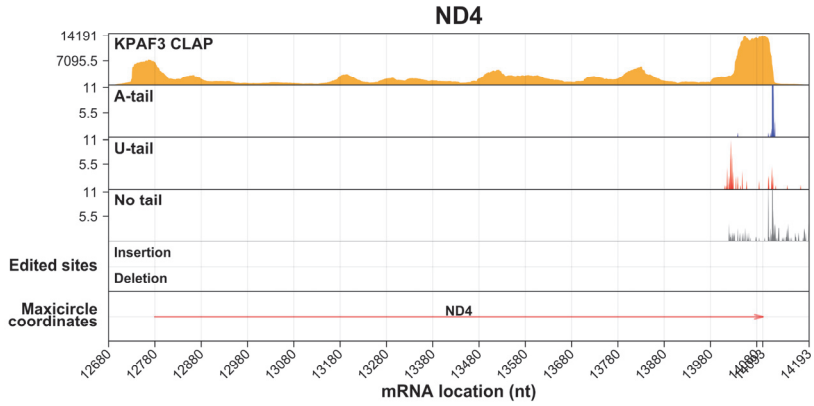
Unedited mRNAs

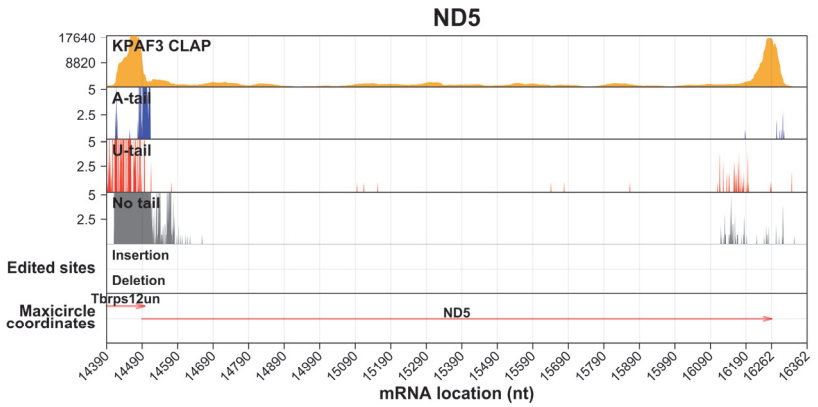




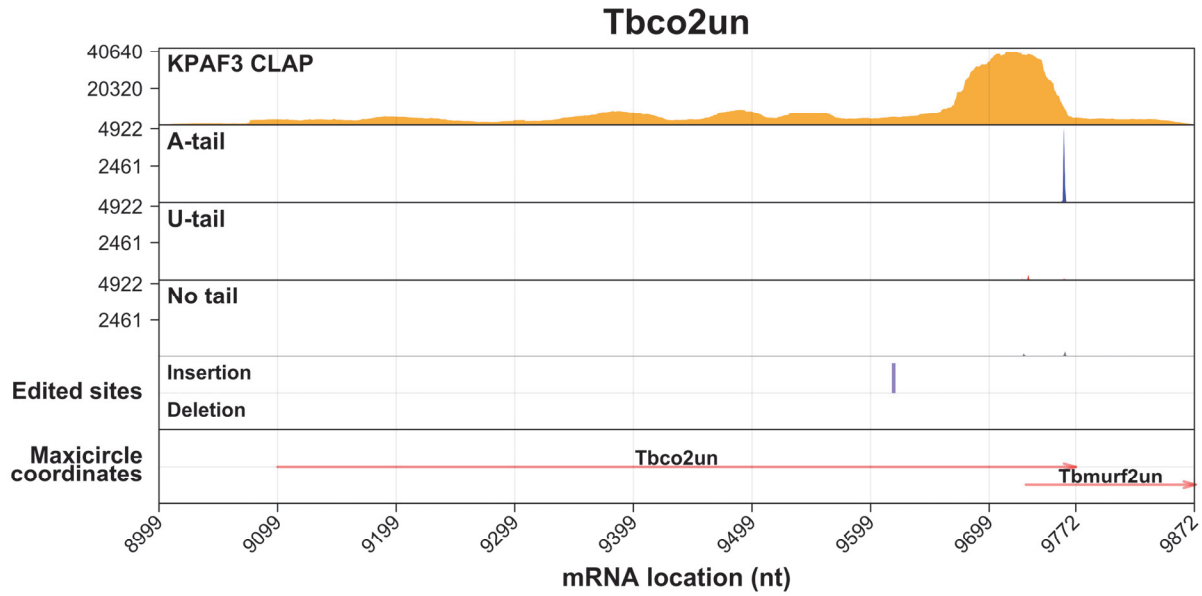


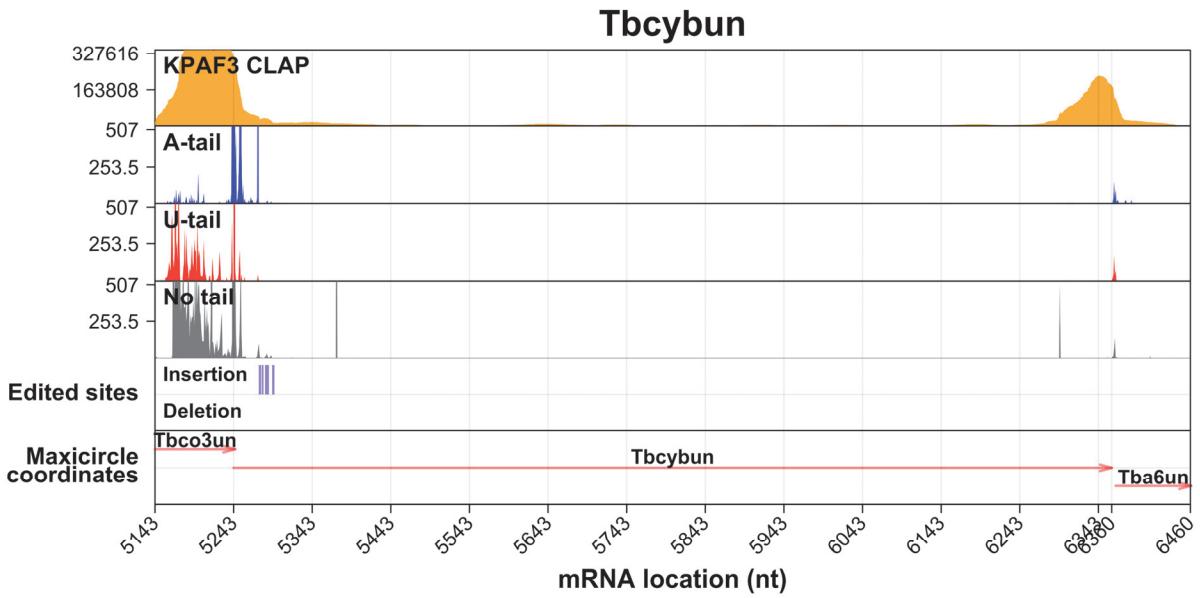


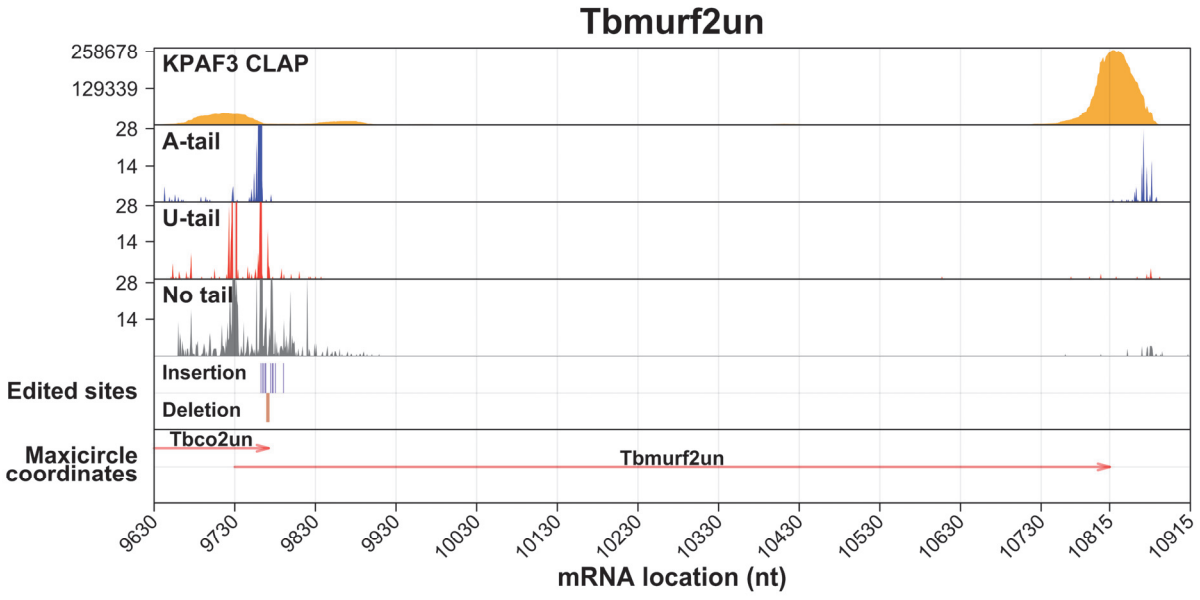




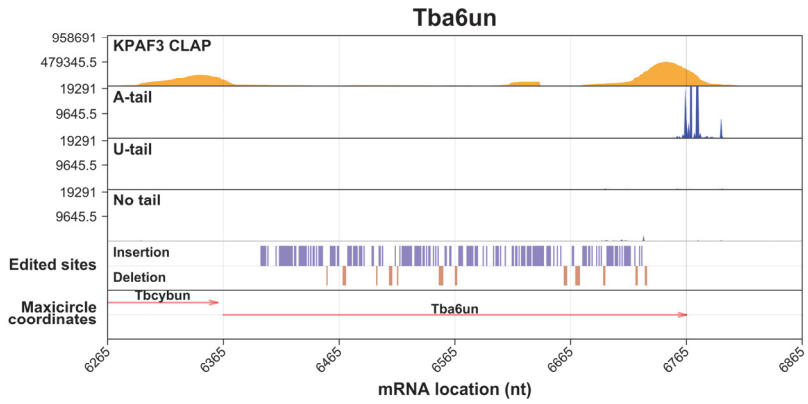
Moderately-edited mRNAs

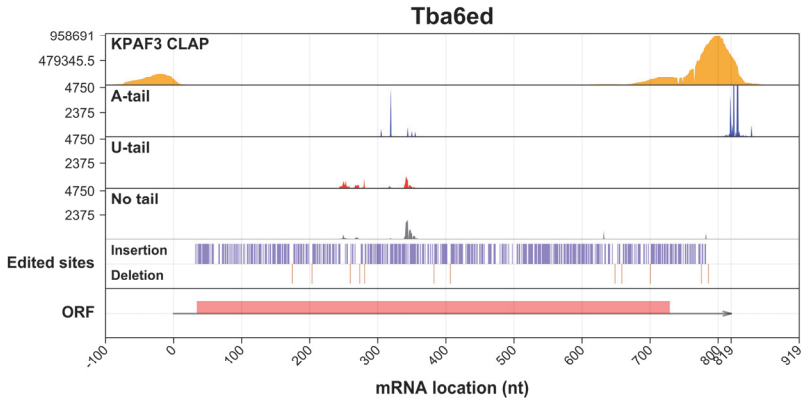


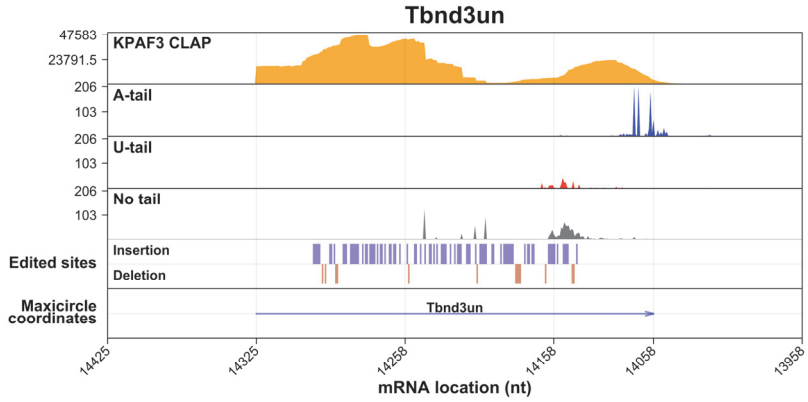


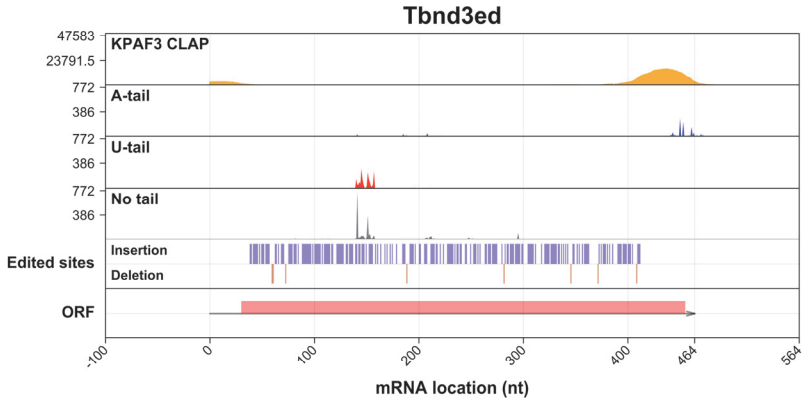


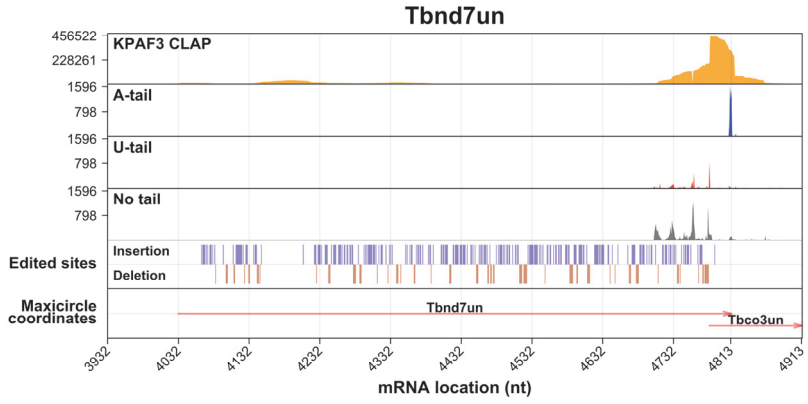
Pan-edited mRNAs

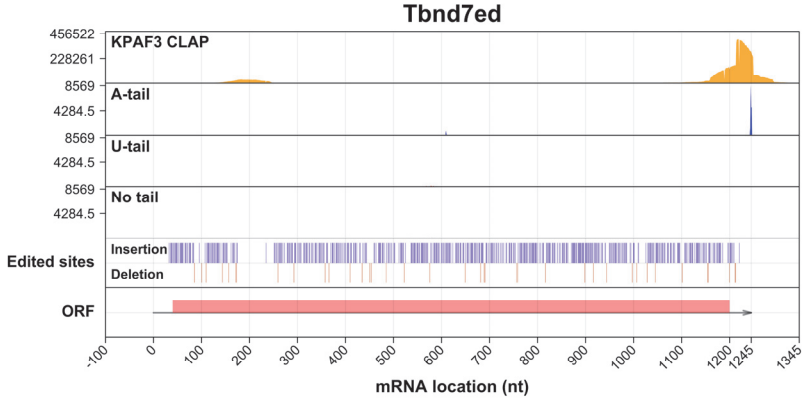


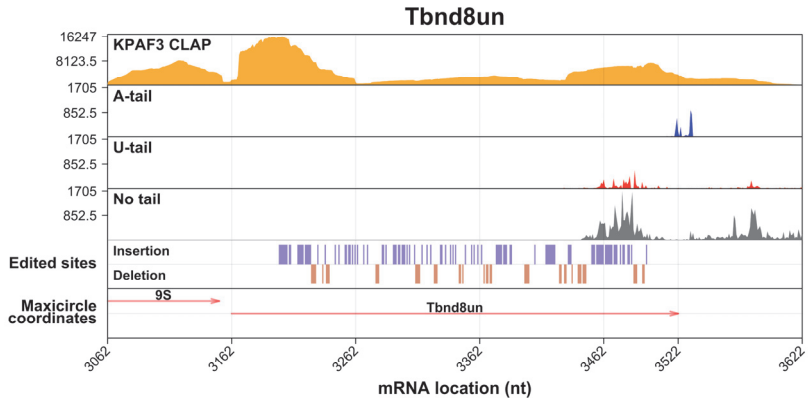


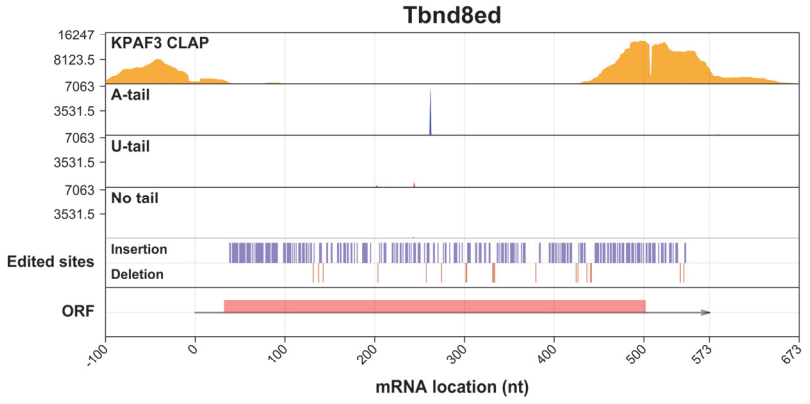












Pan-edited mRNAs