Additional File 1 for:

The *C. elegans* 3′-UTRome V2: an updated genomic resource to study 3′-UTR biology

Steber HS 1, 2, Gallante C3, O’Brien S2, Chiu P.-L4, Mangone M*1, 2.

1Molecular and Cellular Biology Graduate Program, School of Life Sciences 427 East Tyler Mall Tempe, AZ 85287 4501.

2Virginia G. Piper Center for Personalized Diagnostics, The Biodesign Institute at Arizona State University, 1001 S McAllister Ave, Tempe, AZ, USA

3Barrett, The Honors College, Arizona State University, 751 E Lemon Mall, Tempe, AZ 85281

4Center for Applied Structural Discovery, The Biodesign Institute at Arizona State University, 1001 S McAllister Ave, Tempe, AZ, USA

*To whom correspondence should be addressed. Tel: +1(480) 965-7957; Fax: +1(480) 965-3051; Email: mangone@asu.edu

Present Address: Marco Mangone, Arizona State University, Biodesign Institute Building A, 1001 S McAllister Ave, Tempe, AZ 85281 USA

This PDF includes:

- Figures S1-S14
- Supplemental Materials and Methods
Table of Contents

**Additional File 1**

**Figure S1:** Protein alignment of members of the CPC complex in five species .................................................................3

**Figure S2:** Results of the RNAi experiments of the *C. elegans* CPC..................14

**Figure S3:** Bioinformatic pipeline used in this study .......................................15

**Figure S4:** Comparison with Tourasse et al., 2017..............................16

**Figure S5:** PAS site usage in genes with multiple 3'-UTR isoforms...........17

**Figure S6:** Detection of the PAS element in genes lacking a canonical AAUAAA hexamer................................................18

**Figure S7:** GO term analysis for genes with 1, 2 or 3 3'-UTR isoforms ........19

**Figure S8:** Detection of the ‘UGUA’ element in *C. elegans* 3'-UTRs........20

**Figure S9:** Detection of enriched elements in 3'-UTRs of genes with 2 3'-UTR isoforms...................................................21

**Figure S10:** Nucleotide binding site of the human CPSF160-WDR33-CPSF30 complex .........................................................22

**Figure S11:** *In vivo* cleavage assay for M03A1.3 ...........................................23

**Figure S12:** *In vivo* cleavage assay for Y106G6H.9.................................24

**Figure S13:** *In vivo* cleavage assay for ges-1 ...........................................25

**Figure S14:** miRNA target analysis in genes with 2 3'-UTR isoforms which either gain or lose a miRNA binding site.................26

**Supplemental Materials and Methods:** .................................................27
Supplemental Figure S1: Protein alignment of members of the CPC complex in five organisms. Amino acid sequence alignment for the members of the CPC from five organisms produced with Clustal Omega Multiple Sequence Alignment. Conserved domains produced with the Batch Conserved Domain Search on NCBI are represented by the highlighted regions of each figure. Known domains determined from previously published literature are outlined.
CPSF Complex

Supplemental Fig. S1

CPSF-4 (CPSF4)

- Zinc Finger 1 (Clerici et al., 2017)
- Zinc Finger 2, contacts A1 and A2 of PAS (Clerici et al., 2017, Sjun et al., 2018)
- Zinc Finger 3, contacts A4 and A5 of PAS (Clerici et al., 2017, Sjun et al., 2018)
- Zinc Finger 4 (Clerici et al., 2017)
- Zinc Finger 5 (Clerici et al., 2017)
- Zinc Knuckle (Clerici et al., 2017)

FIPP-1 (FIP1L1)

- Conserved Domain contacts CPSF30 (Clerici et al., 2017)
- RE/D region interacts with CFIm68/59 (Zhu et al., 2018)
PFS-2 (WDR33)

- N-terminal domain, contacts CPSF160 and CPSF30 (Clerici et al., 2017)
- WD40 repeats, contact U3 and A6 of PAS (Clerici et al., 2017)
### CPF-1 (CST1)

- **WD40 Domain**
- **CSTF1 Dimer**

| Domain       | Normal | Human | Mouse | Rat   | Fly   |
|--------------|--------|-------|-------|-------|-------|
| N-terminal   |        |       |       |       |       |
| WD40 Repeats |        |       |       |       |       |

### CPF-2 (CSTF2)

- **RRM SF**
- **CSTF2 HINGE**
- **CSTF C**

| Domain     | Normal | Human | Mouse | Rat   | Fly   |
|------------|--------|-------|-------|-------|-------|
| C-terminal |        |       |       |       |       |
| HINGE      |        |       |       |       |       |
| RRM        |        |       |       |       |       |

---

**N-terminal domain (Yang et al., 2018)**

**WD40 Repeats (Yang et al., 2018)**

**C-terminal domain (Yang et al., 2018)**

**HINGE domain (Yang et al., 2018)**
### Proline-rich segment that facilitates interactions with CstF-64 and CstF-55 (Bai et al., 2007)

### HAT-N Domain (Bai et al., 2007)

### HAT-C Domain that interacts with CPSF160 (Bai et al., 2007)
Supplemental Fig. S1

CLPF-1 (CLP1)

| N-terminal domain (Dikfidan et al., 2014) | Polynucleotide kinase domain (Dikfidan et al., 2014) | C-terminal domain (Dikfidan et al., 2014) | Residues that crosslink to PCF-11 (Schäfer et al., 2018) |
|------------------------------------------|--------------------------------------------------|------------------------------------------|----------------------------------------------------------|

| Species | Accession | N-terminal domain | Polynucleotide kinase domain | C-terminal domain | Residues that crosslink to PCF-11 |
|--------|-----------|-------------------|-----------------------------|-------------------|----------------------------------|
| Human  | NP_009077 |                   |                             |                   |                                  |
| Fly    | XP_348021 |                   |                             |                   |                                  |
| Resort | NP_009077 |                   |                             |                   |                                  |
| Mouse  | NP_009077 |                   |                             |                   |                                  |
| N-terminal domain (Dikfidan et al., 2014) | Polynucleotide kinase domain (Dikfidan et al., 2014) | C-terminal domain (Dikfidan et al., 2014) | Residues that crosslink to PCF-11 (Schäfer et al., 2018) |

| Species | Accession | N-terminal domain | Polynucleotide kinase domain | C-terminal domain | Residues that crosslink to PCF-11 |
|--------|-----------|-------------------|-----------------------------|-------------------|----------------------------------|
| Human  | NP_009077 |                   |                             |                   |                                  |
| Fly    | XP_348021 |                   |                             |                   |                                  |
| Resort | NP_009077 |                   |                             |                   |                                  |
| Mouse  | NP_009077 |                   |                             |                   |                                  |
| N-terminal domain (Dikfidan et al., 2014) | Polynucleotide kinase domain (Dikfidan et al., 2014) | C-terminal domain (Dikfidan et al., 2014) | Residues that crosslink to PCF-11 (Schäfer et al., 2018) |

| Species | Accession | N-terminal domain | Polynucleotide kinase domain | C-terminal domain | Residues that crosslink to PCF-11 |
|--------|-----------|-------------------|-----------------------------|-------------------|----------------------------------|
| Human  | NP_009077 |                   |                             |                   |                                  |
| Fly    | XP_348021 |                   |                             |                   |                                  |
| Resort | NP_009077 |                   |                             |                   |                                  |
| Mouse  | NP_009077 |                   |                             |                   |                                  |
| N-terminal domain (Dikfidan et al., 2014) | Polynucleotide kinase domain (Dikfidan et al., 2014) | C-terminal domain (Dikfidan et al., 2014) | Residues that crosslink to PCF-11 (Schäfer et al., 2018) |

| Species | Accession | N-terminal domain | Polynucleotide kinase domain | C-terminal domain | Residues that crosslink to PCF-11 |
|--------|-----------|-------------------|-----------------------------|-------------------|----------------------------------|
| Human  | NP_009077 |                   |                             |                   |                                  |
| Fly    | XP_348021 |                   |                             |                   |                                  |
| Resort | NP_009077 |                   |                             |                   |                                  |
| Mouse  | NP_009077 |                   |                             |                   |                                  |
| N-terminal domain (Dikfidan et al., 2014) | Polynucleotide kinase domain (Dikfidan et al., 2014) | C-terminal domain (Dikfidan et al., 2014) | Residues that crosslink to PCF-11 (Schäfer et al., 2018) |
Supplemental Fig. S1

Domain With No Name (DWNN), ubiquitin-like (Pugh et al., 2006)
Proline-rich domain (Pugh et al., 2006, Lee et al., 2014)
SR domain (Pugh et al., 2006)
Rb-binding domain (Pugh et al., 2006)
p53-binding domain (Pugh et al., 2006)
RING finger domain (Chibi et al., 2008, Lee et al., 2014)
Zinc knuckle (Lee et al., 2014)
Supplemental Fig. S1

PAP-1 (PAP1N)

N  PAP Central  PAP RNA Binding  C

Legend:

Mouse
Fly
Human
Cat
**Supplemental Fig. S1**

### SYMK-1 (Symplekin)

| Organism | Accession | Description |
|----------|-----------|-------------|
| **Fly**  | SF50100   | Symplekin C DUF3453 |
Supplemental Figure S2: Results of the RNAi experiments of the *C. elegans* CPC. Twelve genes for the members of the *C. elegans* CPC were knocked-down using RNAi. Clones/rows are color-coded as from Figure 1. The human orthologs of each gene are shown in parenthesis in the first column. For each RNAi experiment we use 15 worms, and the number of eggs unhatched vs hatched at the end of the experiment were counted. The percent lethality was consistently high across all tested clones. *pkc-3* RNAi was used as a negative RNAi control, since it is known to induce strong embryonic lethality.

| Name   | Experiment # | Eggs/Hatched | Lethality (%) |
|--------|--------------|--------------|---------------|
| *cpsf-1* (CPSF1) | 1 | 163/11 | 93.7 |
| | 2 | 238/5 | 97.9 |
| | 3 | 149/1 | 99.3 |
| *cpsf-2* (CPSF2) | 1 | 68/37 | 64.8 |
| | 2 | 178/28 | 86.4 |
| | 3 | 221/25 | 89.8 |
| *cpsf-4* (CPSF4) | 1 | 323/5 | 98.5 |
| | 2 | 699/17 | 97.6 |
| | 3 | 204/3 | 98.6 |
| *cpf-1* (CST1) | 1 | 251/76 | 76.8 |
| | 2 | 200/64 | 75.8 |
| | 3 | 185/25 | 88.1 |
| *cpf-2* (CSTF2) | 1 | 446/54 | 89.2 |
| | 2 | 138/14 | 90.8 |
| | 3 | 61/3 | 95.3 |
| *cfin-1* (NUDT21) | 1 | 249/23 | 91.5 |
| | 2 | 196/20 | 90.7 |
| | 3 | 154/2 | 98.7 |
| *cfin-2* (CPSF6) | 1 | 137/1 | 99.3 |
| | 2 | 282/27 | 91.3 |
| | 3 | 260/32 | 89.0 |
| *symk-1* (Symplekin) | 1 | 62/4 | 93.9 |
| | 2 | 19/4 | 82.6 |
| | 3 | 114/5 | 95.8 |
| *rbpl-1* (RBBP6) | 1 | 293/0 | 100 |
| | 2 | 344/0 | 100 |
| | 3 | 116/0 | 100 |
| *pcf-11* (PCF11) | 1 | 141/5 | 96.6 |
| | 2 | 105/0 | 100 |
| | 3 | 172/5 | 97.2 |
| *clpf-1* (CLP1) | 1 | 286/23 | 92.6 |
| | 2 | 289/18 | 94.1 |
| | 3 | 208/17 | 92.4 |
| *pkc-3* (negative control) | 1 | 5/0 | 100 |
| | 2 | 8/0 | 100 |
| | 3 | 18/1 | 94.7 |
| | 4 | 51/1 | 98.1 |
| | 5 | 3/0 | 100 |
| | 6 | 22/0 | 100 |
| | 7 | 53/0 | 100 |
| | 8 | 6/0 | 100 |
| | 9 | 13/0 | 100 |
| | 10 | 19/0 | 100 |
| | 11 | 5/0 | 100 |
| | 12 | 4/0 | 100 |
**Supplemental Figure S3: Bioinformatic Pipeline used in this study.** The pipeline uses raw transcriptome datasets downloaded from the public repository SRA trace archive to extract and map 3'-UTR end clusters to the closest protein-coding genes in the correct orientation. The pipeline is divided in three large steps: A) Acquisition/Mapping, B) 3'-UTR cluster preparation and C) 3'-UTR isoforms mapping.

In the acquisition/mapping step, we used custom made Perl scripts to extract reads with 23 consecutive As at the 3'-end or 23 consecutive Ts at the 5'-end and then mapped these filtered reads to the WS250 version of the *C. elegans* genome (Bowtie 2). We then sorted and indexed the reads for visualization purposes.

In the 3'-UTR cluster preparation step, we extracted SAM reads with 100% match to the WS250 and used them to prepare a new bedGraph file (BEDTools). We then merged the reads and discarded the clusters with less than 5 reads. Restrictive parameters for cluster identification and 3'-UTR end mapping included the discard of clusters with an adenosine content of <35% downstream of its end.

Clusters were assigned to a mapped 3'-UTR end and attached to the closest gene with 2,000 nt in the same orientation. At the completion of these steps we performed the 3'-UTR isoform mapping step, which consists of the counting and assignment the total number of 3'-UTR isoforms to a given gene.

We discarded clusters with a density of less than 30% of the total number of reads.
Supplemental Figure S4: Comparison with Tourasse et al., 2017. We have downloaded the poly(A) sites mapped in Tourasse et al. and performed a comparison with the 3'-UTRs present in our 3'-UTRome v2. A) Top Panel: number of mapped poly(A) sites from Tourasse et al. that match our stringent and raw datasets within +/-1 nt, +/-5 nt, +/-10 nt or +/-20 nt. Bottom Panel: number of poly(A) sites in common between Tourasse et al. and Mangone et al. and between Tourasse et al. and Jan et al. within +/-1nt, +/-5nt, +/-10nt or +/-20nt. B) Bar chart showing the % of similarity of the two datasets in Panel A. C. Venn diagrams comparing the 3'-UTRs shared (+/- 20nt) between Tourasse et al., and the UTRome v1 (Mangone et al., - (green), this study (UTRome v2 - red), and Jan et al. (orange). We have used our unfiltered dataset to compare UTRome v2 with Tourasse et al.
Supplemental Figure S5: PAS site usage in genes with multiple 3’-UTR isoforms. A) In genes with only two 3’-UTR isoforms with a difference of at least 10 nt between isoforms, 373 pairs of isoforms had canonical PAS elements in both isoforms with an average of 246 nt difference between isoforms while 665 pairs had variant PAS elements in both isoforms with an average of 125 nt difference between them. In isoform pairs where the type of PAS element switches, 71% have a shorter isoform with a variant PAS element and a longer isoform with a canonical PAS element with an average of 209 nt between them while the remaining 29% have a canonical PAS element on the shorter isoform and a variant PAS element on the longer isoform with an average of 322 nt between them. B) In genes with three or more 3’-UTR isoforms, genes where the longest and the shortest isoform both have canonical PAS elements have an average of 163 nt between them while genes where the longest and the shortest isoforms both have variant PAS elements have an average of 211 nt between them. 72% of genes switch from a variant PAS elements in the short isoform to a canonical PAS element in the long isoform, with an average of 181 nt between them. 28% of genes have canonical PAS elements in the short isoforms and variant PAS elements in the long 3’-UTR isoform, with an average of 211 nt between the two.
Supplemental Figure S6: Detection of the PAS element in genes lacking a canonical AAUAAA hexamer. A) We have searched for the most common RRYRRR motifs located within the last 30 nt in genes with 1 (top) or 2+ (bottom) 3'-UTR isoforms with no canonical PAS and with a detectable RRYRRR motif. The left chart shows the occurrences of the seven most common PAS elements identified in these groups, and the right logo shows the identified PAS motif. Apart from a slight increase in percentage of adenosines and guanosines in 3' UTRs of genes with 2+ 3' UTRs, both results are similar. The overwhelming majority of PAS conform with the AAU(G/A)AA with a tolerance of 1 nt purine-purine or pyrimidine-pyrimidine replacement. B) PAS site usage in genes with 1 or 2 3'-UTR isoforms with non canonical PAS and no RRYRRR motif. The pie chart shows the 10 most common hexamers located within 30 nt upstream of the the cleavage site. Right logo plot shows the nucleotide conservation of the intergenic region encompassing the cleavage site from -30 to +20 nt. Arrow marks the cleavage site. The buffer region and the PAS is marked.
### Supplemental Figure S7: GO term analysis for genes with 1, 2 or 3 3’-UTR isoforms.

#### 1 3’-UTR isoform (n=8,537)

| adj.Pval   | nGenes | Pathways                          |
|------------|--------|-----------------------------------|
| 5.7e-28    | 165    | phosphorus metabolic process      |
| 7.2e-28    | 220    | single-organism metabolic process |
| 2.5e-26    | 159    | phosphate-containing compound metabolic process |
| 8.7e-23    | 149    | macromolecule modification        |
| 1.3e-17    | 98     | phosphorylation                   |
| 1.3e-16    | 130    | cellular protein modification process |
| 4.9e-16    | 164    | cellular protein metabolic process |
| 6.2e-14    | 75     | protein phosphorylation           |
| 4.7e-13    | 90     | oxidation-reduction process       |

#### 2 3’-UTR isoform (n=4,741)

| adj.Pval   | nGenes | Pathways                          |
|------------|--------|-----------------------------------|
| 3.6e-20    | 135    | single-organism metabolic process |
| 1.4e-13    | 38     | immune response                   |
| 1.4e-13    | 38     | immune system process             |
| 5.0e-13    | 63     | oxidation-reduction process       |
| 8.7e-12    | 35     | innate immune response            |
| 7.5e-11    | 82     | phosphorus metabolic process      |
| 4.0e-10    | 41     | defense response                  |
| 1.1e-09    | 78     | phosphate-containing compound metabolic process |
| 9.9e-08    | 41     | carbohydrate derivative metabolic process |
| 1.0e-07    | 51     | small molecule metabolic process  |

#### 3+ 3’-UTR isoform (n=1,530)

| adj.Pval   | nGenes | Pathways                          |
|------------|--------|-----------------------------------|
| 2.3e-03    | 19     | small molecule metabolic process  |
| 2.3e-03    | 21     | organonitrogen compound metabolic process |
| 2.3e-03    | 8      | ribonucleoside monophosphate metabolic process |
| 2.3e-03    | 8      | nucleoside monophosphate metabolic process |
| 2.3e-03    | 11     | innate immune response            |
| 2.3e-03    | 11     | innate immune response            |

#### Pathways Examples

- **Supplemental Fig. S7**: GO term analysis for genes with 1, 2 or 3 3’-UTR isoforms.
Supplemental Figure S8: Detection of the ‘UGUA’ element in *C. elegans* 3′-UTRs. A) Logo plot of the transcript’s region within the cleavage site in genes with only one 3′-UTR isoform and with a canonical or variant PAS element. B) Identification of the “UGUA” motif (red) within 100 nt upstream of the cleavage site in genes with one or two 3′-UTR isoforms. C) Binned frequency distribution of the occurrences of the AAUAAA and UGUA elements in distal 3′-UTR isoforms as in the right heatmap in Panel B (green and red) vs. the occurrences of the UGUA motif in a randomly generated 3′-UTR dataset (blue) (n=785).
Supplemental Figure S9: Detection of enriched elements in *C. elegans* 3'-UTRs of genes with 2 3'-UTR isoforms (only distal) (n=785). These motifs have been detected using the meme suite (Bailey et al., 2015.)
Supplemental Figure S10: Nucleotide binding site of the human CPSF160-WDR33-CPSF30 complex. Ribbon representation of the cryo-EM structure of human CPSF160-WDR33-CPSF30 complex (PDB code: 6DNF) (Sun et al., 2018). The nucleotides of the bound RNA fragment do not show a specific interaction with either CPSF30 or WDR33. The interactions are mostly established by \( \pi-\pi \) ring stacking. Color gray shows the CPSF160, pink for WDR33, and light green for CPSF30. Sticks represent the RNA molecules bound with CPSF30 and WDR33. Surfaces in the inlets are for individual nucleotides.
**Supplemental Figure S11: In vivo cleavage assay for M03A1.3.**

A) M03A1.3 genomic region cloned downstream of the GFP reporter. Blue: terminal portion of the M03A1.3 ORF. Green: STOP codon. Gray: 3'-UTR. Red: mutated terminal adenosine nucleotides. The transgenic worms expressing the Pmyo-3::GFP::M03A1.3_3’-UTR wt and mutant cassette are shown below.

B) At the completion of the experiment, we recovered the total RNA and performed RT-PCR experiments using a forward primer annealing within the GFP ORF and a reverse polydT primer with two anchors containing Invitrogen Gateway adapters. The resultant amplicons were then subcloned in gateway vectors and sequenced to detect the cleavage site. An example of resultant trace files is shown.

C) Examples of 10 clones identified in this study for M03A1.3. The removal of the terminal genomic adenosine nucleotide induces a cleavage site 3 nt upstream of the canonical cleavage site in three clones (arrows), which also contain a terminal adenosine nucleotide. The PAS element is boxed in blue color.
Supplemental Figure S12: In vivo cleavage assay for Y106G6H.9.

A) Y106G6H.9 genomic region cloned downstream of the GFP reporter. Blue: terminal portion of the Y106G6H.9 ORF. Green: STOP codon. Gray: 3'-UTR. Red: mutated terminal adenosine nucleotides. Red Asterisk: position of the cryptic cleavage site (see below). The transgenic worms expressing the Pmyo-3::GFP::Y106G6H.9_3'-UTR wt and mutant cassette are shown below.

B) At the completion of the experiment we recovered the total RNA and performed RT-PCR experiments using a forward primer annealing within the GFP ORF and a reverse polydT primer with two anchors containing Invitrogen Gateway adapters. The resultant amplicons were then subcloned in gateway vectors and sequenced to detect the cleavage site. An example of resultant trace files is shown.

C) Examples of several clones identified in this study for Y106G6H.9. In the wt we were able to detect two classes of cleavage sites, both ending within 4 nt of each other with a terminal adenosine nucleotide. In the double mutant, the removal of the terminal genomic adenosine induces a cleavage skip in two clones (arrows). In one case (red arrow), the cleavage occurs 20 nt downstream of the PAS element. Two of the mutant clones also shown an occurrence of a new cryptic cleavage site 100 nt upstream of the natural site (red asterisks), which also contain a terminal adenosine nucleotide at their 3'end. The PAS element is boxed in blue color.
Supplemental Figure S13: In vivo cleavage assay for ges-1.
A) ges-1 genomic region cloned downstream of the GFP reporter. Blue: terminal portion of the ges-1 ORF. Green: STOP codon. Gray: 3' UTR. Red: mutated terminal adenosine nucleotides. The transgenic worms expressing the Pmyo-3::GFP::ges-1_3'UTR wt and mutant cassette are shown below.

B) At the completion of the experiment we recovered the total RNA and performed RT-PCR experiments using a forward primer annealing within the GFP ORF and a reverse polydT primer with two anchors containing Invitrogen Gateway adapters. The resultant amplicons were then subcloned in gateway vectors and sequenced to detect the cleavage site. An example of resultant trace files is shown.

C) Examples of 10 clones identified in this study for ges-1. The removal of the terminal genomic adenosine nucleotide does not alter the cleavage site but makes it more variable. The PAS element is boxed in blue color.
**A**

**GO Term: Biological Process**

| adj.Pval | nGenes | Pathways                          |
|----------|--------|-----------------------------------|
| 7.2e-05  | 7      | oxidation-reduction process       |
| 5.9e-04  | 2      | deoxyribonucleoside diphosphate metabolic process |
| 1.7e-03  | 8      | single-organism metabolic process |

**GO Term: Molecular Component**

| adj.Pval | nGenes | Pathways                    |
|----------|--------|------------------------------|
| 1.2e-03  | 5      | oxidoreductase activity      |

**Functional Enrichment: Kegg**

| adj.Pval | nGenes | Pathways                                 |
|----------|--------|------------------------------------------|
| 1.7e-03  | 1      | Pentose and glucuronate interconversions |
| 1.7e-03  | 1      | Fructose and mannose metabolism         |

**B**

| miRNA   | Sequence                                | Hits |
|---------|-----------------------------------------|------|
| miR-272 | uguagggcauggguguuug                     | 7    |
| miR-2217a | cagaguggcagucguggucgauuc               | 6    |
| miR-2217b | cagagcggcagucggugucgaauc              | 6    |
| miR-5553 | ucaauggguagcagcgguucaaga                 | 6    |
| miR-265  | uagggaggagaaaggggggguau                 | 5    |
| miR-34   | aggcaguguguuacggguug                   | 5    |
| miR-44   | agaucuagacacauucacgcu                  | 5    |
| miR-4935 | ggccggagagagggggggagcg                 | 5    |
| miR-795  | uagggagauguuacgagcgagcuu             | 5    |
| miR-8190 | cggggaaacgcuuugggaauuccagga          | 5    |
| miR-8194 | auggggugguuuuaaaaggguacgg             | 5    |
| miR-1822 | aguuucucagggaaagcuuacggc             | 4    |
| miR-71   | ugaagagcauggguagugagacg              | 4    |

Supplemental Figure S14: miRNA target analysis in genes with 2 3'-UTR isoforms which either gain or lose a miRNA binding site. A) GO Term analysis of genes with multiple 3'-UTRs which gain or lose a miRNA target as predicted using our miRanda 'stringent' dataset (n=132). B) Most common miRNAs, their sequences, and number of occurrences.
Additional Materials and Methods

Comparative analysis of *C. elegans* members of the CPC

We have downloaded the protein sequences of each known member of the human CPC and used BLAT algorithm to identify *C. elegans* genes with high homology to their human counterparts. We then performed a protein BLAST analysis using the tools available at the NCBI website to obtain the amino acid sequences for the fly, rat, and mouse orthologs. These amino acid sequences were then aligned using Clustal Omega Multiple Sequence Alignment with standard parameters. At the completion of the analysis, we used the Batch NCBI Conserved Domain Search (Batch CD-Search) against the database CDD- 52910 PSSMs using standard parameters to identify the conserved domains across the aligned protein sequences. We then used these results to populate the location of these elements within the alignment shown in **Supplemental Figure S1**. We were unable to identify the *C. elegans* homolog of the human gene CPSF7.

Plasmid DNA isolation, sequencing and visualization

All plasmids used in this study were prepared from cultures grown overnight in LB using the Wizard Plus SV Minipreps DNA Purification System (Promega) according to the manufacturer’s instructions. DNA samples were sequenced with Sanger sequencing performed at the DNASU Sequencing Core Facility (The Biodesign Institute, ASU, Tempe, AZ).
RNAi experiments

RNAi experiments were performed in standard NGM agar containing 1mM IPTG and 50 µg/ml ampicillin. These plates were seeded with 75 µl of RNAi clone bacteria and allowed to induce for a minimum of 16 hours. 5 N2 C. elegans at the L1 stage were aliquoted for each RNAi clone tested. Three days after plating, the progeny was scored for embryonic lethality. Each RNAi experiment was performed in triplicate. The total number of hatched and not hatched eggs was the following: cpsf-1(CPSF160) n=567; cpsf-2(CPSF100) n=557; cpsf-4(CPSF30) n=1,251; cpf-2(CstF64) n= 716; cpf-1(CstF50) n=801; cfim-1(CFlm25) n=644; cfim-2(CFlm68) n=739; symk-1(symplekin) n=208; tag-214(RBBP6) n=753; pcf-11(CPF11) n=428; clpf-1(CLP1) n=841.

Mutagenesis of 3’-UTRs cleavage sites

The mutagenesis reactions to remove the adenosine nucleotides near the cleavage sites were carried out using the QuikChange Site-Directed Mutagenesis Kit (Agilent). The mutagenesis DNA primers for the site mutation reactions are available in Supplemental Table S2. Each mutagenesis reaction was followed by DNA digestion using Dpn-1 enzyme and transformed in Top10 competent cells (Thermo Fisher Scientific) in agar plates containing 20mg/µL of kanamycin. We validated the nucleotide mutation using Sanger sequencing approach. Wild type and mutant 3’-UTRs cloned in pDONR P2RP3 were then shuttled into destination vectors using the Gateway LR Clonase II Plus Enzyme Mix (Invitrogen, Carlsbad, CA). The finalized destination vectors contained the C. elegans pharynx promoter (Pmyo-2) in the first position, a GFP
sequence with a mutated STOP codon in the second position, and the wt or mutant 3' UTRs used in this study in the third position. The resultant recombined constructs were then transformed in Top10 competent cells (Thermo Fisher Scientific) and plated on 10mg/µL ampicillin plates overnight. The success of the recombination reaction was confirmed using Sanger sequencing with the M13F DNA primer.

**Preparation of transgenic worm lines**

EG6699 strain worms were kindly provided by Christian Frokjaer-Jensen (Frokjaer-Jensen et al. 2008). These worm strains were maintained at 18°C on nematode growth media (NGM) agar plates and propagated on plates seeded with OP50-1 bacteria. To synchronize worms for injections, EG6699 worms were bleached with bleaching solution (1 M NaOH) four days before injections. Each construct was mixed with an injection master mix containing pCFJ601 (25 ng/µl), pgH8 (10 ng/µl), and pCFJ104 (5 ng/µl) vectors. Injection needles were loaded with the injection mixture and mounted to the Leica DMI300B microscope. The needle was pressurized with 22 psi through the FemtoJet (Eppendorf). Young adult EG6699 worms were picked onto an agarose pad covered with mineral oil on a glass coverslip. Injected worms were rescued onto an NGM plate and rinsed with M9 buffer. Two days post-injections, the F1 progeny were screened with a Leica DMI3000B microscope for both unc-119 rescues and expression of the red fluorescence produced by the co-injection marker and then isolated onto individual plates. These worms were allowed to lay eggs, and then the F2 progeny was screened for fluorescence. Once 75% of the progeny on a single plate were transgenic, the strains were used for further experimentation.
Worm genotype validation

Populations obtained from single worms from each of the seven strains were lysed using worm lysis buffer (EDTA, 0.1 M Tris, 10% Triton-X, Proteinase K, 20% Tween 20). These worms were subjected to heating in a Bio-Rad T100 Thermal Cycler. To confirm that the mutated cleavage site was present in the injected strains, we used PCR approach using Platinum Taq polymerase (Invitrogen) with a forward DNA primer binding the beginning of the GFP sequence and 3'-UTR-specific reverse DNA primers. The PCR product was then sequenced using Sanger sequencing with a forward DNA primer binding to the GFP sequence present in the injected construct.

Detection of the 3'-UTR cleavage skipping

Total RNA was extracted from transgenic strains using the Direct-zol RNA MiniPrep Plus kit (RPI) according to the manufacturer's instructions. We tested approximately 10 independent wt and mutant clones for each 3'-UTR. Approximately 50 µL of worm pellet was used for extraction. cDNA was synthesized using a reverse transcription reaction using Superscript II enzyme (Invitrogen). The first strand reaction was performed using a reverse poly dT DNA primer containing two anchors and the attB Gateway BP recombination element (Invitrogen). The second strand of the cDNA was synthesized using a PCR with HiFi taq polymerase (Thermo Fisher Scientific) and the forward DNA primer containing the pDONR P2RP3 Gateway element (Invitrogen), which binds to GFP and the same reverse poly dT DNA primer used in the first strand reaction. The BP Gateway kit (Invitrogen) was once again used to clone the cDNA which contains the
polyA tail into pDONR P2RP3. These constructs were then transfected into Top10 competent cells (Thermo Fisher Scientific) and plated on agar plates containing 20mg/µL of Kanamycin. About 8-10 colonies were then sequenced with Sanger sequencing using the M13F DNA primer to map the location of the cleavage site.

**Updated miRanda Predictions**

We downloaded a complete list of *C. elegans* miRNAs from miRBase (Griffiths-Jones et al. 2006) and the miRanda algorithm v3.3a (John et al. 2004) from the microrna.org website. We queried the 3’-UTRome v2 with the miRanda algorithm using both standard and stringent parameters. The stringent query used was ‘-strict -sc -1.2’. The standard query produced 58,330 putative miRNA targets; the stringent query produced 12,136 putative miRNA targets. Both these predictions are included in WormBase (Lee et al. 2018) as individual tracks.

**Homology model building**

Homology modeling was performed using SWISS_MODEL (Waterhouse et al. 2018) with a matched templated of human CPSF160-WDR33-CPSF30 complex (PDB code: 6DNF) (Sun et al. 2018). The molecular graphics were prepared using the UCSF ChimeraX software (version 0.8) (Goddard et al. 2018).