POS-1 promotes endo-mesoderm development

by inhibiting the cytoplasmic

polyadenylation of *neg-1* mRNA

SUPPLEMENTAL INFORMATION

Figure S1. *gld-3(RNAi)* restores gut specification in *pos-1* embryos (Related to Figure 1) (A) *in situ* hybridization of *gfp* mRNA in representative *med-1::gfp* (wildtype) and *med-1::gfp; pos-1(RNAi)* early embryos. *med-1* expression is reduced in EMS and its daughter cells upon *pos-1* knockdown (B) Embryos of *pos-1(-)* are gutless and arrest devoid of gut granules. Birefringent gut granules are restored in *pos-1; gld-3(RNAi)* embryos (white arrow). (C) The wildtype lineage of EMS show follows a characteristic architecture, expresses the endo-mesoderm marker *pha-4* (red) and includes single programmed cell death (green x). These three attributes are lost in *pos-1(-)* embryos and restored in *pos-1(-), gld-3(-)* embryos (**Experimental Procedures**).

Figure S2. NEG-1 is expressed in germline and is localized to the chromatin of early embryonic nuclei (Related to Figure 3) (A) NEG-1::GFP is expressed in germline with localization restricted to distal germ cells, maturing and mature oocytes. An inset image obtained by confocal microscopy indicates NEG-1::GFP localization to the distal nuclei (compare GFP+ and GFP- nuclei). Gonad was dissected and imaged without fixation. Asterisks mark the distal end of the germline. (B) Subnuclear NEG-1::GFP localization is most intense along chromatin.

Figure S3. 3'UTR^{neg-1} regulates asymmetric *neg-1* expression in early embryos (Related to Figure 3) (A) Arrows highlight asymmetric expression in 4-cell stage. Solid triangles highlight expression in early embryos (AB8 and AB16 stages), whereas empty triangles highlight reduced expression in later embryos (post AB16 stage). Live embryos imaged without fixation. (B) MEX-5 competes POS-1 in binding to *neg-1* UTR fragment (RBPc WT). Top right, MEX-5 (236-350) is titrated into a fixed concentration of POS-1 (80-180). Top left, POS-1 (80-180) is

titrated into a fixed concentration of MEX-5 (236-350). The fixed concentration of the proteins is chosen to be a concentration at which the proteins are 70% bound. Bottom, quantifications of the competition experiments. Left, the fraction of bound RNA as MEX-5 is titrated. Open circles denote MEX-5 bound to RNA and the filled circles denote POS-1 bound to RNA. Right, the fraction of bound RNA as POS-1 is titrated. Open circles denote MEX-5 bound to RNA and the filled circles denote POS-1 bound to RNA. (C, D) Electrophoretic mobility shift assays and graphical representations of fluorescence polarization data obtained from incubating the fluorescent probes (P1M3B) with increasing amounts of MEX-3 or POS-1. The apparent dissociation constant, $K_{d,app}$ represents the average of three independent replicates ± the standard deviation, SD.

Figure S4. PHA-4::GFP is ectopically epressed in the AB lineage of *neg-1* embryos (Related to Figure 5) Six lineages of AB blastomere in *neg-1(tm6077)* embryos. Red sublineages note PHA-4::GFP expression. Lineages were conducted using strain RW10425 (**Experimental Procedures**) crossed with *neg-1(tm6077)* and according to Du *et al* 2014.

Figure S5. *neg-1* and its homologous are novel genes that encode proteins with skewed WYF and RK sequences. (Related to Figure 1 and 6) (A) *neg-1* and its two homologues (CBN07554 and CJA09859) reside in the same genomic neighborhood. Four genomic loci stretching ~ 25kbp from (top to bottom) *C. brenneri*, *C. elegans*, *C. japonica* and *C. briggsae*. *neg-1* (grey) is located in near *fipp-1* (red) and *grl-13* and two upstream genes (orange). F13A2.9 (blue) and its homologs demarcate one end of the genomic neighborhood. The locus in *C. brigssae* includes all landmarks but no *neg-1* homolog. Instead a larger predicted gene (CBG23621) resides where *neg-1* would be expected. (B) Multiple alignment of NEG-1 and *C. brenneri* (CBnNEG-1) and *C. japonica* (CJpNEG-1) homologs. Asterisks (*) note identical amino acids, colons (:) note similar and periods (.) note weakly similar amino acids. Putative nuclear localization sequences are boxed. Aromatic residues W, F and Y are in white with a blue

background. Basic residues R and K are in white with a red background. (C) NEG-1, its two homologs (CBN07554 and CJA09859), MED-1, MED-2, END-1 and END-3 all exhibit a biased distribution of aromatic (W, Y and F blue boxes) and basic (R and K red boxes) amino acids. GATA zinc fingers are noted (black lines) based on Pfam annotation. (D) Table summarizing amino acid bias. Amino acids W, Y, F, R and K were counted in the N-terminal and C-terminal halves.

Table S1: Suppressors of *pos-1* gutless phenotype (Related to Figure 1)

Gut granule scoring of progeny of *pos-1(zu148)* hermaphrodites fed RNAi food targeting specific genes (*n* = number of independent hermaphrodites assayed, *e* = number of embryos scored)

Genotype	Gut granules	Annotation/Function
pos-1(zu148)	1.0% ± 2.3 (*n*=9, *e*=270)	Tandem CCCH Zinc finger
pos-1(zu148); gld-2(RNAi)	41.2% ± 12.3 (*n*=4, *e*=38)	Cytoplasmic poly(A) polyermase
pos-1(zu148); neg-1(RNAi)	78.8% ± 10.8 (*n*=7, *e*=152)	unknown
pos-1(zu148); rev-1(RNAi)	76.7% ± 14.4 (*n*=4, *e* 90)	Translesion DNA polymerase
pos-1(zu148); rba-1(RNAi)	46.7% ± 29.2 (*n*=8, *e*=126)	RBAp48 related
pos-1(zu148); rpn-8(RNAi)	23.8% ± 5.8 (*n*=4, *e*=44)	Proteasome regulatory particle
pos-1(zu148); cye-1(RNAi)	19.1% ± 17.7 (*n*=5, *e*=110)	Cyclin E
pos-1(zu148); lrr-1(RNAi)	29.9% ± 33.1 (*n*=6, *e*=113)	Leucine-Rich Repeat protein

Table S2. GLD-3/2 activate and POS-1 represses transcripts enriched in the anterior blastomere AB (Related to Figure 6)

Comparison of PAT-seq data analysis and single-cell transcriptomics (AB or P1 enrichment) based on Osborne-Nishimura *et al*.

Table S3: PAT-seq results and analysis (Related to Figure 6)

Tab: **mello-genewise—counts.csv** (PAT-seq data summary provided by Traude Beilharz)

Coloumn A: gene name.

Columns B –M: average poly(A) tail length in N2, pos-1(RNAi), gld-3(RNAi), gld-2(RNAi), mex-5(RNAi) and mex-6(RNAi) early embryos. In duplicates.

Columns O – T: average of corresponding duplicates.

Column V: "Enrichment rank" in AB or P1 blastomeres (e.g. AB_1 highest transcript level in AB compared to P1). Based on Osborne-Nishimura *et al*. personal communication. Note that not all ranks are included (e.g. AB_2 and AB_3 are missing) since I only included here genes that are represented in PAT-seq dataset.

Column W: gene name (e.g. *neg-1* is the fourth most enriched transcript in AB when compared to its levels in P1).

Tab: **P1RG3A** (POS-1 repressed, GLD-3 activated genes) SORTED according to Column P

Column A: gene name.

Columns B –M: average poly(A) tail length in N2, pos-1(RNAi), gld-3(RNAi), gld-2(RNAi), mex-5(RNAi) and mex-6(RNAi) early embryos. In duplicates.

Columns O – T: average of corresponding duplicates. Sheet is sorted according to Column P to (i.e. genes with shorter poly(A) tails in *pos-1(RNAi)* early embryos are first).

Column U: PR = POS-1 repressed (i.e. poly(A) tail length is longer in *pos-1(RNAi)* early embryos compared to wildtype N2.

Column V: G3A = GLD-3 activated (i.e. poly(A) tail length is shorter in *gld-3(RNAi)* early embryos compared to wildtype N2.

Column X: "Enrichment rank" in AB or P1 blastomeres (e.g. AB_1 highest transcript level in AB compared to P1). Based on Osborne-Nishimura *et al*. Note that not all ranks are included (e.g. AB_2 and AB_3 are missing) since I only included here genes that are represented in PAT-seq dataset.

Column Y: gene name (e.g. *neg-1* is the fourth most enriched transcript in AB when compared to its levels in P1). Highlighted cells refer to genes that are represented in this sheet (i.e. POS-1 repressed GLD-3 activated)

Tab: **P1RG2A** (POS-1 repressed, GLD-2 activated genes) Same as tab P1RG3A, but for GLD-2.

Tab: **P1RG23A** common (POS-1 repressed, GLD-2/3 activated genes). The common genes in previous two sheets

Tab: **POS-1 repressed_activated** SORTED according to Column W

Column A: gene name.
Columns B –M: average poly(A) tail length in N2, pos-1(RNAi), gld-3(RNAi), gld-2(RNAi), mex-5(RNAi) and mex-6(RNAi) early embryos. In duplicates.
Columns O – T: average of corresponding duplicates. Sheet is sorted according to Column P to (i.e. genes with shorter poly(A) tails in *pos-1(RNAi)* early embryos are first).
Column U: PA = POS-1 activated (i.e. poly(A) tail length is shorter in *pos-1(RNAi)* early embryos compared to wildtype N2). PR = POS-1 repressed (i.e. poly(A) tail length is longer in *pos-1(RNAi)* early embryos compared to wildtype N2).
Column W: deltapos-1 (column P – column O, i.e. the difference in poly(A) tail length in pos-1(RNAi) early embryos compared to wildtype N2. A negative number reflects reduction in poly(A) tail length, from which we infer that POS-1 promotes or protects tail length in wildtype early embryos. A positive number reflects increase in poly(A) tail length, from which we may infer that POS-1 curtails tail length in wildtype early embryos.
Column Y: "Enrichment rank" in AB or P1 blastomeres (e.g. AB_1 highest transcript level in AB compared to P1). Based on Osborne-Nishimura *et al*. personal communication. Note that not all ranks are included (e.g. AB_2 and AB_3 are missing) since I only included here genes that are represented in PAT-seq dataset.
Column Z: gene name (e.g. *neg-1* is the fourth most enriched transcript in AB when compared to its levels in P1). Cells highlighted in green refer to genes that are POS-1 repressed. Cells highlighted in red refer to genes that are POS-1 activated.
Column AA: deltapos-1 value from column W for each corresponding gene.

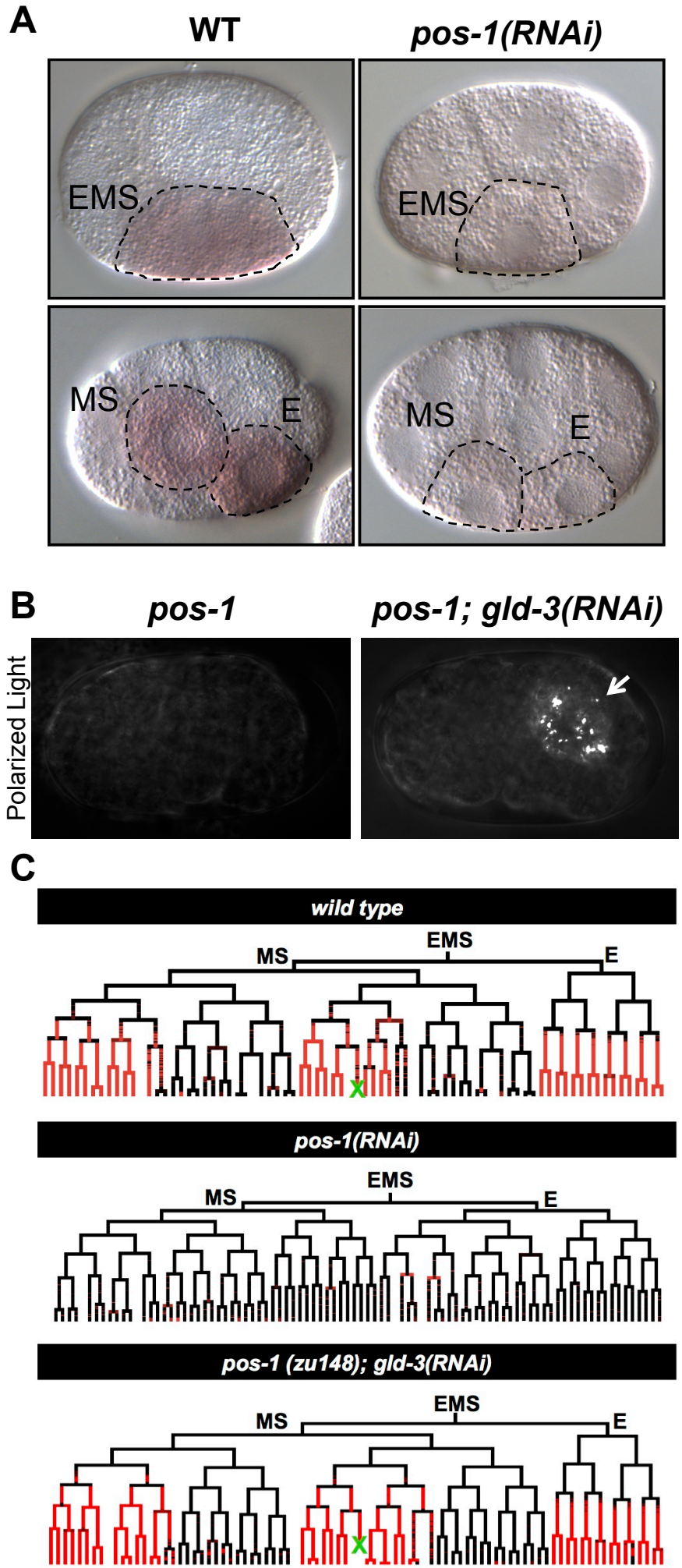
Supplemental Movie 1: NEG-1::GFP expression in early embryo (Related to Figure 2)

Confocal time-lapse imaging of early embryos expressing NEG-1::GFP.

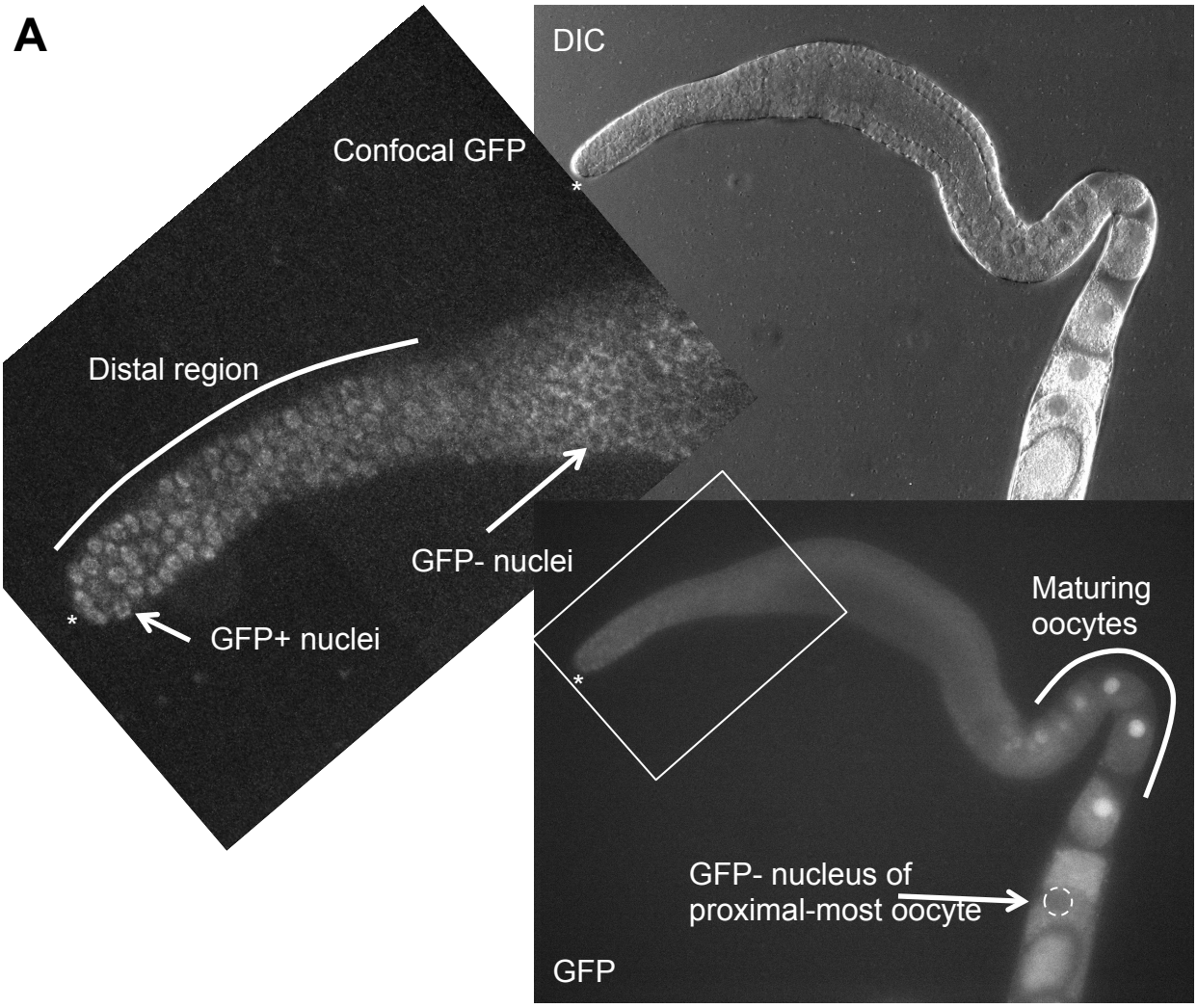
Supplemental Movie 2: *gfp*::3'UTR^{neg-1} expression in early embryo (Related to Figure 2)

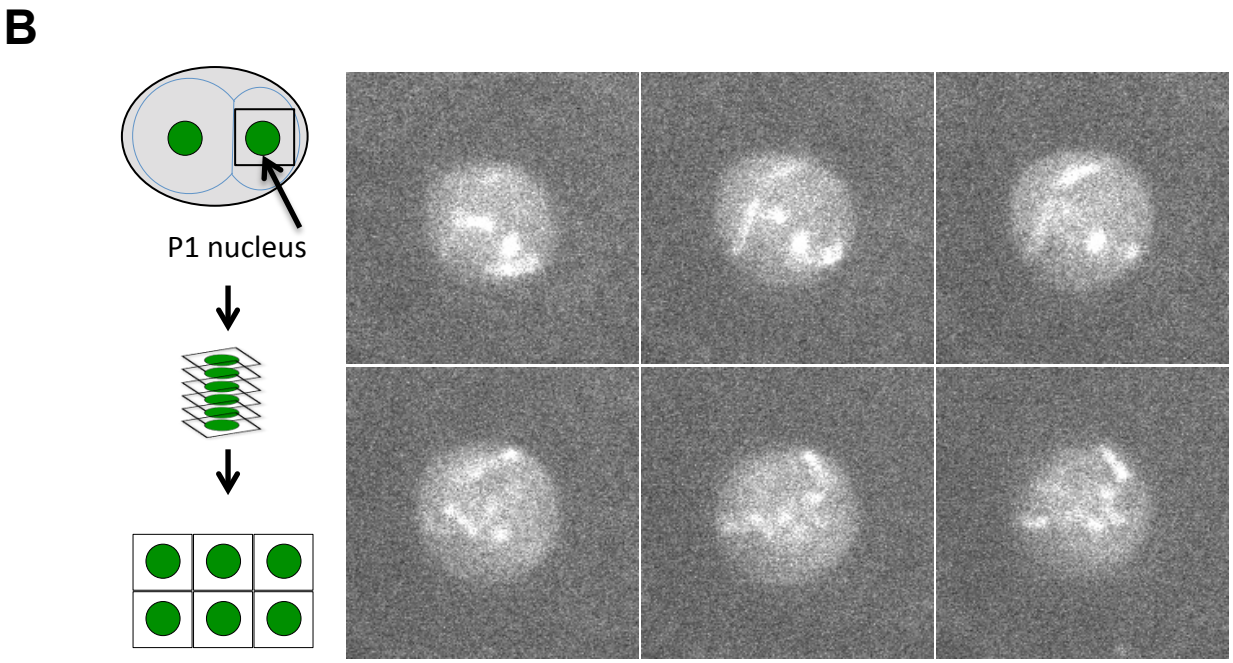
Confocal time-lapse imaging of early embryos expressing histone::GFP under the regulation of the *neg-1* 3' UTR.

Supplemental Figure 1

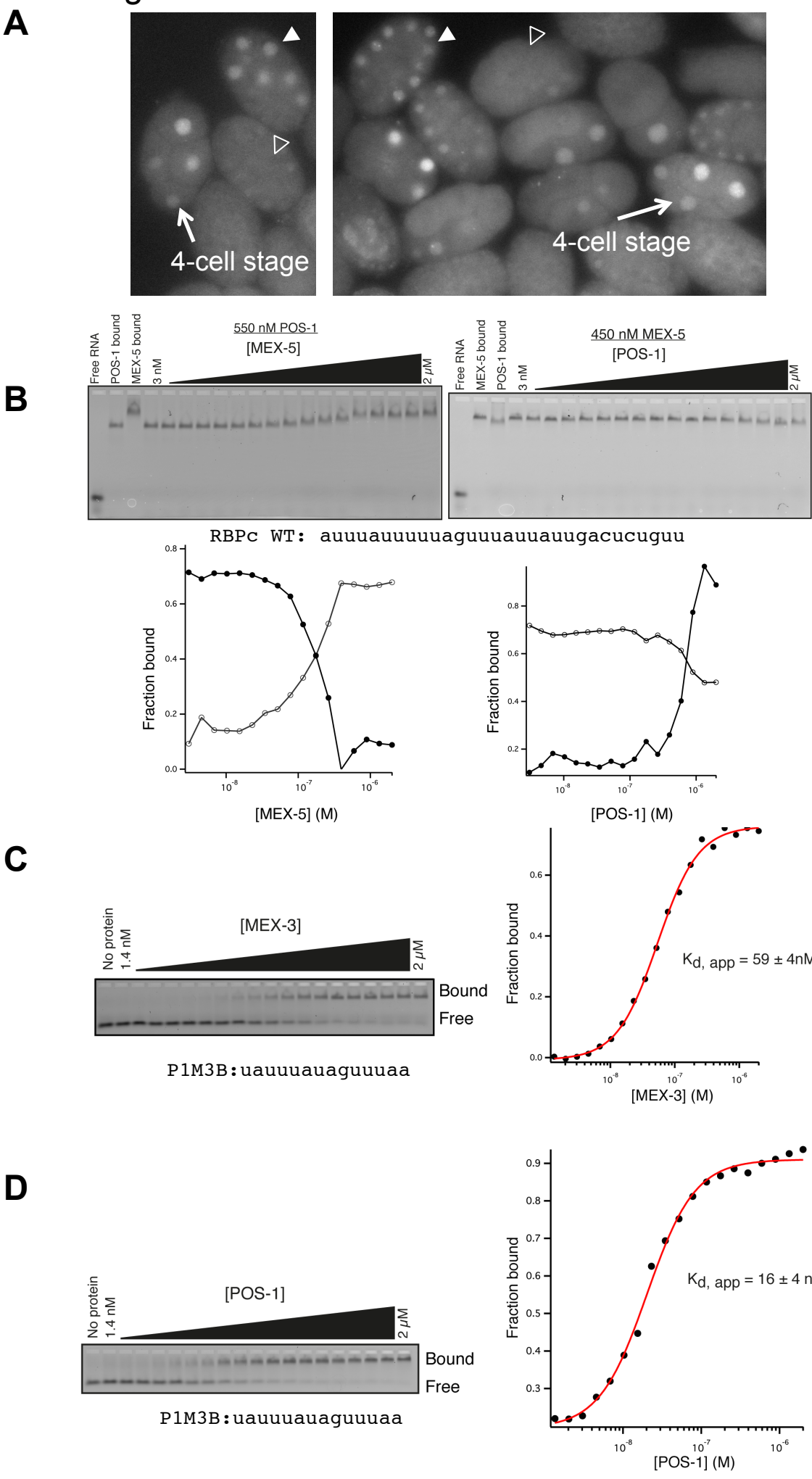


Supplemental Figure 2

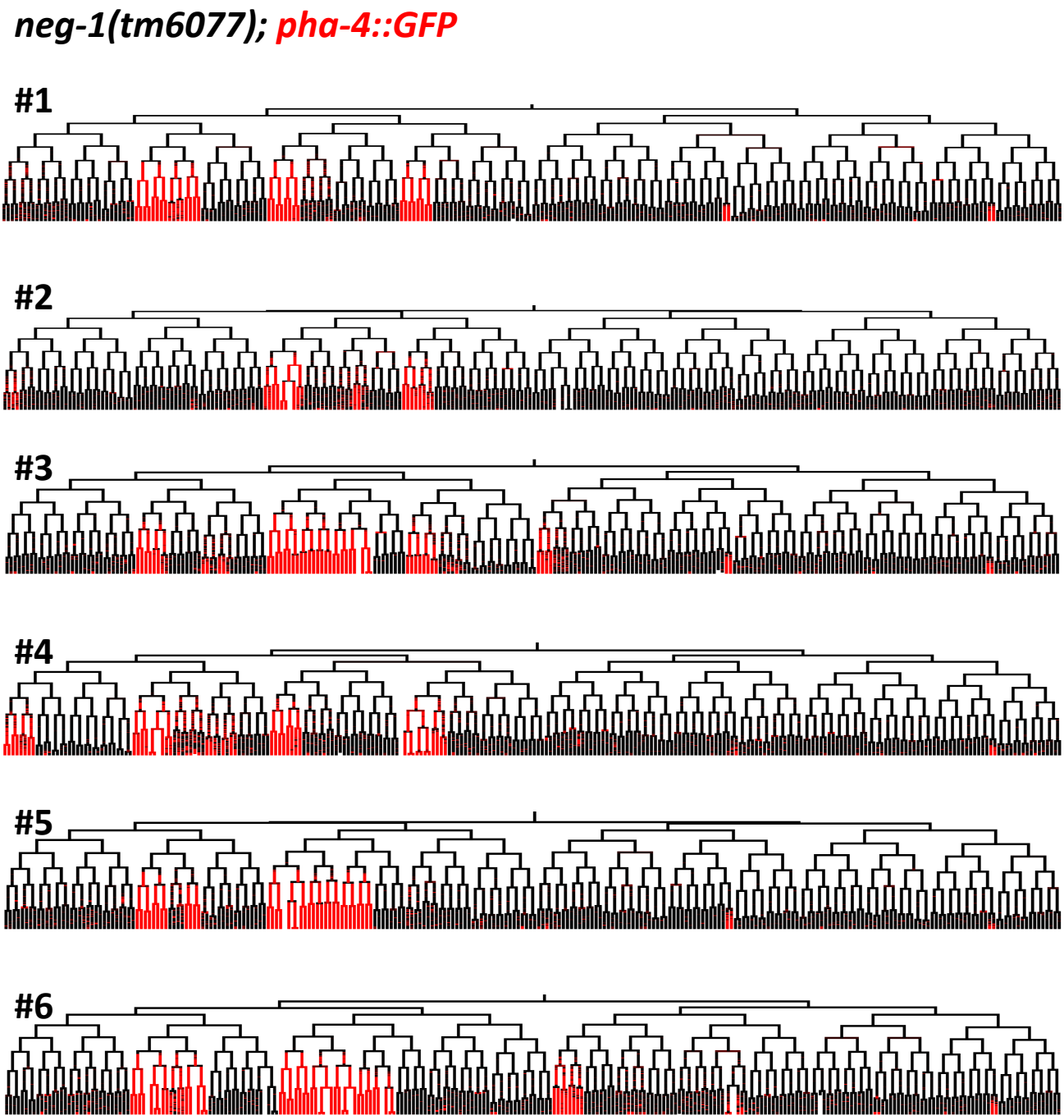




Supplemental Figure 3

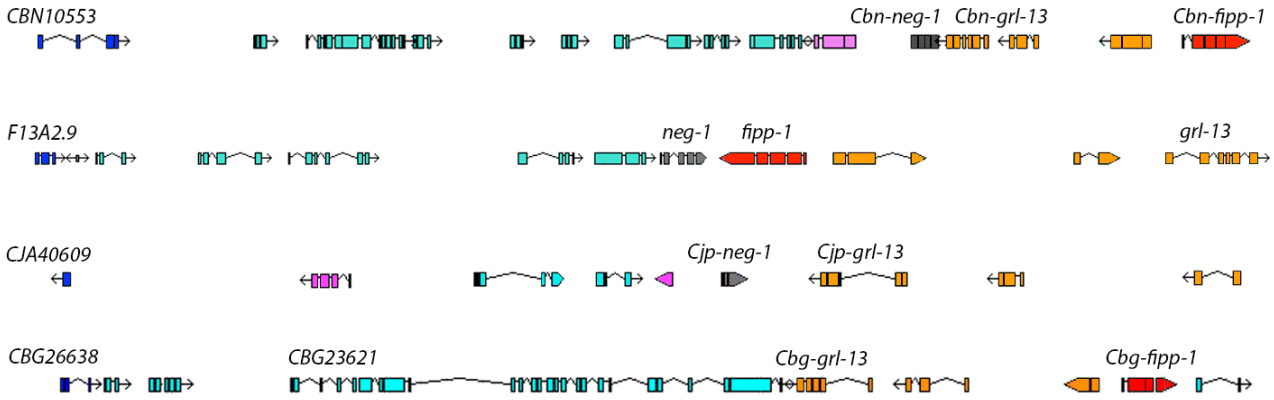


Supplemental Figure 4

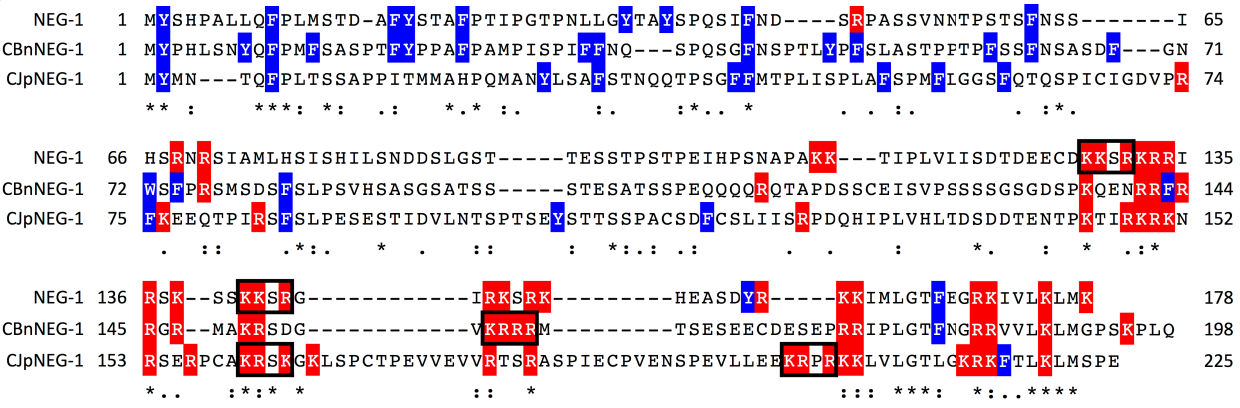


Supplemental Figure 5

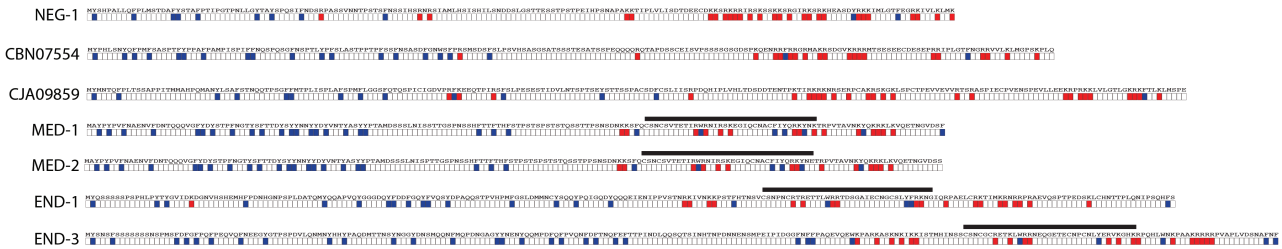
A



B



C



D

Protein	Length	N-half WYF	C-half WYF	N-half RK	C-half RK
NEG-1	176	9	2	3	24
Cbn-NEG-1	196	18	2	1	19
Cjp-NEG-1	223	12	2	3	23
MED-1	174	22	7	0	15
MED-2	174	22	6	0	14
END-1	221	16	5	1	19
END-3	242	22	9	0	21