

Supplemental Data for:

**Alternative polyadenylation directs tissue specific miRNA targeting in *Caenorhabditis elegans* somatic tissues**

Stephen M Blazie, Heather C Geissel, Henry Wilky, Rajan Joshi, Jason Newbern and Marco Mangone\*

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

**This PDF includes:**

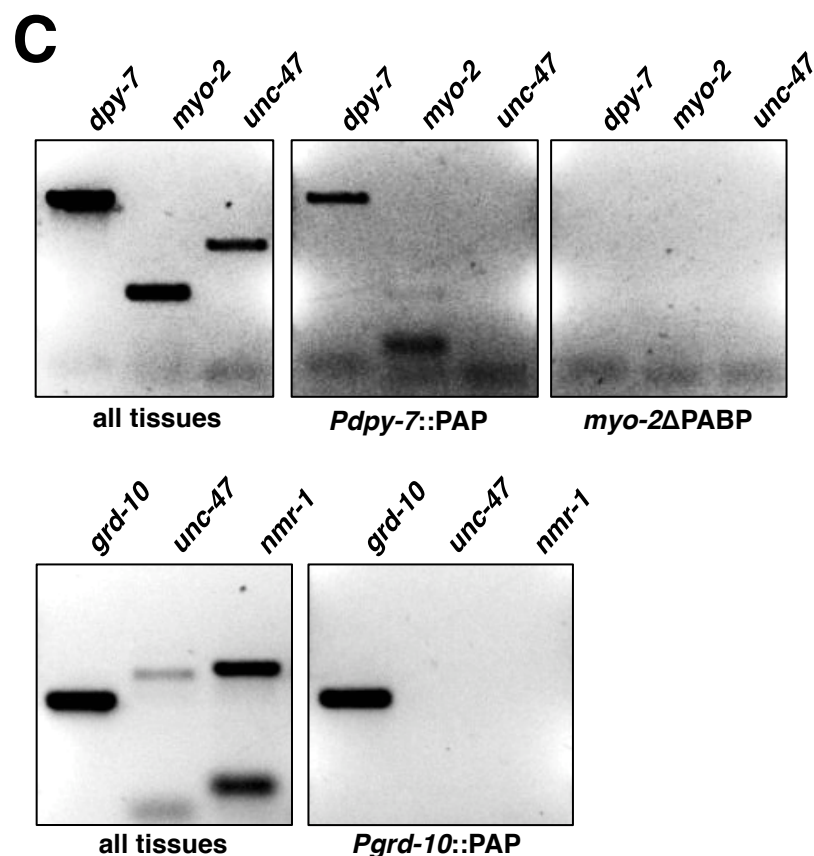
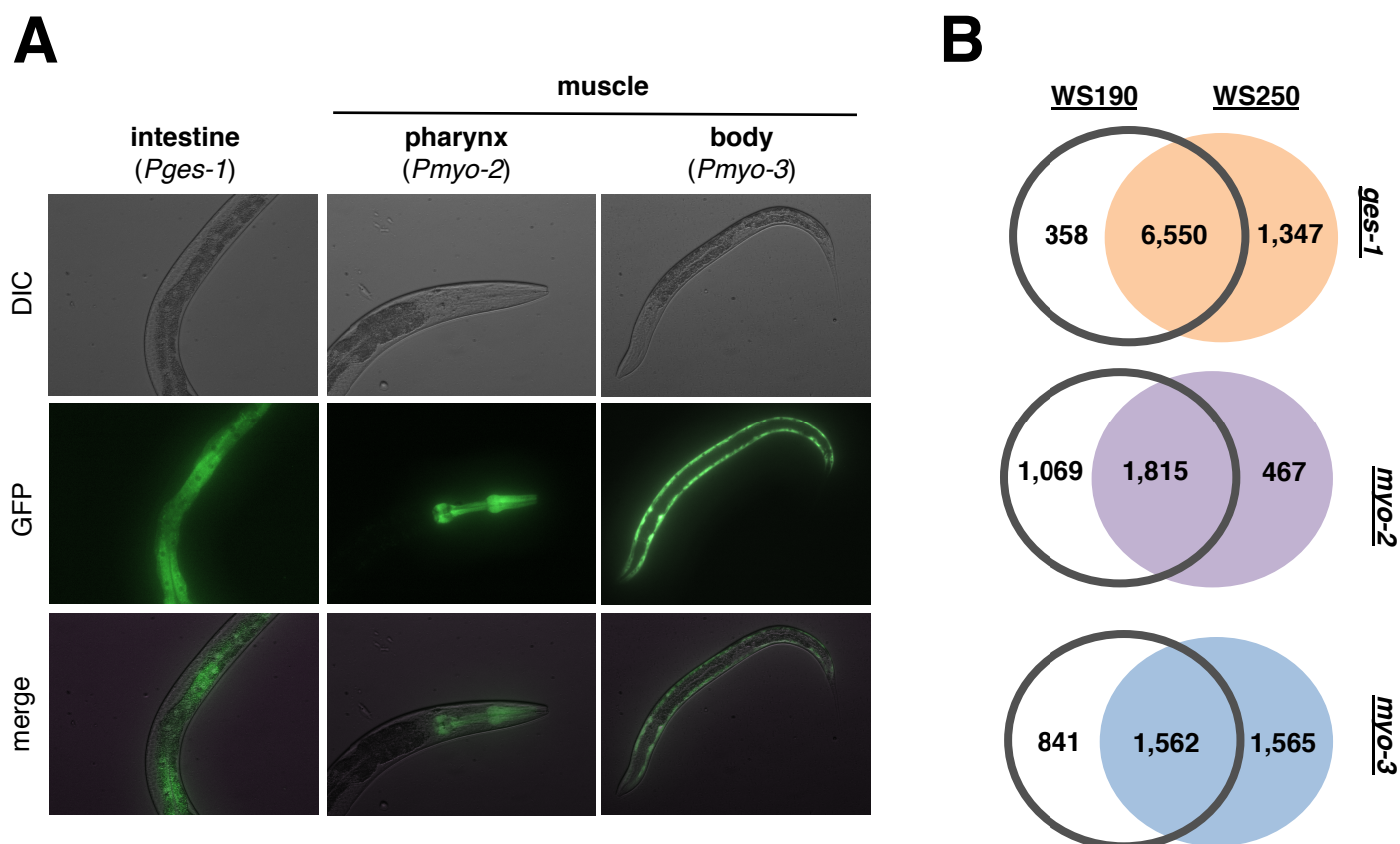
**Figures S1-S7**

**Table S1 - Summary of results from PAT-Seq after deep sequencing.**

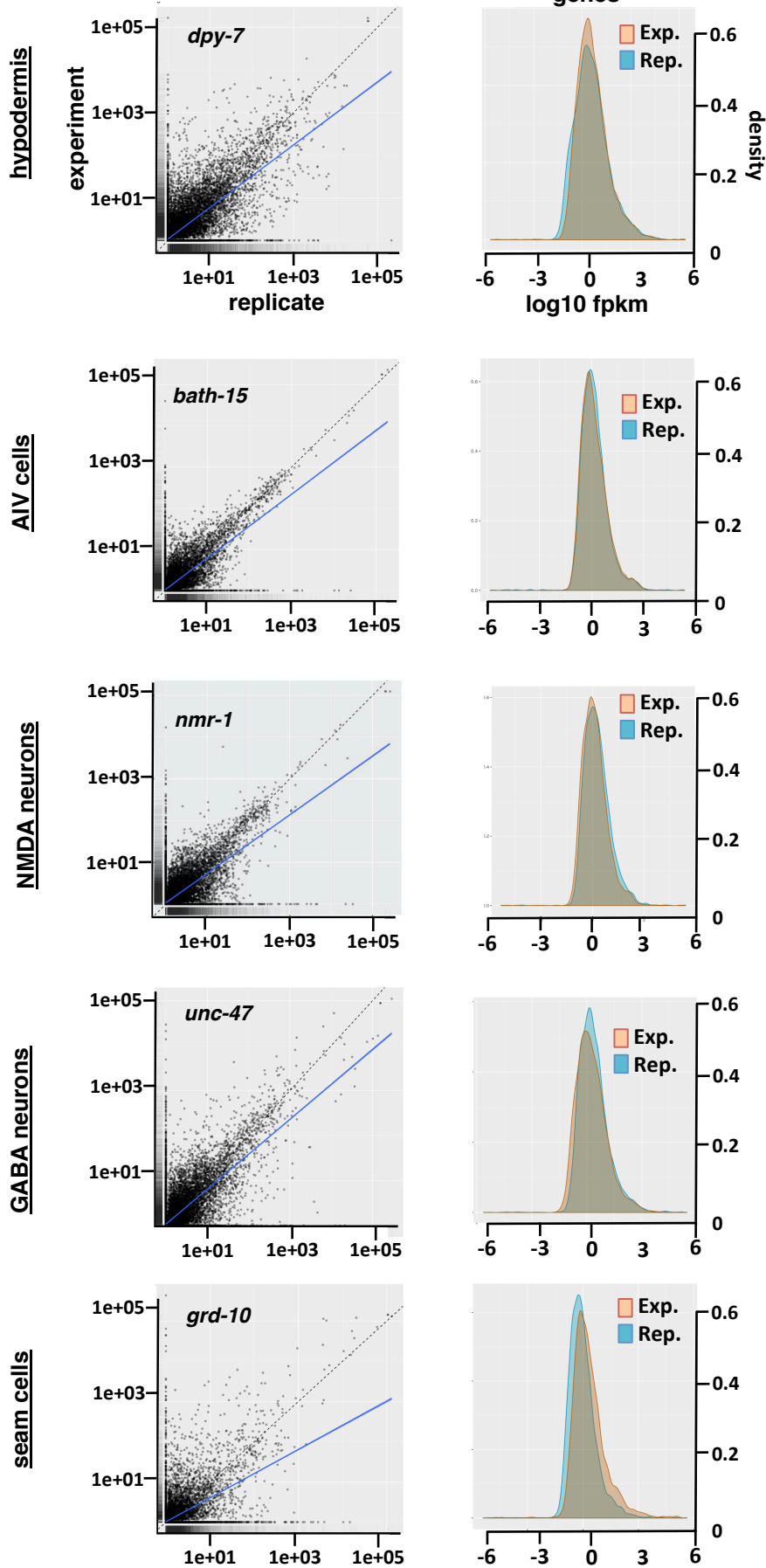
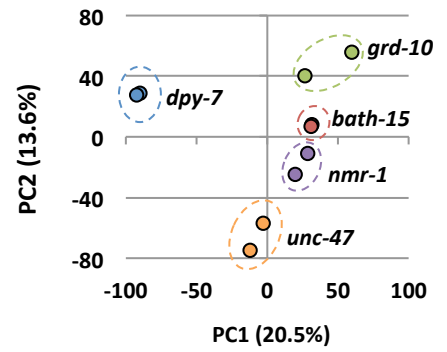
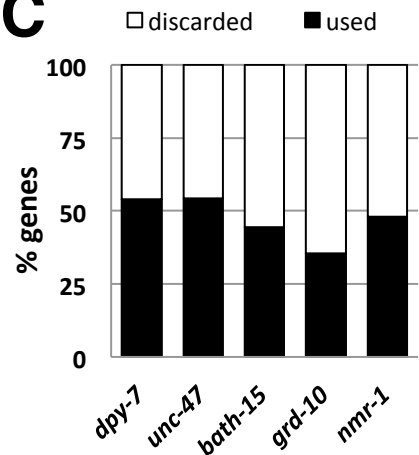
**Table S2 - Summary of sequencing results after mapping genes to WS250**

## Table of Contents

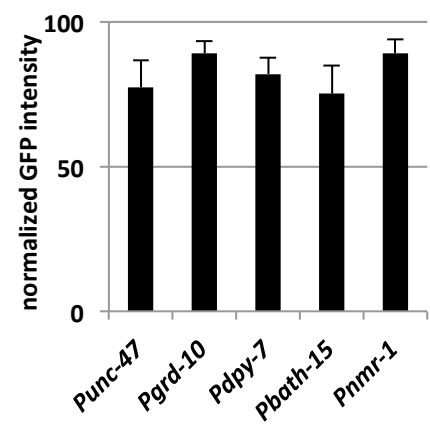
<b>Figure S1:</b> Poly(A)-Pull validation and comparison with WS250 .....	3
<b>Figure S2 :</b> PAT-Seq sequencing results.....	4
<b>Figure S3:</b> PAT-Seq comparison with other published mRNA-tagging derived datasets.....	6
<b>Figure S4:</b> 3'UTR length in each tissue and PAS usage in tissue-restricted versus commonly transcribed genes.....	7
<b>Figure S5:</b> APA drives loss of distinct miRNA family targets.....	8
<b>Figure S6:</b> <i>rack-1</i> and <i>tct-1</i> are expressed with a short 3'UTR isoform in the body muscle and long 3'UTR isoform in the intestine....	9
<b>Figure S7:</b> <i>rack-1 3'UTR ΔPAS1;ΔmiR-85</i> construct rescues GFP expression in <i>miR-85(n4117)</i> .....	10
<b>Table S1:</b> Summary of results from PAT-Seq after deep sequencing.....	11
<b>Table S2:</b> Summary of sequencing results after mapping genes to WS250.....	12



**Figure S1: Poly(A)-Pull validation and comparison with WS250.** **A)** Transgenic worms expressing the Poly(A)-Pull construct in intestine, pharynx, and body wall muscles. We previously applied PAT-Seq to study the transcriptomes of these three tissues (Blazie et al., 2015), which are now included in this study. **B)** Genes mapped in intestine and muscles from Blazie et al. 2015 using WS190 annotations compared to those we have now remapped to WS250. **C)** RT-PCR experiments validating the specificity of mRNA pulldown using Poly(A)-pull expressed in hypodermis (*dpy-7*) and seam cells (*grd-10*). We detected *dpy-7*, *myo-2*, and *unc-47* transcripts in total RNA from all tissues, while *dpy-7* is specifically enriched in mRNA prepared from *dpy-7::PAP* worms. The same transcripts were not detected in mRNA immunoprecipitated using our negative control construct lacking PABPC (*myo-2ΔPABP*). We detected seam cell specific *grd-10*, but not neuronal genes *unc-47* or *nmr-1* from mRNA immunoprecipitated from worms expressing Poly(A)-Pull in the seam cells.

**A****B****Figure S2****C****D****Extra chromosomal array transmission frequency**

strain	Transmission (n)
<i>Punc-47</i>	87.2% (86)
<i>Pgrd-10</i>	100% (102)
<i>Pdpy-7</i>	92.4% (79)
<i>Pbath-15</i>	100% (77)
<i>Pnmr-1</i>	97.67% (172)

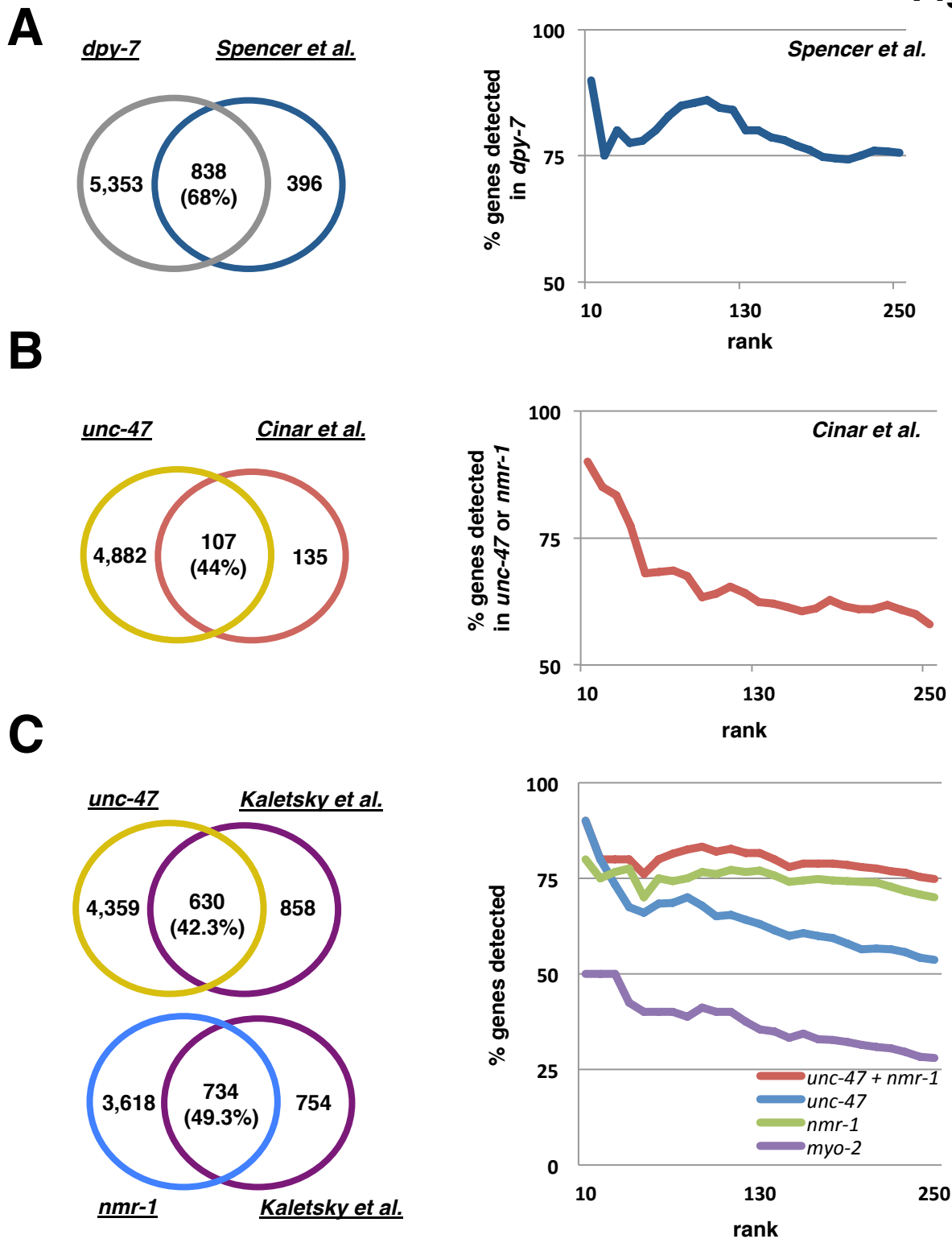
**E****GFP expression variation**



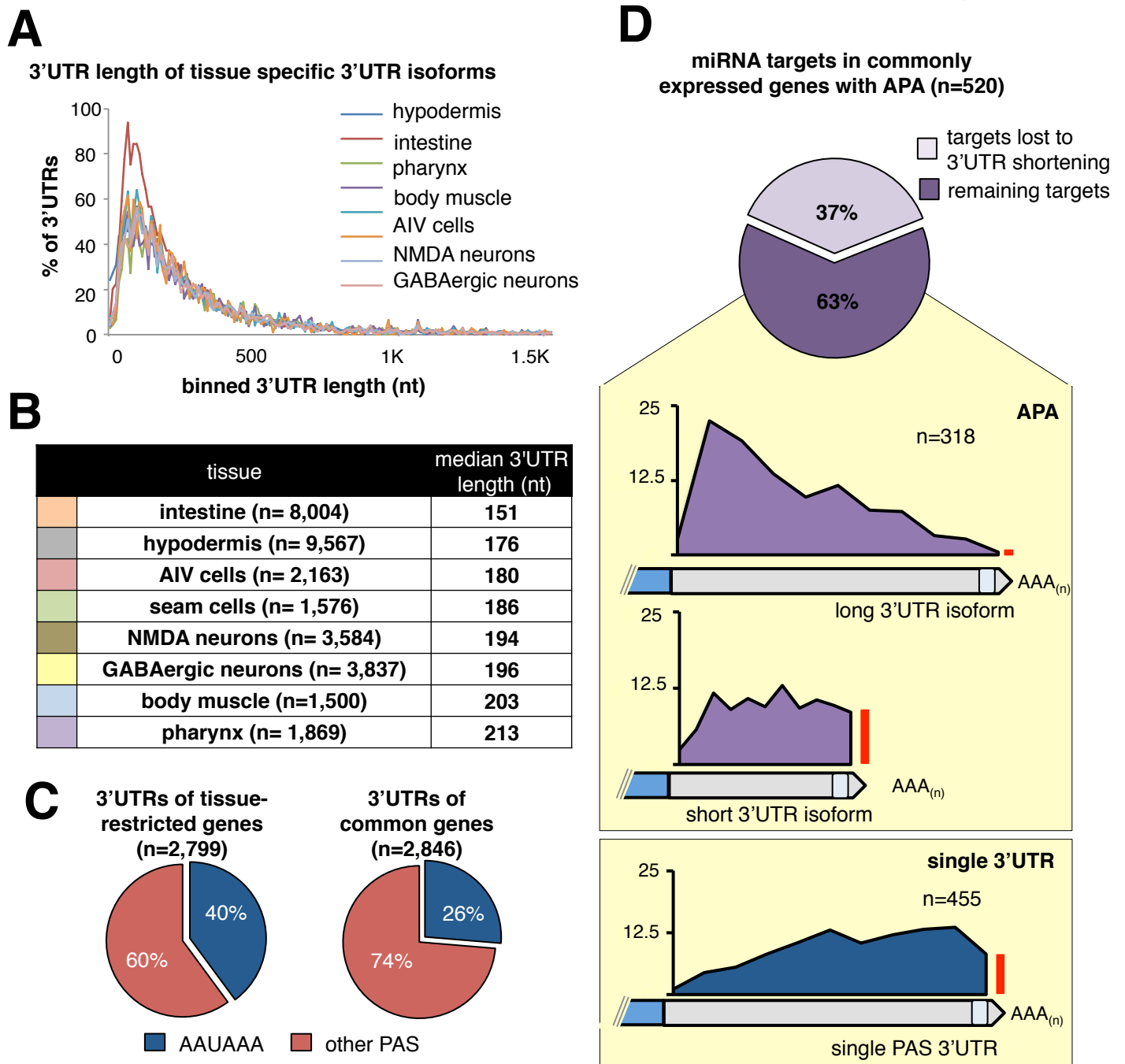
## Figure S2

**Figure S2: PAT-Seq sequencing results.** **A)** Scatter plot of mapped genes from each tissue dataset displayed by fpkm value detected in each replicate on a logarithmic ( $\log_{10}$ ) scale to highlight similarity of detection between replicates. *dpy-7* (hypodermis), *bath-15* (AIV cells), *nmr-1* (NMDA neurons), *unc-47* (GABAergic neurons) and *grd-10* (seam cells). The right panels show the distribution of the fpkm values for all genes in control and replicate biological samples for each tissue. The plots were generated using the cummeRbund package v. 2.0. **B)** Principal Component Analysis (PCA) shows high correlation among duplicates within our datasets. **C)** In this study we have used only the top ~40-50% positive hits produced by Cufflinks. **D)** The transmission frequency of the extra chromosomal arrays in all strains is higher than 87%. **E)** Quantification of GFP expression using imageJ analysis on five different animals from each strain shows minimal variation across animals within each line.

**Figure S3**



**Figure S4**



**Figure S4: 3'UTR length in each tissue and PAS usage in tissue-restricted versus commonly transcribed genes.** A) Histogram displaying the distribution of 3'UTR length in each tissue. The intestine expresses shorter 3'UTRs, on average. B) Median 3'UTR length for all genes expressed in each indicated tissue, sorted by ascending median length. Muscle tissue genes have a large median 3'UTR length. C) Pie charts displaying the proportion of 3'UTR isoforms having canonical PAS elements 'AAUAAA' or other PAS variants in tissue-restricted genes (left chart) or commonly transcribed genes (right chart). Commonly transcribed genes use non-canonical PAS elements more often than tissue-restricted genes. D) *Top*: Pie chart showing the portion of all miRNA targets in commonly transcribed genes that are lost due to expression of the short 3'UTR isoforms of each gene due to APA. Commonly transcribed genes lose ~43% of total miRNA targets due to tissue-specific APA. *Bottom*: We mapped the position of the 57% predicted miRNA targets in the longest and shortest 3'UTR isoforms (purple peaks), normalized by percent distance from the STOP codon. Many predicted miRNA targets overlap the PAS at the 3'-end of the transcript in the short 3'UTR isoform compared with the long isoform (compare height of red bars). The same analysis performed on a dataset of the same number of 3'UTRs of genes that do not undergo APA (single PAS 3'UTR, blue peak) reveals an enrichment of miRNA targets near the PAS (red bar) that is more similar to the short 3'UTR isoforms of commonly transcribed genes.

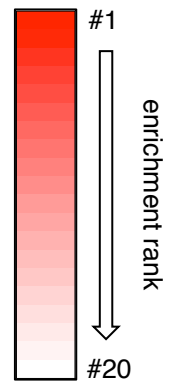
# Figure S5

predicted miRNA family targets enriched in each tissue transcriptome

		tissues							
		intestine	body muscle	pharynx	GABAergic neurons	NMDA neurons	seam cells	hypodermis	AIV cells
<b>1</b>	<b>1</b>	1	1	1	1	1	1	1	1
<b>2</b>	<b>4</b>	3	2	2	3	2	2	2	2
<b>3</b>	<b>2</b>	2	3	3	2	3	3	3	3
<b>4</b>	<b>3</b>	4	4	4	6	4	4	4	4
<b>5</b>	<b>6</b>	9	6	6	4	8	10	10	10
<b>6</b>	<b>7</b>	6	7	8	10	7	5	5	5
<b>7</b>	<b>8</b>	10	9	7	9	5	8	8	8
<b>8</b>	<b>12</b>	11	5	11	7	10	7	7	7
<b>9</b>	<b>10</b>	7	11	9	8	9	9	9	9
<b>10</b>	<b>5</b>	5	8	5	5	6	6	6	6
<b>11</b>	<b>9</b>	8	10	12	11	12	11	11	11
<b>12</b>	<b>13</b>	13	13	13	13	13	13	13	13
<b>13</b>	<b>11</b>	12	12	10	12	11	12	12	12
<b>14</b>	<b>15</b>	15	14	16	16	14	15	15	15
<b>15</b>	<b>14</b>	16	16	14	15	15	14	14	14
<b>16</b>	<b>16</b>	14	15	15	17	16	16	16	16
<b>17</b>	<b>17</b>	17	17	18	14	17	18	18	18
<b>18</b>	<b>18</b>	18	18	17	18	18	17	17	17
<b>19</b>	<b>20</b>	20	20	20	20	20	20	20	20
<b>20</b>	<b>19</b>	19	19	19	19	19	19	19	19

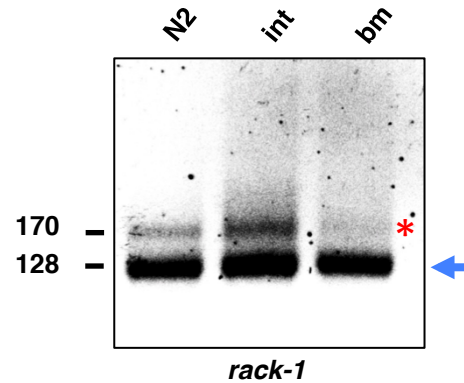
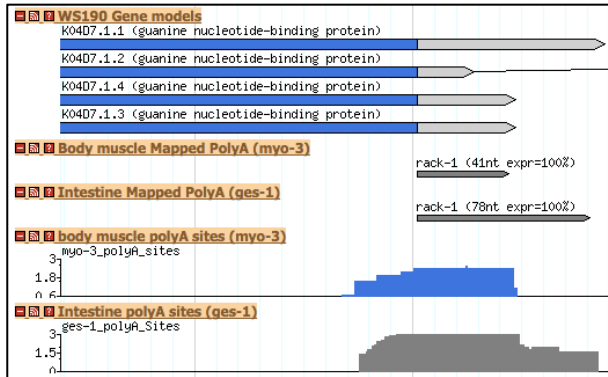
predicted miRNA family targets lost to APA

		tissues							
		intestine	body muscle	pharynx	GABAergic neurons	NMDA neurons	seam cells	hypodermis	AIV cells
<b>mir-2</b>	<b>18</b>	13	7	13	13	6	6	11	11
<b>mir-58</b>	<b>13</b>	20	14	14	16	13	4	15	15
<b>let-7</b>	<b>11</b>	19	6	12	5	15	18	14	14
<b>mir-44</b>	<b>12</b>	9	4	9	6	8	17	8	8
<b>mir-63</b>	<b>7</b>	8	15	2	4	10	5	10	10
<b>mir-50</b>	<b>2</b>	6	10	7	2	3	19	7	7
<b>mir-75</b>	<b>15</b>	7	13	18	11	17	20	9	9
<b>mir-1</b>	<b>19</b>	11	11	17	19	9	9	17	17
<b>mir-35</b>	<b>6</b>	12	2	10	3	5	10	4	4
<b>mir-51</b>	<b>20</b>	15	8	19	10	16	15	5	5
<b>mir-86</b>	<b>17</b>	14	17	5	9	7	13	13	13
<b>mir-72</b>	<b>4</b>	16	1	6	1	2	8	1	1
<b>mir-232</b>	<b>5</b>	3	5	3	8	1	1	3	3
<b>mir-87</b>	<b>9</b>	2	3	1	7	4	3	6	6
<b>mir-251</b>	<b>16</b>	17	18	15	12	14	2	16	16
<b>mir-49</b>	<b>1</b>	1	12	11	17	18	12	18	18
<b>mir-34</b>	<b>10</b>	18	9	8	14	12	14	19	19
<b>mir-46</b>	<b>8</b>	10	16	4	20	11	11	12	12
<b>lin-4</b>	<b>3</b>	4	20	20	15	19	7	2	2
<b>mir-231</b>	<b>14</b>	5	19	16	18	20	16	20	20

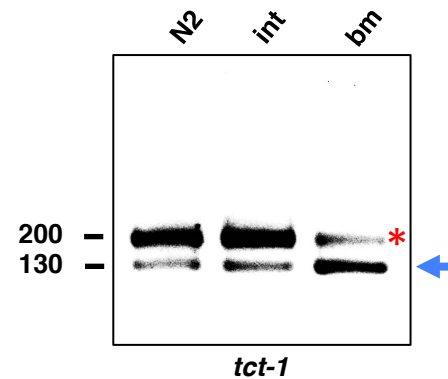
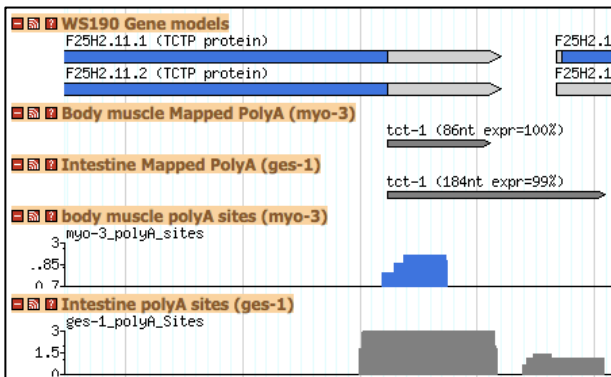


**Figure S5: APA drives loss of distinct miRNA family targets.** We ranked the overall abundance of PicTar and miRanda predicted targets for each of the twenty *C. elegans* miRNA families (Alvarez et. al, 2010) in genes expressed in each tissue transcriptome (left) and those lost due to 3'UTR shortening through APA in each tissue (right). miRNA families that are lost to APA are largely different than those that are most enriched in each tissue transcriptome overall. Predicted targets of the *mir-72*, *mir-232* and *mir-87* miRNA families are frequently lost among tissues (dotted box).

A

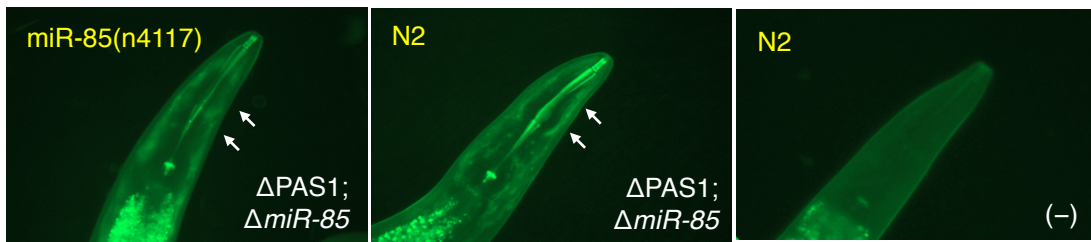


B

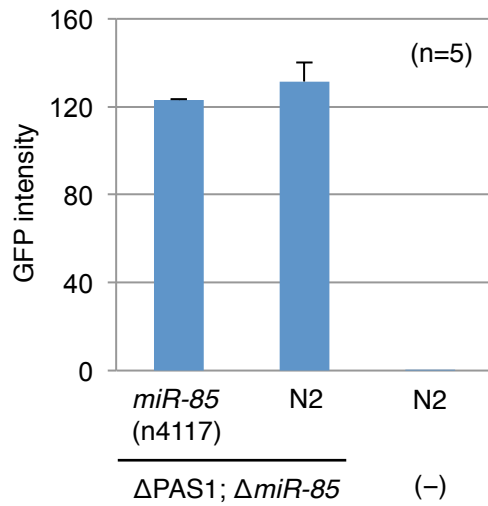


**Figure S6: *rack-1* and *tct-1* are expressed with a short 3'UTR isoform in the body muscle and long 3'UTR isoform in the intestine.** A) *Left*: diagram of clusters mapped for *rack-1* in body muscle (blue) and intestine (gray) tissues. *Right*: Results of 3'RACE experiments to amplify the 3'ends of *rack-1* in total RNA extracted from all tissues (N2), or mRNA samples prepared from intestine (int) or body muscle (bm) tissues using the PolyA-Pull immunoprecipitation (Blazie et. al, 2015). The long 3'UTR isoform of *rack-1* (red asterisk) was detected only in the intestine-specific mRNA prep, while only the short 3'UTR isoform (blue arrow) was detected in the body muscle. B) *Left*: diagram of clusters mapped for *tct-1* in body muscle (blue) and intestine (gray) tissues. *Right*: Results of 3'RACE experiments to amplify the 3'ends of *tct-1* in total RNA extracted from all tissues (N2), or mRNA samples prepared from intestine (int) or body muscle (bm) tissues using the PolyA-Pull immunoprecipitation (Blazie et. al, 2015). The long 3'UTR isoform of *tct-1* (red asterisk) is most abundant in the intestine mRNA prep, while only the short 3'UTR isoform (blue arrow) is instead most abundant in the body muscle.

**A**



**B**



**Figure S7: *rack-1* 3'UTR  $\Delta$ PAS1; $\Delta$ *miR-85* construct rescues GFP expression in *miR-85* (n4117).** A) We have injected our  $\Delta$ PAS1; $\Delta$ *miR-85* construct in the *miR-85* deletion strain *miR-85*(n4117), and in N2 *wild type* animals. As expected, this construct is able to rescue GFP expression in *miR-85* (n4117) similarly to what observed in Figure 5 Panel C iv. B) Quantification of GFP intensity from Panel A (n=5).

**Table S1**

samples (tissue)		total reads	mapped (%)	not mapped	average depth	
neurons	GABAergic neurons ( <i>Punc-47</i> )	<i>experiment</i>	7,412,370	4,997,158 (67.42)	2,415,212	34.3x
		<i>replicate</i>	13,834,418	7,031,857 (50.83)	6,802,561	32.8x
	NMDA-type neurons ( <i>Pnmr-1</i> )	<i>experiment</i>	7,048,055	5,171,858 (73.38)	1,876,197	30.1x
		<i>replicate</i>	6,526,023	4,241,115 (64.99)	2,284,908	21.9x
hypodermis	seam cells ( <i>Pgrd-10</i> )	<i>experiment</i>	6,563,296	4,852,499 (73.93)	1,710,797	32.3x
		<i>replicate</i>	9,456,923	2,406,168 (25.44)	7,050,755	79.9x
	hypodermis ( <i>Pdpy-7</i> )	<i>experiment</i>	31,509,607	10,288,238 (32.6)	21,221,369	44x
		<i>replicate</i>	14,090,409	6,222,500 (44.16)	7,867,909	31.4x
epithelial	AIV cells ( <i>Pbath-15</i> )	<i>experiment</i>	6,297,159	4,591,797 (72.92)	1,705,362	31.5x
		<i>replicate</i>	8,151,065	5,850,351 (71.77)	2,300,714	30.1x
muscle	pharynx ( <i>Pmyo-2</i> )	<i>experiment</i>	31,370,767	27,174,301 (86.6)	4,196,466	12.5x
		<i>replicate</i>	29,596,195	23,954,109 (80.1)	5,642,086	25.9x
	body muscle ( <i>Pmyo-3</i> )	<i>experiment</i>	30,993,699	24,557,134 (79.2)	6,436,565	12.6x
		<i>replicate</i>	33,770,647	27,760,544 (82.2)	6,010,103	9.9x
intestine	intestine ( <i>Pges-1</i> )	<i>experiment</i>	36,296,454	25,478,178 (70.2)	10,818,276	60.5x
		<i>replicate</i>	29,897,356	20,870,860 (69.8)	9,026,496	46.9x

**Table S1: Summary of results from PAT-Seq after deep sequencing.**

Raw reads derived from tissue-specific mRNA libraries on the Illumina Hi-Seq Instrument, mapped to the *C. elegans* WS250 genome annotation.

**Table S2**

	samples (tissue)		genes	
neurons	GABAergic neurons ( <i>Punc-47</i> )	<i>experiment</i>	4,399	3,011 <sup>(*)</sup>
		<i>replicate</i>	4,608	
	NMDA-type neurons ( <i>Pnmr-1</i> )	<i>experiment</i>	3,817	2,483 <sup>(*)</sup>
		<i>replicate</i>	4,153	
hypodermis	seam cells ( <i>Pgrd-10</i> )	<i>experiment</i>	3,051	1,208 <sup>(*)</sup>
		<i>replicate</i>	1,972	
	Hypodermis ( <i>Pdpy-7</i> )	<i>experiment</i>	5,393	3,786 <sup>(*)</sup>
		<i>replicate</i>	5,596	
epithelial	AIV cells ( <i>Pbath-15</i> )	<i>experiment</i>	3,539	2,112 <sup>(*)</sup>
		<i>replicate</i>	3,397	
muscle	pharynx ( <i>Pmyo-2</i> )	<i>experiment</i>	6,741	4,836 <sup>(*)</sup>
		<i>replicate</i>	6,166	
	body muscle ( <i>Pmyo-3</i> )	<i>experiment</i>	5,975	4,277 <sup>(*)</sup>
		<i>replicate</i>	6,158	
intestine	intestine ( <i>Pges-1</i> )	<i>experiment</i>	8,807	7,970 <sup>(*)</sup>
		<i>replicate</i>	9,092	

**Table S2: Summary of sequencing results after mapping genes to WS250.**

Mapped reads from the tissue-specific mRNA libraries on the Illumina Hi-Seq Instrument. Genes and isoforms are mapped to the *C. elegans* WS250 genome annotation. Genes and isoforms marked with an asterisk correspond to genes and isoforms enriched in both biological duplicates (fpkm>=1).