An updated *C. elegans* body muscle transcriptome for studies in muscle formation and function

Anna L. Schorr[1,2,3]

Email: aschorr@asu.edu

Alejandro Felix Mejia[3]

Email: Aflelixme@asu.edu

Martina Y. Miranda[4]

Email: MMiranda26@my.gcu.edu

Marco Mangone[1,2,3*]

* Corresponding author

Email: mangone@asu.edu

[1)] Molecular and Cellular Biology Graduate Program, Arizona State University, Tempe, AZ; [2)] Virginia G. Piper Center for Personalized Diagnostics, The Biodesign Institute at Arizona State University 1001 S McAllister Ave, Tempe, AZ; [3)] School of Life Sciences, Arizona State University, 751 E Lemon Mall, Tempe, AZ 85287; [4)] Helios Scholars at the Translational Genomics Research Institute, 445 N 5th St 4th Floor, Phoenix, AZ 85004

Running Title: Body muscle miRNAs in *C. elegans*

Keywords: miRNAs, 3'UTR, muscle tissue, interactome, *C. elegans*

ABSTRACT

The body muscle is an important tissue used in organisms for proper viability and locomotion. The contractile unit of the muscle is the sarcomere, which is ultimately responsible for the contraction reaction leading to movement. Although this tissue is generally well studied and characterized, and many pathways have been elucidated throughout the years, we still lack a comprehensive understanding of its transcriptome, and how it controls muscle development and function. Here, we have updated a nuclear FACS sorting-based approach to isolate and sequence a high-quality muscle transcriptome from *C. elegans* mixed stage animals. We have identified 2,848 muscle-specific protein-coding genes, including 78 transcription factors and 206 protein-coding genes containing an RNA binding domain. We studied their interaction network, performed a detailed promoter analysis, and identified novel muscle-specific *cis*-acting elements. We have also identified 16 high-quality muscle-specific miRNAs, studied their function *in vivo* using fluorochrome-based analyses, and developed the first high-quality miRNA Interactome in a living organism, incorporating other muscle-specific datasets produced by our lab and others.

Our study expands our understanding of how muscle tissue functions in *C. elegans* and in turn, provide results that can in the future be applied to humans to study muscular-related diseases.

INTRODUCTION

The body muscle is an important tissue used in organisms for proper viability and locomotion. Muscle tissue's structure, shape, and size change depending on its function and location. In humans, there are three main types of muscles: skeletal, smooth, and cardiac muscles, each responsible for a different form of contraction. Smooth and cardiac tissue contract involuntarily and are mainly located on the surface of organs. In contrast, skeletal muscles only contract voluntarily.

The contractile unit of the muscle is the sarcomere. It is composed of at least thirty different proteins, of which the most abundant are myosin and actin, and is ultimately responsible for the contraction reaction leading to movement.

The model organism *C. elegans* is ideal for studying muscle development and morphogenesis due to its well-characterized transcriptome, fully mapped cell lineage, and the availability of several methods for genetic engineering.

The physiology of its body muscle shares a high degree of similarity to the skeletal muscle of vertebrates with several mutually conserved genes (Lecroisey et al. 2007; Gieseler et al. 2017; Hrach et al. 2020; Ellwood et al. 2021b). In addition, *C. elegans* is a popular choice for studying disorders affecting humans, including muscular dystrophy (Hrach et al. 2020; Chaya et al. 2021; Ellwood et al. 2021a), since many human disease genes when introduced to worms phenocopy the symptoms (Markaki and Tavernarakis 2020; Natale et al. 2020; Chandler et al. 2021; Yue et al. 2021).

C. elegans possess only two large muscles: the pharynx and the body muscle. The pharynx is localized to the front of the animal and is responsible for food intake and physical crushing of the worm bacterial diet (Song and Avery 2013). It comprises eight

layers of non-striated muscles, surrounded by epithelial and neural tissue. The body muscle tissue comprises 95 striated muscle cells localized throughout the animal's body and is functionally equivalent to the vertebrate skeletal muscles (Gieseler et al. 2017). The muscle's basic functional unit is the sarcomere. It is highly conserved between nematodes and vertebrates and is composed of repeated units of a complex mesh of thin (I-band) and thick (A-band) filaments. Conversely, the thin filaments are mainly composed of actin. They anchor each sarcomere to the dense bodies and stabilize the entire structure. Instead, thick filaments are mostly made of myosin and are attached to the M-line, and mediate the tension generated during muscle contraction.

To properly develop and maintain functional body muscle tissue, gene expression must be strictly regulated. Our group and others in the past finely characterized the transcriptome of the *C. elegans* body muscle in the adult stage (Blazie et al. 2015; Blazie et al. 2017) and during the embryonic development (Fox et al. 2007), and identified important regulatory mechanisms used by muscle cells to maintain their homeostasis.

In addition to the identification of important genes in this tissue, our group, and others recently identified that tissue-specific alternative polyadenylation (APA), a poorly characterized mechanism that produces genes with different 3'untranslated region (3'UTR) isoforms, is an important regulatory mechanism used by cells to evade the negative regulation effected by microRNAs (miRNAs) (Blazie et al. 2015; Blazie et al. 2017). Still, the complexity and depth of this form of regulation are currently unknown. miRNAs are short ncRNAs that target regulatory elements in 3'UTRs, repressing gene expression in the cytoplasm (Lee et al. 1993; Reinhart et al. 2000).

Since their discovery, several groups, including our lab, have shown that miRNAs influence gene regulation considerably (Bartel 2018). The canonical miRNA biogenesis pathway begins with the transcription of pri-miRNAs in the nucleus, which fold into a hairpin structure (Bartel 2018). The microprocessor complex then cleaves the pri-miRNAs into a shorter stem-loop structure, the pre-miRNAs, which are then exported to the cytoplasm (Bartel 2018). Once in the cytoplasm the pre-miRNAs are cleaved into mature miRNAs which associate with an Argonaute protein as part of the RNA-induced silencing complex (RISC) (Bernstein et al. 2001; Kawamata and Tomari 2010; Bartel 2018). When loaded into the RISC, miRNAs bind to complementary target sequences within the 3'UTRs of genes to be repressed (Bartel 2018). This implies that the 3'UTRs and their isoforms are integral for miRNA-based regulation.

Several strategies have been developed to sequence tissue-specific transcriptomes and miRNAs in *C. elegans*. These methodologies utilize either fluorescence-activated cell sorting (FACS) (Fox et al. 2007; Haenni et al. 2012), immunoprecipitation (Blazie et al. 2015; Blazie et al. 2017; Brosnan et al. 2021), or RNA modification techniques (Alberti et al. 2018). Each of these approaches has unique advantages but also present several caveats. Methodologies utilizing FACS allow for high purity and stringency when isolating RNA from cells or nuclei. These approaches rely on tissue-specific promoters to drive fluorochromes expression in cells or in nuclei. After isolation, the fluorescent cells or nuclei may be readily sorted using FACS, then either sequenced (Haenni et al. 2012) or processed using microarrays (Fox et al. 2007).

Immunoprecipitation and antibody-based approaches have also been widely used to isolate the transcriptomes of specific tissues. This approach provides

considerably high yields of RNA compared to FACS-based methods and is generally much more cost-efficient.

Brosnan *et al*. (2021) used an epitope labeled Argonaute protein expressed in the specific *C. elegans* tissues to immunoprecipitate tissue-specific miRNAs populations from the body muscle, intestine, and neurons (Brosnan et al. 2021).

Alberti *et al*. (2018) developed a methodology named “microRNAome by methylation-dependent sequencing (Mime-seq)” to identify tissue-specific miRNA populations from complex animals at the level of single-cell specificity (Alberti et al. 2018). The Mime-seq approach expresses a transgenic methyltransferase (Ath-HEN1) from *Arabidopsis thaliana* in an animal system which allows the labeling of small RNAs. Mime-seq has been utilized in *C. elegans* and *Drosophila* and improved tissue-specific miRNA identification when compared with previous immunoprecipitation-based methods (Alberti et al. 2018).

Our group also used an immunoprecipitation-based approach to identify miRNA targets within the *C. elegans* intestine and body muscle (Kotagama et al. 2019), and observed transcripts were targeted differentially between the two tissues. While identifying the miRNA targets provided valuable information regarding tissue-specific gene regulation, a major limitation of this study was that we were unable to identify the specific miRNAs involved.

Here, we describe a novel nuclear FACS sorting-based approach, we named “Nuc-Seq,” to allow the isolation and sequencing of high-quality muscle transcriptomes and miRNA populations from *C. elegans* mixed stage animals. Using this method, we have identified 2,848 muscle-specific protein-coding genes, studied their interaction

network, performed a detailed promoter analysis, and identified novel muscle-specific *cis*-acting elements. Using fluorochrome-based analyses *in vivo*, we developed a high-quality muscle miRNA interactome, incorporating other muscle-specific datasets produced by our lab and others.

RESULTS

Nuc-Seq: an updated approach to identify muscle-specific transcriptome

We wanted to improve the annotation of the muscle transcriptome by developing a nuclear FACS-based strategy to isolate and sequence transcriptomes from *C. elegans* body muscle nuclei. We named this approach 'Nuc-Seq'. Our final goal was to identify the body muscle transcriptome and its miRNA population expressed explicitly in this tissue.

To test the feasibility and optimize this approach, we first decided to perform experiments using the *C. elegans* strain BN452, which ubiquitously expresses the mCherry fluorochrome fused to the histone H2B ortholog gene *his-58* (Gomez-Saldivar et al. 2016) (**Figure 1**). We fractionated this worm strain using mechanical stress and separated two fractions: one cytoplasmic and one enriched with mCherry fluorescent nuclei. (**Figure 1A-B and Materials and Methods**). We successfully tested these two fractions for the presence of known cytoplasmic and nuclear genes using a Western blot approach (**Figure 1C**). Importantly, these nuclei are stable over time with minimal degradation after several hours (**Supplemental Figure S1**). We then performed the FACS sorting step and successfully isolated a large pool of mCherry positive nuclei,

with over 34% of all the events being mCherry-positive (**Figure 1C**). We named this updated approach Nuclear Sequencing (Nuc-Seq).

With these results in our hands, we decided to move forward and limit the expression of the *mCherry*::*his-58* cassette in the body muscle tissue. We cloned the *his-58* gene, fused it to the mCherry fluorochrome, and forced its expression only in the body muscle tissue using the promoter region of the *C. elegans* ortholog of the myosin heavy chain gene (*myo-3*) (**Figure 2A upper panel**). As expected, the resultant transgenic worm restricts the expression of the mCherry fluorochrome in the body muscle nuclei (**Figure 2A lower panel**), with fewer fluorescent nuclei when compared to the experiments in the BN452 strain (**Figure 2A**). We then performed the FACS sorting step (**Figure 2B**). Although the number in sorted nuclei is significantly less than those obtained with the positive control BN452, we could still isolate a large population of mCherry-positive nuclei across from each of our replicates, which we then extracted and sequenced RNA from.

The identification of the C. elegans body muscle transcriptome

We performed three sequencing reactions in duplicates (technical triplicates), two from our body muscle-expressing nuclei (body muscle enriched), and one from nuclei isolated from BN452 (negative controls) (total of six samples). We obtained approximately 55M mappable reads for each sample (including our technical replicates). For most of the samples, we could map more than 97% of the total reads (**Supplemental Figure S2A**). The results obtained with our biological replicates correlate well (**Supplemental Figure S1B-C**).

We have identified 2,849 protein-coding genes in the *C. elegans* body muscle tissue, corresponding to ~14% of all *C. elegans* protein-coding genes (20,362 protein-coding genes; WS250) (**Fig. 3 and Supplemental Table S1**). While some of our top hits have not been characterized yet, as expected, we detected many known components of the sarcomere, such as many of the proteins that form the thick filaments (myosin heavy chain genes, *myo-3*, *unc-54*) and myosin light chain (*mlc-1*), and sarcomere organization and structural factors (*unc-15*, *unc-45*), as well as those that form the thin filaments, including several actin genes (*act-1, act-3, act-4*) and structural factors (*unc-60*, *unc-73*) (**Main Figure 3A and Supplemental Figure S3)**. We also detect orthologs of the tropomyosin gene *lev-11*, which in resting sarcomere blocks the binding of myosin to actin, and other notable genes previously known to be in the *C. elegans* body muscle tissue (**Main Figure 3A and Supplemental Figure S3)**. The genes in this tissue are highly interconnected (**Main Figure 3B)** and produce several distinguished sub-networks involved in body muscle formation and maintenance and protein production, and energy storage (**Main Figure 3B)**.

Our dataset identified 78 transcription factors, including *mdl-1*, *skn-1*, *xbp-1*, *mxl-3*, all previously known to be expressed in the body muscle tissue, and 207 non-ribosomal proteins containing an RNA binding domain **Supplemental Table S1**. Within this group, we identified K08D12.3, an orthologous of the human ZNF9, which is mutated in myotonic dystrophy type 2, *mbl-1*, a MUSCLEBLIND-type of mRNA splicing regulator required for muscle dense body organization, locomotion, and vulval morphogenesis, and *etr-1*, an ortholog of the human CUG-binding protein CUGBP1,

required for embryonic muscle development, and has been implicated in myotonic dystrophy **Supplemental Table S1**.

The C. elegans body muscle PROMOTERome

Next, we sought to study the core promoters of the genes identified in this study to pinpoint potential motifs transactivated by muscle-specific transcription factors. We have extracted 500 nucleotides upstream and 100 nucleotides downstream of the transcription start site of the 2,849 genes we mapped in this tissue, and plotted their nucleotide distribution (**Supplemental Figure S4**). We detected an enrichment of thymidine nucleotides just upstream of the transcription start site in these genes (**Supplemental Figure S4 Panel A**). Polythymine motifs (T-blocks) have been previously reported in *C. elegans* core promoters, and their presence correlates with gene expression levels (Grishkevich et al. 2011). We also performed a motif enrichment analysis in these core promoters and identified several novel *cis*-acting elements (**Supplemental Figure S4 Panel B** and **Supplemental Table S2**). We mapped 12 members of the human Krueppel-like transcription factors family (KLF), which are important regulators of gene expression in vertebrate development and are involved in muscle health and disease (Prosdocimo et al. 2015), Myogenin (MYOG), a muscle-specific basic-helix-loop-helix (bHLH) transcription factor involved in muscle development, myogenesis, and repair (Nabeshima et al. 1993), and NHLH1, a helix-loop-helix transcription factor that is involved in growth and development of a wide variety of tissues and species (Cogliati et al. 2002) and has been linked to muscle

growth and development (Miao et al. 2021). The complete list is shown in **Supplemental Table S2**.

The C. elegans body muscle miRNAs

We then aimed to detect and study the body muscle-specific miRNA population identified in our study. Using stringent parameters (**see Materials and Methods**), we identified 18 miRNAs expressed in the *C. elegans* body muscle tissue (**Main Figure 5A and Supplemental Table S1**). Several of these miRNAs have also been previously described to be expressed in this tissue by other groups (Martinez et al. 2008), and impairing their function either leads to defects in the muscular fiber formation or shows an uncoordinated phenotype (Miska et al. 2007; Alberti et al. 2018) (**Supplemental Table S1).** We validated the body muscle localization of some of these miRNAs using previously developed strains expressing GFP under the control of specific miRNA promoters (**Main Figure 4B**). Our most interconnected miRNA is *miR-5551*, a not characterized *C. elegans*-specific miRNA that is predicted to target 26 muscle-expressed protein-coding genes we detected in this tissue (**Main Figure 5A**). A GO term analysis shows that its targets control contractile fibers, sarcomere, and myofibril formation (**Supplemental Figure S5**). *Let-7* and the closely related *miR-241* are also present. These miRNAs temporally regulate larval development and have been previously identified in the pharyngeal and vulval muscles (Martinez et al. 2008; Andachi and Kohara 2016). Among the predicted target genes, we identified several genes in the DAF insulin pathway that control muscle health and aging (Mallick et al. 2020; Basu et al. 2021).

DISCUSSION

Here we have developed an updated FACS-based nuclear sorting method we named Nuc-Seq, to identify tissue-specific transcriptomes from cell lysates and applied it to determine the *C. elegans* muscle transcriptome. Nuc-Seq is robust, reproducible, has minimal background noise, and allows the simultaneous identification of both muscle-specific protein-coding genes and ncRNAs, such as miRNAs (**Main Figure 1**). When applied to the *C. elegans* muscle tissue, Nuc-Seq allowed the precise identification of 2,849 protein-coding genes, including 78 muscle transcription factors and 207 non-ribosomal proteins containing an RNA-binding domain. Taken together, the muscle transcriptome we identified covers ~14% of the entire *C. elegans* protein-coding transcriptome (**Main Figure 3**). Some of these genes, such as *unc-54*, *mlc-1*, *unc-27,* and others, have been previously identified in the muscle tissue and play an important role in muscle contraction and muscle health in general. We have identified almost all genes that form the sarcomere structure and are required for muscle contraction (**Supplemental Figure S3**). When clustered using previously published protein-protein interaction data (Remmelzwaal and Boxem 2019), the *C. elegans* muscle transcriptome is highly interconnected (**Main Figure 3**), with several notable subnetworks corresponding to mRNA processing, protein synthesis, ATP metabolism, and muscle cell development (**Main Figure 3B**), all consistent with the function of this tissue. We have also performed a promoter analysis, studying the promoter composition of the genes identified in this study (**Supplemental Figure S4**). We have found a Polythymine motif (T-blocks) directly upstream of the transcription start site and, using

prediction software (Meme Suite), we have identified many potential elements used by muscle-specific transcription factors to control gene expression in this tissue.

Our approach also allowed us to identify 18 high-quality muscle-specific miRNAs. Unfortunately, since our Nuc-Seq identifies nuclear pre-miRNAs, we can only identify the miRNAs, but not which strand (5p or 3p or both) is loaded in the Argonaute protein. Because of this limitation, one of the caveats of our miRNA prediction analysis is that we process only the 5p strands through the miRanda algorithm, ignoring the 3p strands. We also ignore miRNAs described in miRbase with less than 1,000 reads. Another stringent filter we used excluded the miRNAs located within introns of other genes (mirtron). The *C. elegans* WS250 release contains 257 miRNAs, including 118 mirtrons. We decided to remove the mirtrons because the density of reads within their genomic regions in genes with small introns made it too difficult to map them with high confidence. This step automatically removed 46% of all *C. elegans* miRNAs from our analysis and may explain why several body muscle miRNAs identified by others (Alberti et al. 2018) are not present in our list.

In addition, we used very stringent filters during the miRNA identification step using the miRanda algorithm, which may have also reduced the number of target predictions.

We believe that all these stringent filters and rules, while significantly reducing the number of miRNAs and their identified target genes, increase the quality of our hits.

We have validated the body muscle localization of several miRNAs identified in our study (**Main Figure 4B**).

The body muscle miRNA interactome produced by clustering our 16 high-quality miRNAs and the miRanda predictions using 3'UTRs data from our 2,849 protein-coding genes identified in this study, showed seven subnetworks. One large subnetwork contained eight miRNAs (**Main Figure 4)**. Most of the 88 genes in this large subnetwork are connected to single miRNAs, but exceptions occur with *gld-1*, *daf-16* and *lin-14* targeted by three miRNAs (*miR-241*, *let-7,* miR-392, and *miR-555*1), and *flh-1*, *ntl-9*, *rga-4*, *mboa-2*, *alh-7*, *lgg-2*, and *cab-1*, each connected to two miRNAs. More experiments need to be performed to validate these hits and understand how and why these genes are specifically regulated in the body muscle tissue.

In conclusion, we have produced an updated *C. elegans* body muscle transcriptome and miRNA Interactome, which will allow future studies to better understand the function of this tissue in normal states and diseases.

MATERIALS AND METHODS

Preparation of the transgenic strains: The *C. elegans* strain bqSi189 II; mel-28(bq5[GFP::mel-28]) III, or BN452, which ubiquitously expresses GFP and mCherry in all nuclei, was obtained from the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). The genomic DNA was extracted from mixed stage BN452 worms, and the mCherry-tagged *his-58* sequence was cloned using PCR using BP Gateway-flanked primers specific for the Gateway entry vector pDONR221. The *mCherry*::*his-58* sequence was successfully cloned using Gateway technology into the entry vector, pDONR221, as evidenced by sequencing data. To assemble the finalized clone, we performed a Gateway LR Clonase II plus reaction (cat. 12538-013;

Invitrogen) using the destination vector pCFJ150 (Frøkjær-Jensen *et al.* 2012) and entry clones containing the body-muscle-specific promoter *myo-3*, the *mCherry*::*his-58* sequence, and the *unc-54* 3′UTR as previously published (Blazie *et al.* 2017). The pCFJ150 construct and pCFJ601 (50 ng/µl), which contains a Mos1 transposon, were injected into the *C. elegans* strain EG6699 [ttTi5605 II; unc-119(ed3) III; oxEx1578] (Frøkjær-Jensen et al. 2012), which is designed for MosI-mediated single-copy integration (MosSCI), using standard injection techniques.

Nuclei extraction: Several medium plates of mixed stage worms were washed 4 times at 1,500 rpm for 3 min and resuspended in 10 mL of ice-cold NPB buffer (10 mM HEPES pH 7.6, 10 mM KCl, 1.5 mM MgCl2, 1 mM EGTA, 0.25 mM sucrose) per 250 µl volume of worm pellet. Nuclei were released by douncing 8-12 times in a chilled Wheaton stainless-steel tissue grinder (clearance 0.0005 inches = 12.5 µm) in 5 mL batches. The homogenate was sequentially filtered through 40 µm, 20 µm, and 10 µm nylon filters (pluriStrainer), then centrifuged for 10 min at 2,500 *g* at 4 °C to pellet the nuclei. The supernatant was aspirated, and the pellets containing nuclei were resuspended in 4 volumes of ice-cold NPB buffer. To remove larger debris, the resuspended nuclei pellets were centrifuged at 300 *g* for 1 min, then the supernatant containing the nuclei was transferred to a new tube on ice.

Nuclei stability: The nuclei were isolated from the *myo-3::mCherry::his-58::unc-54* strain as described above. Nuclei were imaged in triplicates using DAPI and mCherry filters using a Leica DMi8 inverted microscope over eight hours. Images were obtained using 1 s exposure times. The result of this experiment is shown in **Supplemental Figure S1**.

***FACS sorting**:* Two replicates of resuspended nuclei samples for the body muscle tissue were FACS isolated into TRIzol solution using a BD Biosciences FACSAria III cell sorter with a 70 µm nozzle, and a 560 nm laser with a temperature-controlled tube holder at 4°C. The Eppendorf tubes were filled until a 1:1 TRIzol reagent, and sorted nuclei prep was achieved after periodic gentle inversion. We set the gating parameters with 10,000 events from our positive control BN452 and negative control N2, and 30,000 events for the *myo-3::mCherry::his-58::unc-54* strain. We sorted 239,000 and 353,000 nuclei for the sample and the replicate for the set1 and 340,000 and 353,000 sorted nuclei for the sample and the replicate of the set2.

RNA extraction and library preparation: The RNA from FACS isolated nuclei was extracted using an RNA MiniPrep kit (ZR2070, Zymo Research) per the manufacturer's protocol. RNA was quantified by using a bioanalyzer. The cDNA libraries for the two replicates were generated using Nextera XT RNA sequencing. We prepared six mixed-stages cDNA libraries from the following worm strains: BN452 (two replicates), *myo-3::mCherry::his-58::unc-54* (four replicated). The libraries were depleted of ribosomal RNA contaminants using the Low Input RiboMinus Eukaryotic System v2 (ThermoFisher Scientific Catalog # A15027). The libraries were prepared from the ribosomal RNA depleted samples using the Nextflex Rapid Directional RNAseq Kit (PerkinElmer Catalog # NOVIA-5138-08)) as per the manufacturer's protocol and sequenced on the Nextseq 500 instrument (Illumina, San Diego, CA), using 75bp SE chemistry. The sequencing runs were performed at Girihlet Inc. Oakland (CA). We obtained between 50M to 60M of mappable reads across all six datasets.

Bioinformatics analysis: The FASTQ files corresponding to the six datasets and the controls (total of six datasets) were mapped to the *C. elegans* gene model WS250 using the Bowtie 2 algorithm (Langmead and Salzberg 2012) with the following parameters: --local -D 20 -R 3 -L 11 -N 1 -p 40 --gbar 1 -mp 3. The mapped reads were converted into a bam format and sorted using SAMtools software using standard parameters (Li et al. 2009). We processed ~285M reads obtained from all our datasets combined and obtained a median mapping success of ~93%. We used the Cufflinks software (Trapnell et al. 2010) to estimate the expression levels of the genes obtained in each dataset as per their BAM files. We calculated the fragment per kilobase per million bases (FPKM) number in each experiment and performed the comparison using the Cuffdiff algorithm (Trapnell et al. 2010). We have pooled the four experimental samples into two sets, named set1 and set2, and the two replicates using the BN452 into a new set named 'BN452 nuclear'. We used the median FPKM value >=5 in the set1 and set2 dataset as a threshold to define positive gene expression levels. The results are shown in **Supplemental Fig. S2 and Supplemental Table S1** using scores obtained by the Cuffdiff algorithm (Trapnell et al. 2010) and plotted using the CummeRbund package.

miRNA target identification: The predicted miRNA targeting network was constructed by extracting the longest 3'UTR sequence for each of the 2,848 protein-coding genes identified in our study. We converted the 3'UTR sequences to FASTA format and parsed the file using the miRanda algorithm (Enright et al. 2003) using a mature 5p *C. elegans* miRNA list containing the 18 miRNAs detected with an FPKM value >=1 (**Supplemental Table S1**), and using stringent parameters (-strict -sc -1.2). We only used miRNAs that have been previously detected in *C. elegans* with more than 1,000

reads in the miRbase database (www.mirbase.org) and that are not present in introns of protein-coding genes. The miRanda algorithm produced high-quality 141 predicted targets for 16 miRNAs. The networks were then built using the Cytoscape software (Shannon et al. 2003) and uploaded to the Network Analyst online software (Xia et al. 2015) to produce the images shown in **Main Figure 4**.

miRNA validation*:* To validate the RNA sequencing data shown in **Main Figure 4B**, three specific transgenic *C. elegans* strains were purchased from the CGC (VT1153, VL347, and VT1189). These strains drive GFP expression under the control of *let-7*, *miR-230,* or *miR-241* promoters respectively. Each strain was imaged using identical settings for brightfield and GFP expression. These transgenic strains were imaged using a Leica DMi8 Inverted microscope.

Western Blot experiments*:* The western blot validation experiments shown in **Main Figure 1C** were performed using total protein for each fraction, and the input was measured using a Bradford assay; 0.75 ng of protein was used for both the nuclear and cytoplasmic fractions, and 3 ng of protein was used for the input. Primary monoclonal antibodies for RFP (6G6-100, Chromotek) (1:1000) and GAPDH (ab125247, Abcam) (1:2000) were used, followed by IRDye 800CW goat: mouse secondary antibodies (LI-COR, 925–32210) (1:5000). The membrane was imaged using the ODYSSEY CLX system (LI-COR Biosciences, NE).

Imaging analysis: Confocal images of isolated nuclei pre-FACS stained with DAPI for the BN452 strain were acquired in the Biodesign Imaging Core, Division of the Arizona State University Bioimaging Facility. Fluorescent microscopy images of isolated nuclei

pre-FACS stained with DAPI for the *myo-3*::*MH58*::*unc-54* strain were acquired using a Leica DMi8 inverted microscope.

Network Analysis: The network shown in **Main Figure 3B** was constructed parsing the top 2,000 hits identified in this study using the STRING algorithm (v. 11.5) (Szklarczyk et al. 2021), run with standard parameters using only 'protein-protein interactions' as input. The produced network possesses 1,995 nodes and 31,449 edges, with an average node degree of 31.5 and an average local clustering coefficient of 0.356.

Promoter Analysis: We extracted 600 nt from the transcription start site for the top 100 genes identified in this study. We then used different custom Perl scripts to calculate the nucleotide distribution. The transcription factor predictions were produced by parsing these promoters to the Simple Enrichment Analysis script from the MEME suite software (Bailey et al. 2015). The results are shown in **Supplemental Figure S4** and **Supplemental Table S2**.

FIGURE LEGEND

Main Figure 1: Nuc-Seq schematics. A) Transgenic worms expressing mCherry fluorochrome fused to histone H2B (*his-58*) are (1) homogenized, (2) then subjected to several rounds of mechanical filtration. (3) The resultant nuclei are then collected and (4) subjected to FACS sorting. At the completion of the sorting step, (5) the mCherry-positive nuclei were incubated into TRIzol reagent and the recovered RNA is sequenced. B) Top panel: The strain BN452 ubiquitously expresses the mCherry::*his-58* transgene. Bottom Panel: mCherry positive nuclei recovered after Step 2 described in Panel A. (Red: mCherry, Blue:

DAPI). C) Western blot of nuclear and cytoplasmic fractions from BN452 worms blotted with α-RFP and α-GAPDH antibodies. D) Left Panel: nuclear FACS profile from N2 and BN452 nuclei. The red dots show the sorted mCherry-positive population. Right Panel: Pie chart showing the percentage of BN452 sorted mCherry-positive nuclei.

Main Figure 2: FACS sorting of body muscle nuclei: A: Top Panel: Schematics of the construct used to isolate the body muscle transcriptome. The body muscle-specific promoter *myo-3* drives the expression of the mCherry:::*his-58* cassette. Bottom Panel: Bright field and fluorescent images of the resultant transgenic worm strain. B) nuclear FACS profile from N2 and *myo3p::mCherry:::his-58::unc-54 3'UTR* nuclei. The red dots show the sorted mCherry-positive population. Right Panel: Pie chart showing the percentage of sorted mCherry-positive and body muscle-specific nuclei.

Main Figure 3: The *C. elegans* body muscle transcriptome. A) Left Panel: Heat map showing the body muscle expression levels log(fpkm) of the 20,362 protein-coding genes in WS250. (a) BN452 positive control, (m) muscle transcriptome. The muscle-specific transcriptome is highlighted with a dashed yellow line. Right Panel: several examples of expression levels in genes identified in the (a) BN452 and (m) muscle-specific nuclear fractions. The vertical axes mark expression levels as the log(FPKM). The muscle-specific genes *unc-54*, *mlc-1*, *unc-27,* and *lev-11* were all detected in our study, while the intestine-specific *elt-2* and the GABA neuronal-specific *unc-47* were not present. B) The *C. elegans* body muscle Interactome. We clustered all 2,849 protein-coding genes

identified in this study using previously published protein-protein interaction data. Each gene is shown as a pink dot. Within this network, we identified six smaller highly interconnected subnetworks shown in different colors. The genes in each subnetwork were independently subjected to GO term enrichment analyses shown with colored panels above each respective subnetwork.

Main Figure 4: The *C. elegans* body muscle Interactome. A) The 3'UTRs of the 2,849 protein-coding genes identified in this study were screened for miRNA targets in their 3'UTRs using 16 high-quality body-muscle miRNAs identified in this study. We identified seven subnetworks. Each miRNA is shown in grayscale and varies in size, depending on its number of predicted targets. Each target is shown in a different color, depending on the number of predicted miRNA targets in its 3'UTR, as from the miRanda algorithm. Red: one target, Light pink: two targets, Orange: three targets. B) transgenic *C. elegans* strains expressing GFP fluorochrome under the control of three miRNAs identified in this study. *Let-7* (VT1153), *miR-230* (VL347), and *miR-241* (VT1189). All three strains show strong body muscle expression in young adult worms.

Supplemental Materials:

Supplemental Figures S1-S4.

Supplemental Table S1: Comprehensive list of transcripts identified in this study.

Supplemental Table S2: Simple Enrichment Analysis of sequence motifs identified in promoters of transcripts identified in this study.

DATA AVAILABILITY

Raw reads were submitted to the NCBI Sequence Read Archive (http://trace.ncbi.nlm.nih.gov/Traces/sra/) with BioProject ID: PRJNA820874 and Submission ID: SUB11236444. The results of our analyses are available in Excel format as **Supplemental Table S1**.

AUTHOR CONTRIBUTION

ALS and MM designed the experiments. ALS developed and performed the nuclear FACS sorting experiments in collaboration with AFM and MYM, and ALS prepared the sequencing reactions. MM performed the bioinformatic analysis. ALS and MM analyzed the data, led the analysis and interpretation, assembled the figures, and wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This work is supported by NIH grant 1R01GM118796 to M.M. Some of the strains used in this study were provided by the CGC, funded by the NIH Office of Research Infrastructure Programs (P40 OD010440).

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

Alberti C, Manzenreither RA, Sowemimo I, Burkard TR, Wang J, Mahofsky K, Ameres SL, Cochella L. 2018. Cell-type specific sequencing of microRNAs from complex animal tissues. *Nat Methods* **15**: 283-289.

Andachi Y, Kohara Y. 2016. A whole-mount in situ hybridization method for microRNA detection in Caenorhabditis elegans. *RNA* **22**: 1099-1106.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Res* **43**: W39-49.

Bartel DP. 2018. Metazoan MicroRNAs. *Cell* **173**: 20-51.

Basu A, Behera S, Bhardwaj S, Dey S, Ghosh-Roy A. 2021. Regulation of UNC-40/DCC and UNC-6/Netrin by DAF-16 promotes functional rewiring of the injured axon. *Development* **148**.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363-366.

Blazie SM, Babb C, Wilky H, Rawls A, Park JG, Mangone M. 2015. Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. *BMC Biol* **13**: 4.

Blazie SM, Geissel HC, Wilky H, Joshi R, Newbern J, Mangone M. 2017. Alternative Polyadenylation Directs Tissue-Specific miRNA Targeting in Caenorhabditis elegans Somatic Tissues. *Genetics* **206**: 757-774.

Brosnan CA, Palmer AJ, Zuryn S. 2021. Cell-type-specific profiling of loaded miRNAs from Caenorhabditis elegans reveals spatial and temporal flexibility in Argonaute loading. *Nat Commun* **12**: 2194.

Chandler RJ, Cogo S, Lewis PA, Kevei E. 2021. Modelling the functional genomics of Parkinson's disease in Caenorhabditis elegans: LRRK2 and beyond. *Biosci Rep* **41**.

Chaya T, Patel S, Smith EM, Lam A, Miller EN, Clupper M, Kervin K, Tanis JE. 2021. A C. elegans genome-wide RNAi screen for altered levamisole sensitivity identifies genes required for muscle function. *G3 (Bethesda)* **11**.

Cogliati T, Good DJ, Haigney M, Delgado-Romero P, Eckhaus MA, Koch WJ, Kirsch IR. 2002. Predisposition to arrhythmia and autonomic dysfunction in Nhlh1-deficient mice. *Mol Cell Biol* **22**: 4977-4983.

Ellwood RA, Hewitt JE, Torregrossa R, Philp AM, Hardee JP, Hughes S, van de Klashorst D, Gharahdaghi N, Anupom T, Slade L et al. 2021a. Mitochondrial hydrogen sulfide supplementation improves health in the C. elegans Duchenne muscular dystrophy model. *Proc Natl Acad Sci U S A* **118**.

Ellwood RA, Piasecki M, Szewczyk NJ. 2021b. Caenorhabditis elegans as a Model System for Duchenne Muscular Dystrophy. *Int J Mol Sci* **22**.

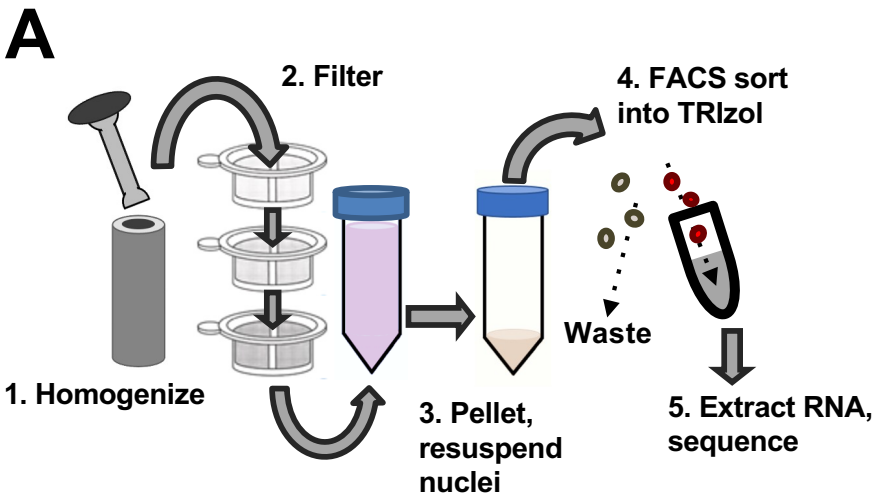
Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. 2003. MicroRNA targets in Drosophila. *Genome Biol* **5**: R1.

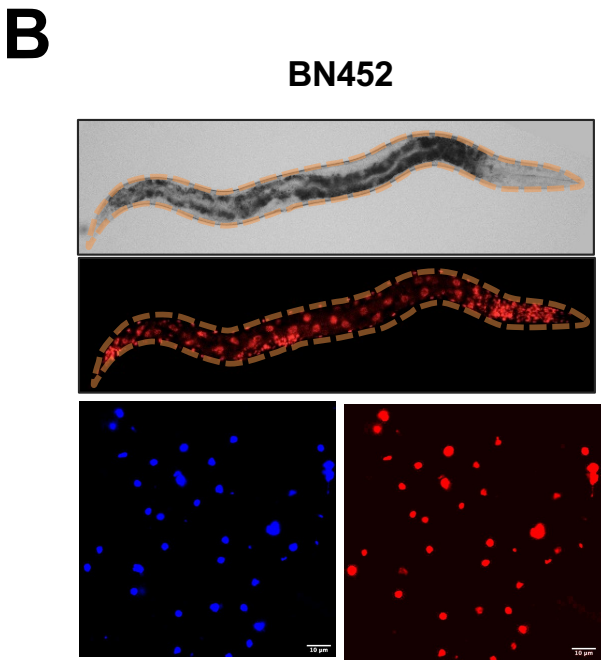
Fox RM, Watson JD, Von Stetina SE, McDermott J, Brodigan TM, Fukushige T, Krause M, Miller DM, 3rd. 2007. The embryonic muscle transcriptome of Caenorhabditis elegans. *Genome Biol* **8**: R188.

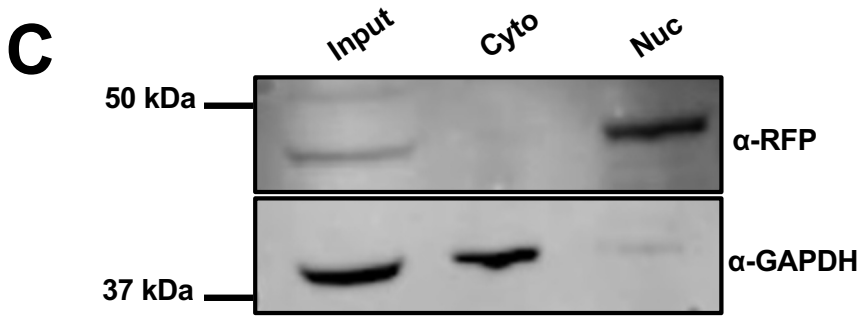
Gieseler K, Qadota H, Benian GM. 2017. Development, structure, and maintenance of C. elegans body wall muscle. *WormBook* **2017**: 1-59.

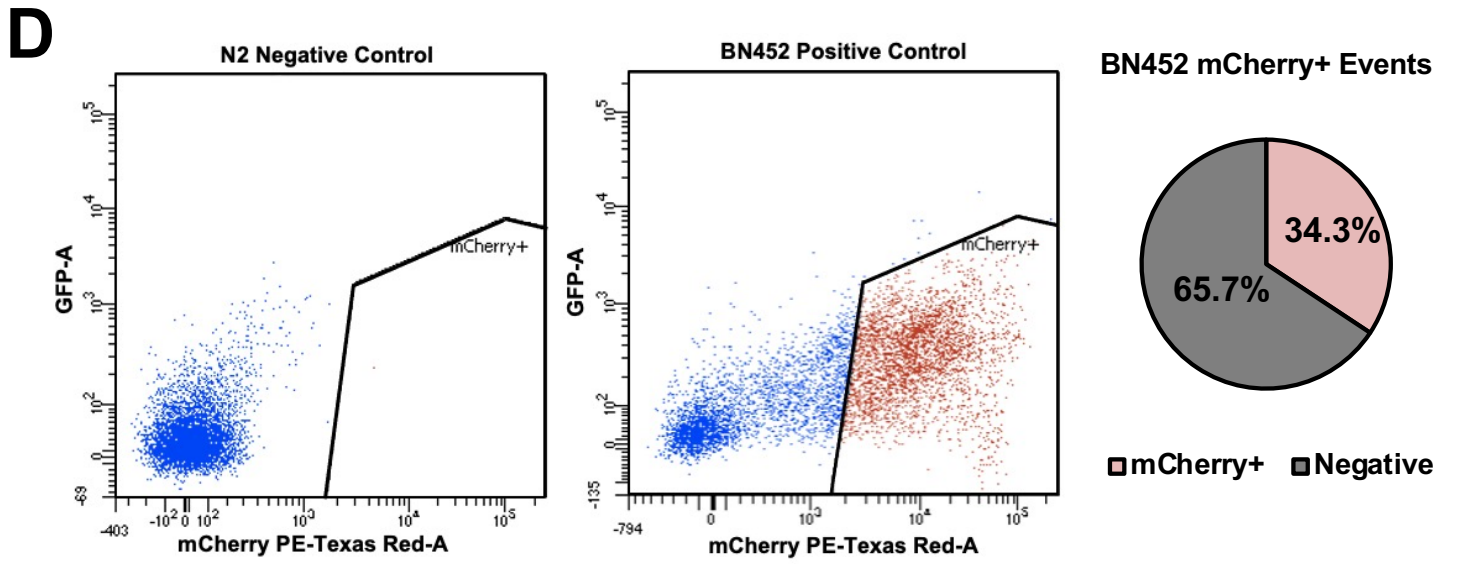
543 Gomez-Saldivar G, Fernandez A, Hirano Y, Mauro M, Lai A, Ayuso C, Haraguchi T, Hiraoka Y,
544 Piano F, Askjaer P. 2016. Identification of Conserved MEL-28/ELYS Domains with
545 Essential Roles in Nuclear Assembly and Chromosome Segregation. *PLoS Genet* **12**:
546 e1006131.
547 Grishkevich V, Hashimshony T, Yanai I. 2011. Core promoter T-blocks correlate with gene
548 expression levels in C. elegans. *Genome Res* **21**: 707-717.
549 Haenni S, Ji Z, Hoque M, Rust N, Sharpe H, Eberhard R, Browne C, Hengartner MO, Mellor J, Tian
550 B et al. 2012. Analysis of C. elegans intestinal gene expression and polyadenylation by
551 fluorescence-activated nuclei sorting and 3'-end-seq. *Nucleic Acids Res* **40**: 6304-6318.
552 Hrach HC, O'Brien S, Steber HS, Newbern J, Rawls A, Mangone M. 2020. Transcriptome changes
553 during the initiation and progression of Duchenne muscular dystrophy in Caenorhabditis
554 elegans. *Hum Mol Genet* **29**: 1607-1623.
555 Kawamata T, Tomari Y. 2010. Making RISC. *Trends Biochem Sci* **35**: 368-376.
556 Kotagama K, Schorr AL, Steber HS, Mangone M. 2019. ALG-1 Influences Accurate mRNA Splicing
557 Patterns in the Caenorhabditis elegans Intestine and Body Muscle Tissues by Modulating
558 Splicing Factor Activities. *Genetics* **212**: 931-951.
559 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:
560 357-359.
561 Lecroisey C, Segalat L, Gieseler K. 2007. The C. elegans dense body: anchoring and signaling
562 structure of the muscle. *J Muscle Res Cell Motil* **28**: 79-87.
563 Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small
564 RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-854.
565 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
566 Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and
567 SAMtools. *Bioinformatics* **25**: 2078-2079.
568 Mallick A, Ranawade A, van den Berg W, Gupta BP. 2020. Axin-Mediated Regulation of Lifespan
569 and Muscle Health in C. elegans Requires AMPK-FOXO Signaling. *iScience* **23**: 101843.
570 Markaki M, Tavernarakis N. 2020. Caenorhabditis elegans as a model system for human
571 diseases. *Curr Opin Biotechnol* **63**: 118-125.
572 Martinez NJ, Ow MC, Reece-Hoyes JS, Barrasa MI, Ambros VR, Walhout AJ. 2008. Genome-scale
573 spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. *Genome*
574 *Res* **18**: 2005-2015.
575 Miao W, Ma Z, Tang Z, Yu L, Liu S, Huang T, Wang P, Wu T, Song Z, Zhang H et al. 2021.
576 Integrative ATAC-seq and RNA-seq Analysis of the Longissimus Muscle of Luchuan and
577 Duroc Pigs. *Front Nutr* **8**: 742672.
578 Miska EA, Alvarez-Saavedra E, Abbott AL, Lau NC, Hellman AB, McGonagle SM, Bartel DP,
579 Ambros VR, Horvitz HR. 2007. Most Caenorhabditis elegans microRNAs are individually
580 not essential for development or viability. *PLoS Genet* **3**: e215.
581 Nabeshima Y, Hanaoka K, Hayasaka M, Esumi E, Li S, Nonaka I, Nabeshima Y. 1993. Myogenin
582 gene disruption results in perinatal lethality because of severe muscle defect. *Nature*
583 **364**: 532-535.
584 Natale C, Barzago MM, Diomede L. 2020. Caenorhabditis elegans Models to Investigate the
585 Mechanisms Underlying Tau Toxicity in Tauopathies. *Brain Sci* **10**.

586 Prosdocimo DA, Sabeh MK, Jain MK. 2015. Kruppel-like factors in muscle health and disease.
587 *Trends Cardiovasc Med* **25**: 278-287.
588 Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G.
589 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis
590 elegans. *Nature* **403**: 901-906.
591 Remmelzwaal S, Boxem M. 2019. Protein interactome mapping in Caenorhabditis elegans. *Curr*
592 *Opin Syst Biol* **13**: 1-9.
593 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T.
594 2003. Cytoscape: a software environment for integrated models of biomolecular
595 interaction networks. *Genome Res* **13**: 2498-2504.
596 Song BM, Avery L. 2013. The pharynx of the nematode C. elegans: A model system for the study
597 of motor control. *Worm* **2**: e21833.
598 Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T,
599 Bork P et al. 2021. The STRING database in 2021: customizable protein-protein
600 networks, and functional characterization of user-uploaded gene/measurement sets.
601 *Nucleic Acids Res* **49**: D605-D612.
602 Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ,
603 Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals
604 unannotated transcripts and isoform switching during cell differentiation. *Nat*
605 *Biotechnol* **28**: 511-515.
606 Xia J, Gill EE, Hancock RE. 2015. NetworkAnalyst for statistical, visual and network-based meta-
607 analysis of gene expression data. *Nat Protoc* **10**: 823-844.
608 Yue Y, Li S, Shen P, Park Y. 2021. Caenorhabditis elegans as a model for obesity research. *Curr*
609 *Res Food Sci* **4**: 692-697.
610

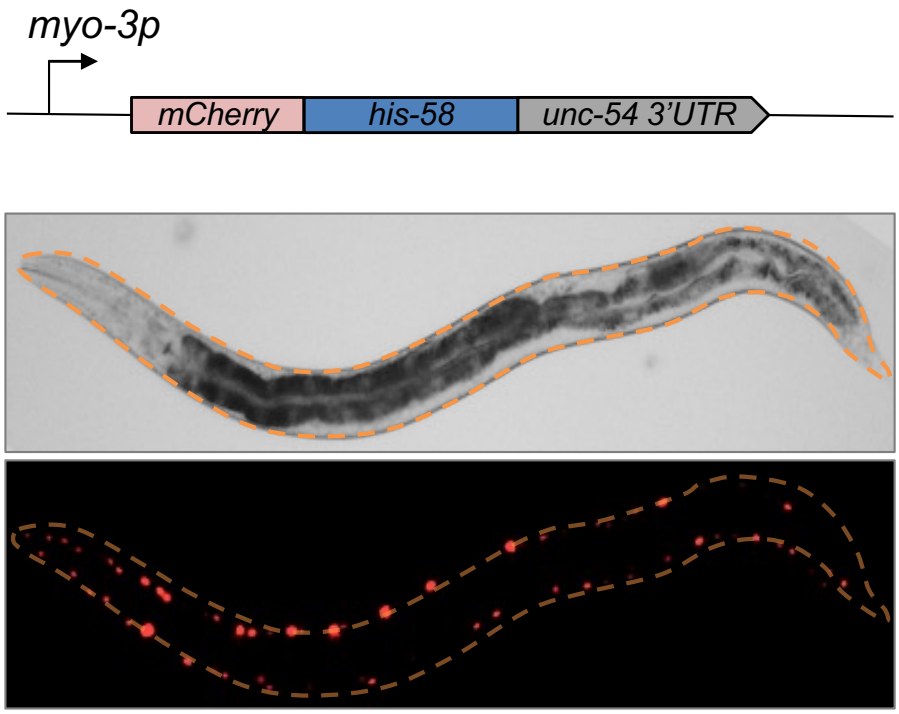

A
2. Filter
4. FACS sort into TRIzol
Waste
1. Homogenize
3. Pellet, resuspend nuclei
5. Extract RNA, sequence


B
BN452

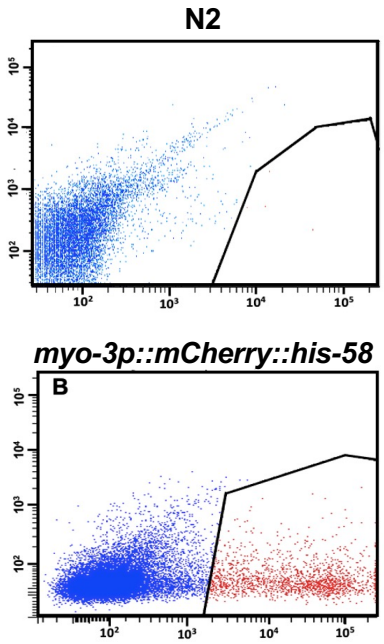

C
Input
Cyto
Nuc
50 kDa
α-RFP
37 kDa
α-GAPDH

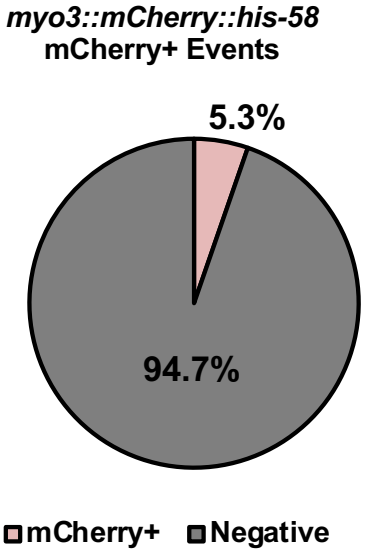

D
N2 Negative Control
GFP-A
mCherry+
mCherry PE-Texas Red-A
BN452 Positive Control
GFP-A
mCherry+
mCherry PE-Texas Red-A
BN452 mCherry+ Events
34.3%
65.7%
mCherry+
Negative

A


myo-3p
mCherry
his-58
unc-54 3'UTR

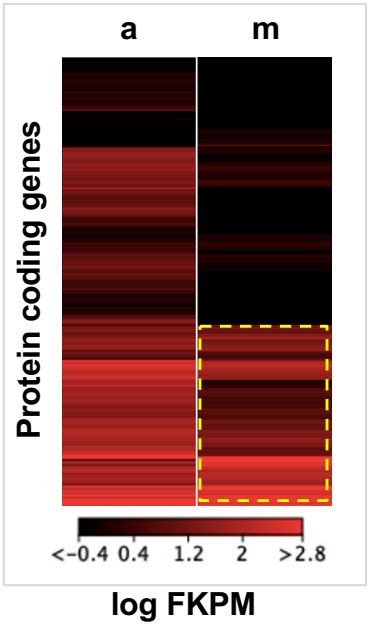
B

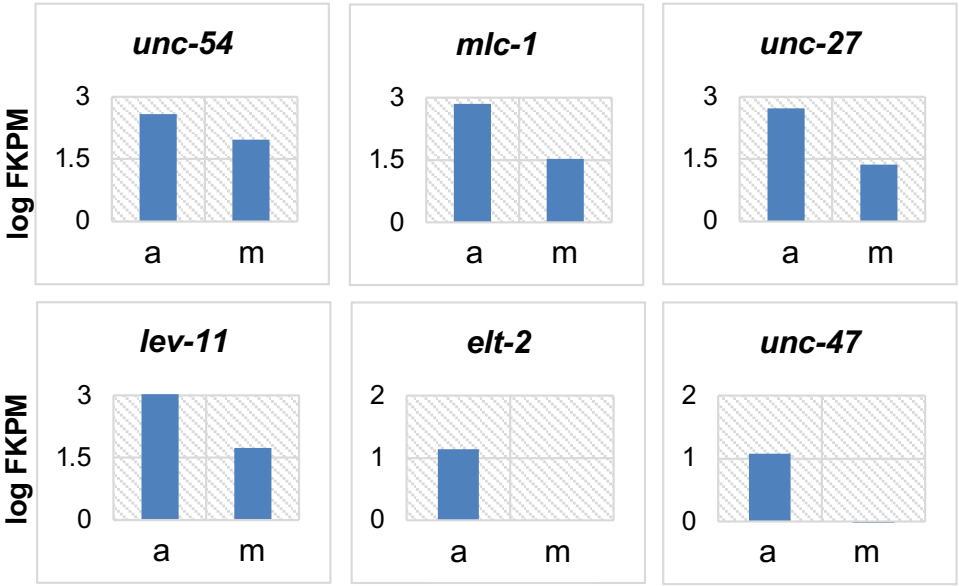

N2
myo-3p::mCherry::his-58
B


myo3::mCherry::his-58
mCherry+ Events
5.3%
94.7%
mCherry+
Negative

Main Figure 3

A



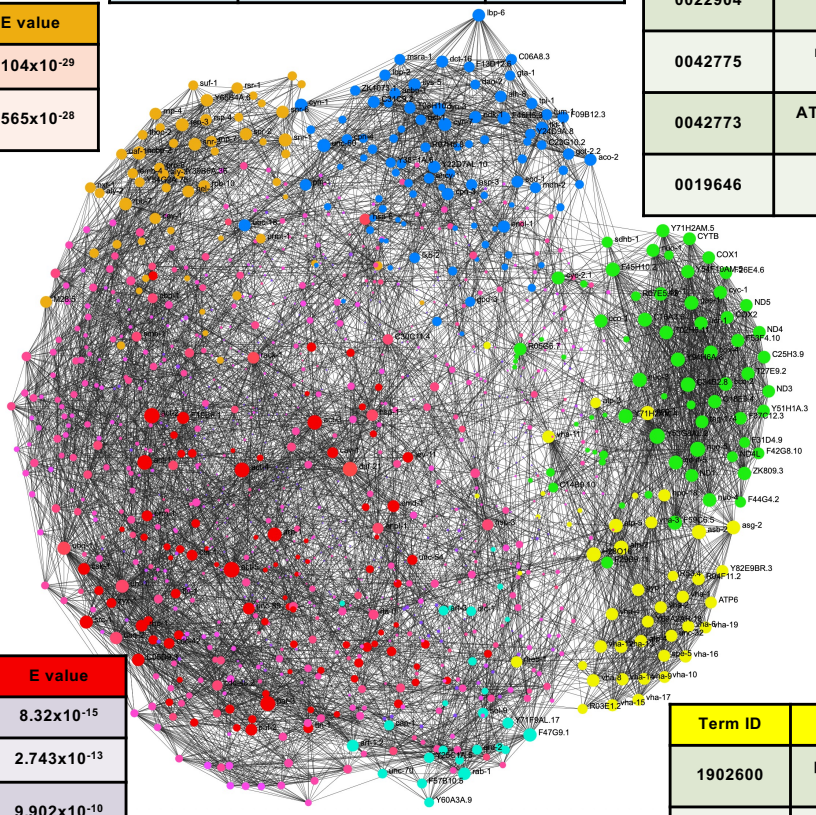


B

Term ID	Term Name	E value
0003824	catalytic activity	3.794×10^{-16}
0016071	small molecule metabolic process	6.546×10^{-14}
0006091	generation of precursor metabolites and energy	6.482×10^{-13}

Term ID	Term Name	E value
0006379	mRNA processing	5.104×10^{-29}
0016071	mRNA metabolic process	2.565×10^{-28}

Term ID	Term Name	E value
0022904	respiratory electron transport chain	6.242×10^{-40}
0042775	mitochondrial ATP synthesis	2.72×10^{-39}
0042773	ATP synthesis coupled electron transport	4.807×10^{-39}
0019646	aerobic electron transport chain	4.807×10^{-39}



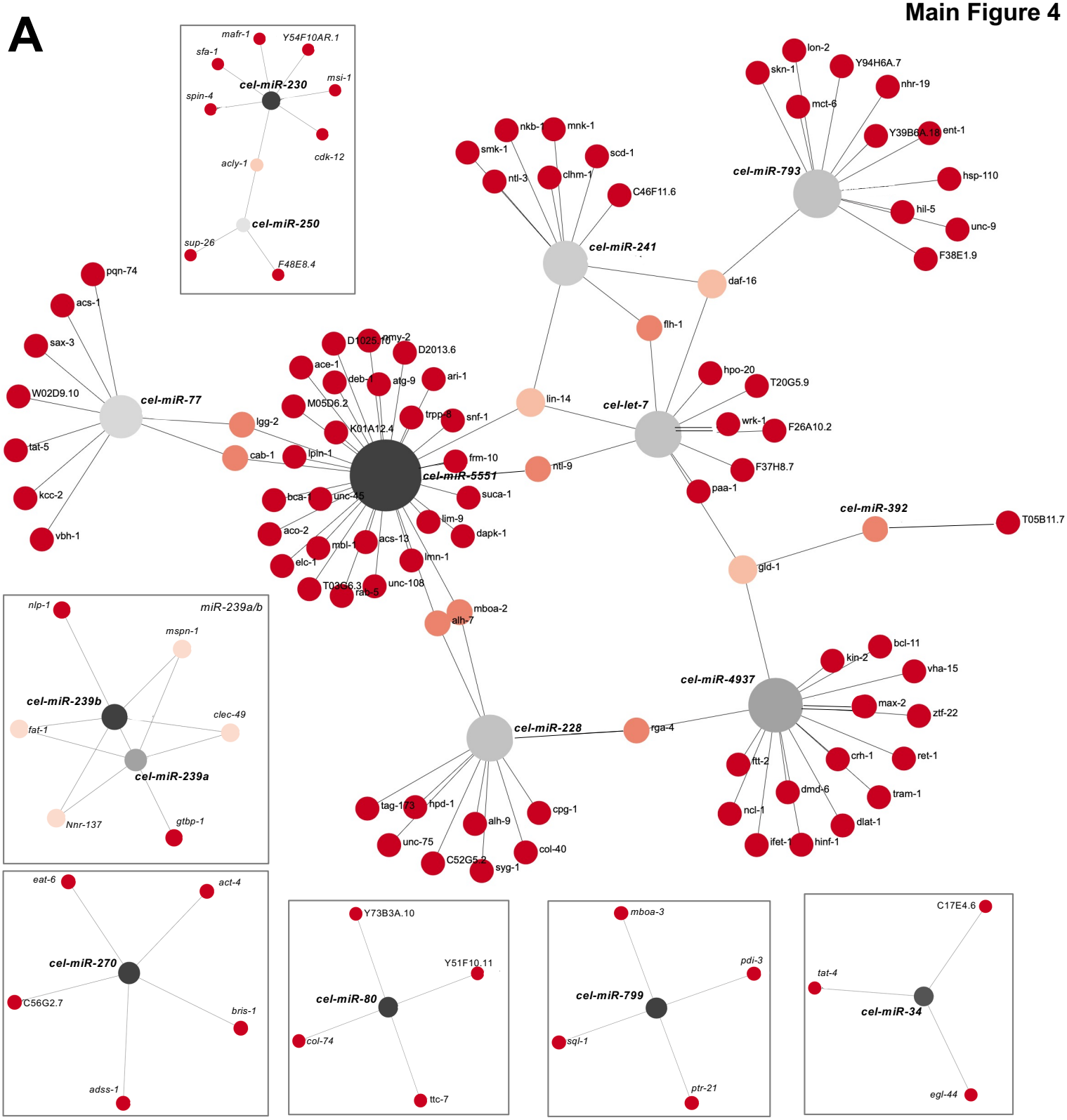
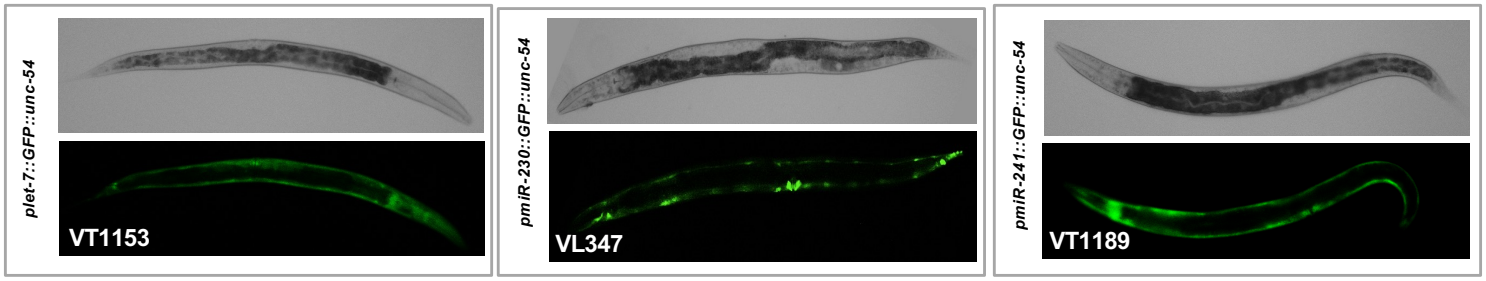
Term ID	Term Name	E value
0051015	actin filament binding	8.32×10^{-15}
003779	actin binding	2.743×10^{-13}
0009653	anatomical structure morphogenesis	9.902×10^{-10}
0030239	myofibril assembly	2.173×10^{-6}
0055001	muscle cell development	3.684×10^{-6}
0055002	striated muscle cell development	2.846×10^{-6}

Term ID	Term Name	E value
0005793	endoplasmic reticulum	4.422×10^{-10}
0030135	coated vesicle	1.219×10^{-9}
0005794	Golgi apparatus	4.269×10^{-9}

Term ID	Term Name	E value
1902600	proton transmembrane transport	2.084×10^{-60}
0098662	inorganic cation transmembrane transport	4.313×10^{-39}
0006754	ATP biosynthetic process	3.425×10^{-32}
0046034	ATP metabolic process	6.021×10^{-26}

Main Figure 4

A



B