Development 136, 1241-1249 (2009) doi:10.1242/dev.030668

Caudal-like PAL-1 directly activates the bodywall muscle module regulator *hlh-1* in *C. elegans* to initiate the embryonic muscle gene regulatory network

Haiyan Lei[1], Jun Liu[2], Tetsunari Fukushige[1], Andrew Fire[3] and Michael Krause[1,*]

Previous work in *C. elegans* has shown that posterior embryonic bodywall muscle lineages are regulated through a genetically defined transcriptional cascade that includes PAL-1/Caudal-mediated activation of muscle-specific transcription factors, including HLH-1/MRF and UNC-120/SRF, which together orchestrate specification and differentiation. Using chromatin immunoprecipitation (ChIP) in embryos, we now demonstrate direct binding of PAL-1 in vivo to an *hlh-1* enhancer element. Through mutational analysis of the evolutionarily conserved sequences within this enhancer, we identify two cis-acting elements and their associated trans-acting factors (PAL-1 and HLH-1) that are crucial for the temporal-spatial expression of *hlh-1* and proper myogenesis. Our data demonstrate that *hlh-1* is indeed a direct target of PAL-1 in the posterior embryonic *C. elegans* muscle lineages, defining a novel in vivo binding site for this crucial developmental regulator. We find that the same enhancer element is also a target of HLH-1 positive auto regulation, underlying (at least in part) the sustained high levels of CeMyoD in bodywall muscle throughout development. Together, these results provide a molecular framework for the gene regulatory network activating the muscle module during embryogenesis.

KEY WORDS: *C. elegans*, Myogenesis, Caudal, MyoD, Embryogenesis

INTRODUCTION

Embryonic skeletal muscle development in vertebrates is dependent on a well-defined transcriptional cascade that is evolutionarily conserved. Cascade components include homeobox-containing factors (PAX-3/7), which activate a set of four related basic helix-loop-helix (bHLH) factors (MyoD, MYF-5, MRF-4 and myogenin; a.k.a. MRFs) that work in concert with MADS box factors (MEF-2 and SRF) to specify muscle cells and activate expression of muscle-specific genes (Buckingham and Relaix, 2007; Charge and Rudnicki, 2004; Tapscott, 2005). A similar, genetically defined pathway has been shown to regulate posterior bodywall muscle development during *C. elegans* embryogenesis (Baugh and Hunter, 2006; Fukushige et al., 2006). The maternally supplied, Caudal-related homeobox transcription factor PAL-1 specifies the posterior somatic blastomeres called C and D (Hunter and Kenyon, 1996; Edgar et al., 2001). PAL-1 acts as a binary developmental switch that is toggled by nuclear levels of POP-1/TCF that, in turn, are regulated by Wnt/MAP kinase signaling. In the presence of high nuclear POP-1, PAL-1 drives hypodermal (skin) cell fate, whereas in the presence of low or absent POP-1, PAL-1 directs bodywall muscle development (Baugh et al., 2005a; Fukushige and Krause, 2005). In the case of myogenesis, PAL-1 is required genetically to activate key transcriptional regulators of myogenesis, including HLH-1/MRF and UNC-120/SRF (Baugh et al., 2005a; Baugh et al., 2005b; Fukushige et al., 2006; Yanai et al., 2008).

Circumstantial evidence suggests that both HLH-1 and UNC-120 are direct targets of PAL-1 in the posterior myogenic lineage. Both HLH-1 and UNC-120 are detectable within posterior, PAL-1-positive muscle lineages within 60 minutes of somatic lineage establishment (Fukushige et al., 2006). Similarly, expression array analysis reveals that ectopically expressed *pal-1* in early embryos results in the activation of both *hlh-1* and *unc-120* within 2 hours (Fukushige and Krause, 2005), and that both genes are temporally downstream of PAL-1 activity in wild-type development (Baugh et al., 2005a; Baugh et al., 2005b; Yanai et al., 2008). Thus, PAL-1 sits at the top of a hierarchy of gene function in posterior embryonic bodywall muscle development, although the molecular details of this transcriptional cascade remain unknown.

To gain insight into the network of maternal and zygotic genes that regulate muscle module transcription factors during *C. elegans* development, we combined embryonic chromatin immunoprecipitation (ChIP) assays with transgene mutagenesis. Our results demonstrate that the maternal factor PAL-1 directly binds in vivo to an enhancer region within the promoter of the potent myogenic regulator *hlh-1*. Mutational analysis of conserved sequences within this enhancer identified two sequence elements, P1 and E1, which bind to PAL-1 and HLH-1, respectively. *hlh-1* is the first direct, zygotic target of PAL-1 to be molecularly identified in vivo, providing information on PAL-1 binding site preferences. Our study also provides details of the temporal-spatial control of *hlh-1*, including the first direct evidence for HLH-1 positive auto-regulation. Together, these results molecularly define a network of maternal and zygotic genes that is responsible for activating and sustaining muscle development; they also suggest how extrinsic cell signals can influence the activity of key components within this network.

MATERIALS AND METHODS

C. elegans strains

The following *C. elegans* strains were used: wild type (N2); heat shock promoter driven *pal-1* (*hs::pal-1*; JA1180) kindly provided by J. Ahringer (University of Cambridge, UK); *hs::hlh-1* (KM267) (Fukushige and Krause, 2005); PD6817: *hlh-1(cc450)/mIn1[dpy-10(e128) mIs14]* II; C-His terminal

[1]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA. [3]Departments of Pathology and Genetics, Stanford School of Medicine, Stanford, CA 94305, USA.

*Author for correspondence (e-mail: mwkrause@helix.nih.gov)

Accepted 9 February 2009

tagged *hs::pal-1* (KM471); 353 bp *hlh-1* enh-1 driving *myo-2* basal promoter (L3136; pPD107.97) (KM480, KM481) and the mutant derivatives P1 site mutant KM483, KM484, E1 site mutant KM485, KM486, P1 and E1 double mutant KM487, KM488; 351 bp *hlh-1* negative control region driving *myo-2* basal promoter (KM489, KM490); four copy JKL26 with the *pes-10* basal promoter (PD4743); and eight copy P1 element driving *myo-2* basal promoter (KM492, KM493).

Generating transgenic lines

Promoter deletion and enhancer assays primarily used the *pes-10* basal promoter vector L3135 (pPD107.94), which drives the expression of coding regions for green fluorescent protein (GFP) fused to *lacZ*. Transgenic strains were generated using 10-100 ng/μl of test plasmid and 50 μg/μl of the selectable dominant *rol-6* plasmid pRF4. Fine resolution mapping of enhancer sequences used concatenated oligonucleotides (between one and eight copies) as simple or complex arrays (see Tables S1 and S2 in the supplementary material) upstream of the *pes-10* basal promoter. Complex array methods used 4-8 μg/ml test plasmid, 50 μg/ml of pRF4 (all test plasmid DNAs digested with *Fsp*I and pRF4 digested with *Sca*I) and 100 μg/ml of *Pvu*II digested N2 genomic DNA.

A full-length *pal-1* cDNA was joined to a 6×His C- or N-terminal tag (primer details can be provided on request) inserted into the *hsp16.41* vector pPD49.83 *Kpn*I/*Nco*I sites. Extrachromosomal transgenic lines were subsequently integrated by gamma irradiation yielding strains KM471, KM472, KM473 and KM474.

Chromatin immunoprecipitation (ChIP)

ChIP assays were performed using the Upstate ChIP kit (cat #17-295). C-terminal 6×His-tagged, *hs::pal-1* transgenic embryos were collected from 10 synchronized 10 cm OP50 feeding plates. After a 34°C heat-shock for 30 minutes, embryos were incubated at room temperature for 3 hours and collected, frozen on dry ice and cracked by three freeze/thaw cycles. Embryos were incubated in a crosslinking solution of 1.5% formaldehyde/PBS/proteinase inhibitor buffer (Sigma, cat #P8340) for 20 minutes at ~22°C, pelleted, washed three times in cold PBS/proteinase inhibitor buffer and used immediately or frozen at –80°C. After washing embryos, warm nucleic SDS lysis buffer (Upstate, cat #20-163) was added, the samples vortexed for 20 seconds and put on ice for 30 minutes to 1 hour with several vortex treatments during incubation. Samples were sonicated on ice (Misonix sonicator, Model #3000) using an output program of 2.5 (6 W), 30 second work/30 second stop/4 minutes to shear DNA to between 200 and 1000 bp, as determined by gel electrophoresis. Samples were centrifuged to remove debris. The supernatant was collected and divided into three fractions: input, sample and IgG control. Each fraction (except input) was diluted 10-fold into ChIP dilution buffer (Upstate, catalog number 20-153) and added to pre-cleaned ssDNA/protein A agarose beads (Upstate, catalog number 16-157c). Sample fractions (1500 μl) were incubated with 5 μl 6×His antibody (Abcam, ab9108) overnight at 4°C with constant rotation while IgG control samples were incubated in parallel with 5 μl normal Rabbit IgG (Upstate, catalog number 12-370). After incubation, 60 μl protein A beads were added to each sample for 1 hour with constant rotation. Beads were washed three times each with a low salt (Upstate, cat #20-154), followed by high salt (Upstate, catalog number 20-155), followed by $LiCl_2$ (Upstate, catalog number 20-156) and finally TE buffer (Upstate, catalog number 20-157). After the last wash, beads were resuspended in fresh elution buffer (1% SDS, 0.1 M $NaHCO_3$) for 30 minutes. Beads were pelleted with a brief spin and the supernatant transferred to a new tube; this was repeated once and elution volumes combined. The eluted volumes and the input (500 μl each) were adjusted with 20 μl 5 M NaCl and heated at 65°C for 4 hours followed by the addition of EDTA, Tris-HCl, proteinase K and incubation for 1 hour at 45°C. The DNA was recovered from samples using the Qiaquick PCR Purification Kit (Qiagen) with elution off the column with 100 μl elution buffer. Eluted DNA (0.4 μl) was used as a template in each 20 μl quantitative PCR (qPCR) reaction.

Heat-shock experiments

Two-cell stage embryos were isolated by hand dissection and incubated at room temperature for 25 minutes before heat-shock. Heat shock consisted of 30 minutes at 34°C with reporter gene expression assayed 4.5 hours later.

RESULTS

Fine mapping of *hlh-1* enhancers

The promoter of the *hlh-1* gene, encoding the lone MRF-related protein in *C. elegans*, has previously defined muscle regulatory regions within 3 kb upstream of the initiation codon, as well as one conserved, but dispensable, enhancer within intron 1 (Krause et al., 1994) (Fig. 1A). To define the enhancer regions at higher resolution, we tested a series of additional genomic fragments to narrow each enhancer to several hundreds of base pairs. Each enhancer was further dissected using a series of partially overlapping (15 bp) 54-mer oligonucleotides that were concatenated and cloned into the *pes-10* basal promoter test plasmid to generate stable transgenic strains. Evolutionarily conserved sequences within muscle positive concatenates were eliminated or mutated and these oligonucleotide derivatives again tested for activity. Our analysis significantly improved the resolution of the map of cis-acting sequence elements that function to regulate *hlh-1* expression. Three muscle enhancer elements can now be defined within 3 kb upstream of the *hlh-1* start that are centered around the following positions relative to the translational start codon: enh-1 (–2446), enh-2 (–1536) and enh-3 (–496) (Fig. 1A). Further details of this analysis are available in Figs S1 and S2, and Tables S1 and S2 in the supplementary material.

Most *hlh-1*-derived enhancers with muscle activity functioned in larval and adult bodywall muscle cells, but showed distinct preferences for certain bodywall muscle lineages in embryos. For example, enh-1 was preferentially active in the posterior C and D lineages, whereas enh-2 was active in C, D and MS muscle lineages (Fig. 1B; see Figs S1 and S2 in the supplementary material). This suggested that although there may be common sequence elements directing general bodywall muscle expression after hatching, lineage-restricted elements probably functioned during embryonic development to direct the correct spatial and temporal regulation of *hlh-1* activation and maintain its expression.

PAL-1 directly binds to the *hlh-1* enh-1 region

One of the important and unresolved issues for *hlh-1* gene regulation was the identity of its activator(s) during embryogenesis. In this regard, enh-1 drew our attention because of its strong preference for directing expression in the posterior C and D bodywall muscle lineages in early embryogenesis (Krause et al., 1990). Previous studies have shown that C and D blastomere specification is dependent on the caudal-related transcription factor PAL-1 (Hunter and Kenyon, 1996). Maternal PAL-1 protein can be first detected in the P2 and EMS blastomeres of four-cell stage embryos and subsequently in the somatic descendents of P2 – namely C and D. Maternal PAL-1 persists in the posterior somatic lineages until the 100-cell stage of embryogenesis, overlapping the onset of zygotic PAL-1 accumulation that persists in these lineages until the end of gastrulation (~350 cells) (Hunter and Kenyon, 1996; Edgar et al., 2001). This temporal and spatial profile of PAL-1 initially proceeds, and is then coincident with, the expression of *hlh-1* in the C and D lineages, putting PAL-1 in the right place and time to activate this crucial myogenic regulatory gene directly. Unfortunately, the available PAL-1 antibody failed to pull down endogenous PAL-1 protein efficiently by immunoprecipitation (C. Hunter, personal communication) and was unsuitable for chromatin immunoprecipitation (ChIP) studies.

We engineered N- and C-terminal 6×His-tagged versions of PAL-1 driven by a heat shock promoter and made stable integrated transgenic strains to assay by ChIP if PAL-1 was directly binding to *hlh-1* enhancer sequences in vivo. To ensure that these 6×His-tagged versions of PAL-1 retained wild type

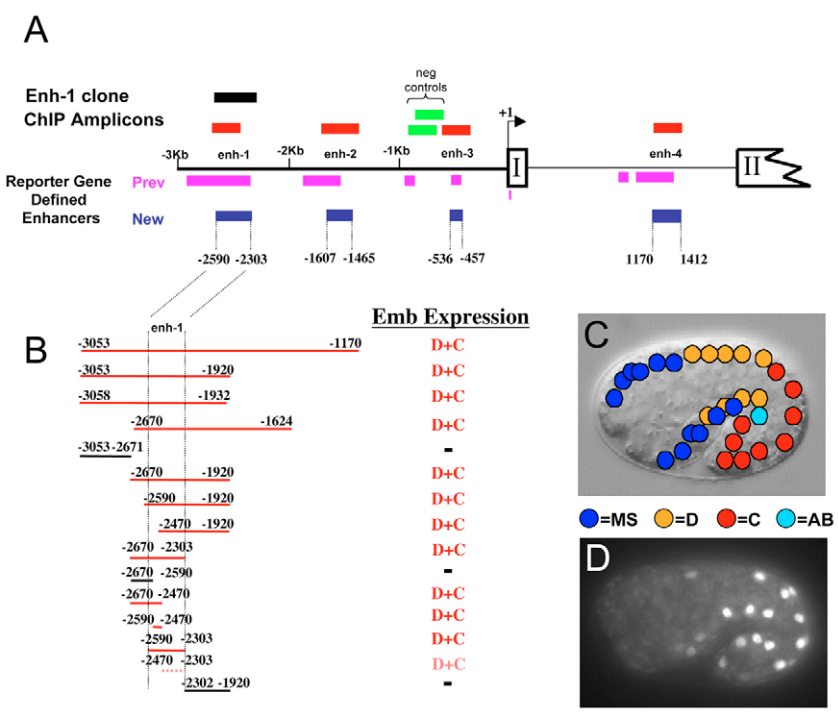


Fig. 1. *hlh-1* enhancer regions. (**A**) The promoter and partial coding region of *hlh-1* (up to exon II) is illustrated with regions of interest highlighted. Shown below the gene are the bodywall muscle enhancer regions defined by this (enh-1 to enh-4; blue boxes) and previous work (purple boxes), with positions indicated relative to nucleotide 1 of the start codon. Shown above the gene are the regions interrogated by qPCR after ChIP, corresponding to newly defined muscle enhancers (red) or negative control regions (green). The black box above enh-1 indicates the cloned genomic region driving reporter gene expression assayed in vivo. (**B**) Reporter gene constructs used to narrow down the location of enh-1. The genomic regions, as indicated by nucleotide positions listed at the ends of each line, were cloned and tested for bodywall muscle enhancer activity as described. Fragments that were positive for muscle lineage expression are underlined in red; a broken underline indicates weak or variable expression. Black underline indicates fragments negative for muscle expression. D+C: D and C founder blastomere embryonic muscle lineage expression; a dash indicates no expression detected. (**C**) Schematic representation of bodywall muscle nuclear positions and lineage of origin indicated by color. (**D**) Example of GFP observed during embryogenesis for an enh-1-driven reporter gene showing strongest expression in D and C lineage bodywall muscles.

activity, we tested each in a biological assay. We employed an in vivo blastomere cell fate conversion assay rather than rescuing the *pal-1* mutant, as the rescue experiments would be complicated by a need for the transgene to be expressed both maternally and zygotically. We have shown previously that ectopic expression of either endogenous or transgenic *pal-1* in early embryos is sufficient to convert almost all blastomeres to a body wall muscle-like fate when POP-1 levels are depleted by RNAi (Fukushige and Krause, 2005). We found that both the N- and C-terminal 6×His-tagged PAL-1 proteins were indistinguishable from wild-type PAL-1 in this assay (see Fig. S3 in the supplementary material), demonstrating that both fusion proteins retained biological activity in vivo. We chose to focus on the C-terminal tagged transgene for all subsequent experiments. We monitored by western blot analysis the level of C-terminal 6×His-tagged PAL-1 protein (His-PAL-1) accumulation following heat shock induction and found that under optimal conditions, heat-shock induced His-PAL-1 accumulation was robust with maximum levels of protein detected 3 hours after induction (see Fig. S4 in the supplementary material).

To determine whether PAL-1 could bind directly to the *hlh-1* promoter, we did ChIP assays in transgenic embryos harboring the heat shock inducible His-PAL-1 transgene. Mixed early stage, transgenic embryos were isolated by bleach treatment and heat shocked at 34°C for 30 minutes and chromatin prepared 3 hours later. Regions corresponding to the defined *hlh-1* muscle enhancers, or negative control regions, were interrogated by ChIP followed by quantitative PCR (qPCR). An enhancer region was considered positive in the ChIP assay if in each of at least three repetitions of the experiment its signal intensity by qPCR following anti-His antibody ChIP was at least twofold higher than the control anti-IgG ChIP. We found that of all of the *hlh-1* promoter regions tested, only enh-1 was identified as clearly being bound by PAL-1, as assayed by ChIP with a signal almost eight times higher than background (Fig. 2A).

To gain additional insight into *hlh-1* regulation and to control for ChIP assay specificity, we also looked for occupancy at any of these *hlh-1* genomic regions by HLH-1 itself, given that previous studies have suggested that HLH-1 positively regulates its own expression (Krause et al., 1994). Our attempts to ChIP endogenous levels of HLH-1 from wild-type embryos or strains carrying a high copy number *hlh-1* promoter transgene failed to generate a signal above background. Consequently, we generated integrated transgenic lines expressing a full-length *hlh-1* cDNA under the control of the heat-shock promoter, induced expression, prepared chromatin as described above and ChIPed with an affinity-purified chicken anti-HLH-1 antibody. Western blot analysis demonstrated that heat-shock-induced HLH-1 protein levels from the transgene accumulated to between 20- and 100-fold higher levels than endogenous HLH-1 (data not shown). ChIP data demonstrated in vivo HLH-1 occupancy of enh-1, enh-3 and enh-4 regions of the *hlh-1* gene, suggesting one or more of these sites may mediate its positive auto regulation (Fig. 2B). The clear distinction between positive enhancer elements from PAL-1 and HLH-1 ChIP experiments, coupled with the lack of binding to our negative control regions, suggested the antibodies used for ChIP had specificity. As both experimental approaches relied on overexpression of the transcription factor prior to ChIP, probably resulting in unnatural binding conditions, the positive results were validated by extensive analysis of functional cis-acting elements.

Both PAL-1 and HLH-1 alone can activate transcription through enh-1

We focused on the role of enh-1 in *hlh-1* gene regulation because this region demonstrated binding by both PAL-1 and HLH-1 in our ChIP assays. We tested whether this region alone was sufficient to drive reporter expression in response to either endogenous factors or ectopic PAL-1 and HLH-1 by cloning a 353 bp genomic fragment upstream of the *myo-2* basal promoter driving *gfp* expression (Okkema et al., 1993). Multiple independent integrated

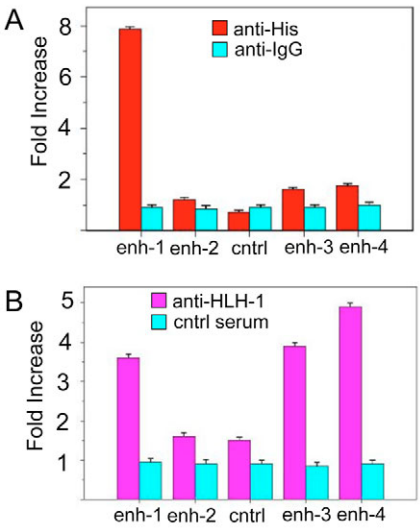


Fig. 2. Chromatin immunoprecipitation (ChIP) results identify potential in vivo binding sites regulating *hlh-1* expression. (**A**) The four *hlh-1* enhancer regions, and a negative control region (cntrl), were interrogated by qPCR after a His-PAL-1 ChIP from mixed stage embryos. Of the regions assayed, only *hlh-1* enh-1 was bound by PAL-1 (red) at levels greater than the IgG control (blue) ChIP. (**B**) ChIP results of *hlh-1* enhancer and control regions following overexpression of HLH-1 in mixed stage embryos and immunoprecipitation with an affinity-purified chicken-anti-HLH-1 antibody. Enhancer regions 1, 3 and 4 show greater than twofold enrichment (purple) compared with pre-immune serum control (blue) ChIP. All results represent the combined data from a minimum of three independent repetitions of chromatin preparation and ChIP.

enh-1*::myo-2::gfp::lacZ* transgenes resulted in GFP in early embryonic C- and D-derived muscle in otherwise wild-type animals (Fig. 3). Two independent enh-1*::myo-2::gfp::lacZ* lines were crossed into animals carrying integrated version of the heat-shock promoter driving either PAL-1-or HLH-1-encoding cDNAs. In the absence of heat shock induction of either of these two factors, the enh-1*::myo-2::gfp::lacZ* reporter genes retained the early embryonic muscle expression pattern that was seen when assayed on their own. By contrast, heat shock induction of either PAL-1 (Fig. 3A) or HLH-1 (Fig. 3B) in the presence of the reporter resulted in widespread and strong GFP in 92% (*n*=176) and 94% (*n*=165) of the embryos, respectively. Heat-shock treatment of these enh-1 reporters in an otherwise wild-type background resulted in 53% (*n*=77) GFP-positive embryos, with a slight decrease in number of positive cells per embryo owing to heat shock disruption of normal development (Fig. 3A,B). As a control for enh-1 specificity to respond to these induced factors, we also tested a negative control region within the *hlh-1* promoter (Fig. 1A) in this assay; this region showed no response to either PAL-1 or HLH-1 (data not shown).

Enh-1 contains functional PAL-1 and HLH-1 binding sites that are evolutionarily conserved

To identify potential cis-acting sequence responsible for enh-1 responsiveness to either PAL-1 or HLH-1, we (1) tested a series of oligonucleotides spanning the region to further refine sequence elements directing expression, (2) performed a phylogenetic comparison of genomic regions from *C. briggsae* and *C. remanei* encoding the HLH-1 homolog to identify conserved sequences, and (3) tested the activities of conserved sequence motifs by deletion or mutation in muscle enhancer assays in vivo.

A series of eight partially overlapping (15 bp) 54-mer oligonucleotides that spanned the enh-1 region were tested as concatenates for their ability to drive embryonic expression in the C and D bodywall muscle lineages (Fig. 4A; see Tables S1 and S2 in the supplementary material). Three of these (JKL26, JKL28 and JKL30) showed strong expression in embryonic C and D lineages. As a test for HLH-1 dependence of these oligonucleotide concatenates, we put those that showed weak (JKL22 and JKL24) or strong (JKL26, JKL28 and JKL30) activity into a balanced *hlh-1(cc450)/mIn1(mIs14)* mutant background and assayed expression in heterozygous and homozygous *hlh-1*-null mutants. JKL22 failed to express in any animals, suggesting the original expression in D+C lineages was a false positive result. JKL24 and JKL26 were active in both heterozygous and homozygous *hlh-1* mutants, demonstrating that neither region was dependent on wild-type HLH-1 activity to enhance muscle expression. By contrast, JKL30 enhanced muscle expression in heterozygous *hlh-1* mutant animals, but was not expressed in *hlh-1(cc450)* homozygous mutants, demonstrating that this construct was dependent on HLH-1. We were unable to generate lines by injecting JKL28 reporters into the *hlh-1* balanced mutant background, so we tested this construct by crossing two independent reporters into the mutant background. Like JKL30, JKL28 required HLH-1 activity for expression. We also tested the ability of JKL26 (strain PD4743) to respond to PAL-1 by knocking down maternal and early embryonic *pal-1* expression by RNAi injection into parental hermaphrodites. Whereas 64% (*n*=76) of embryos from untreated hermaphrodites showed D+C expression of the reporter gene, only 9% (*n*=96) of embryos from *pal-1* RNAi-treated hermaphrodites had GFP-positive cells (see Fig. S5 in the supplementary material), demonstrating that PAL-1 is required for efficient JKL26 bodywall muscle expression in D+C lineages.

A phylogenetic comparison of genomic regions represented by reporter constructs JKL26 through JKL30 revealed four blocks of conserved sequence in the alignment of sequences from three species (Fig. 4C). Two of the conserved blocks (P1 and Block 3) contained sequences that are identical to known Caudal-binding sites [TTTATG (Dearolf et al., 1989; Maurer et al., 2007)], the *Drosophila* homolog of PAL-1 (TESS: http://www.cbil.upenn.edu/cgi-bin/tess/tess). Block 2 had a single TCF/LEF-like binding site [MAMAG (Travis et al., 1991; Waterman et al., 1991)], the vertebrate homolog of POP-1 that functions in concert with PAL-1 in the C and D lineages. This sequence is also similar to a consensus binding site sequence derived from a limited number of known *C. elegans* POP-1 sites (Arata et al., 2006; Korswagen et al., 2000; Lam et al., 2006; Maduro et al., 2005; Shetty et al., 2005). Finally, the E1 site (CAACTG) is a bHLH factor binding E-box sequence (Blackwell and Weintraub, 1990; Kophengnavong et al., 2000) and we have previously shown that HLH-1 can bind canonical E-box sequences in vitro (Krause et al., 1997).

To test the functional significance of these conserved sequence elements, a series of 13 oligonucleotides that removed or mutated bases that were within conserved blocks were tested for muscle enhancer activity as concatenates cloned upstream of the *pes-10* basal promoter (see Tables S1 and S2 in the supplementary material). The results demonstrated that both the P1 and E1 sites were crucial for bodywall muscle enhancer activity in the embryonic C and D lineages. Some mutations within conserved Block 2 and 3 sequences retained muscle enhancer activity, including weak activity for an oligonucleotide lacking the Caudal-like binding site within

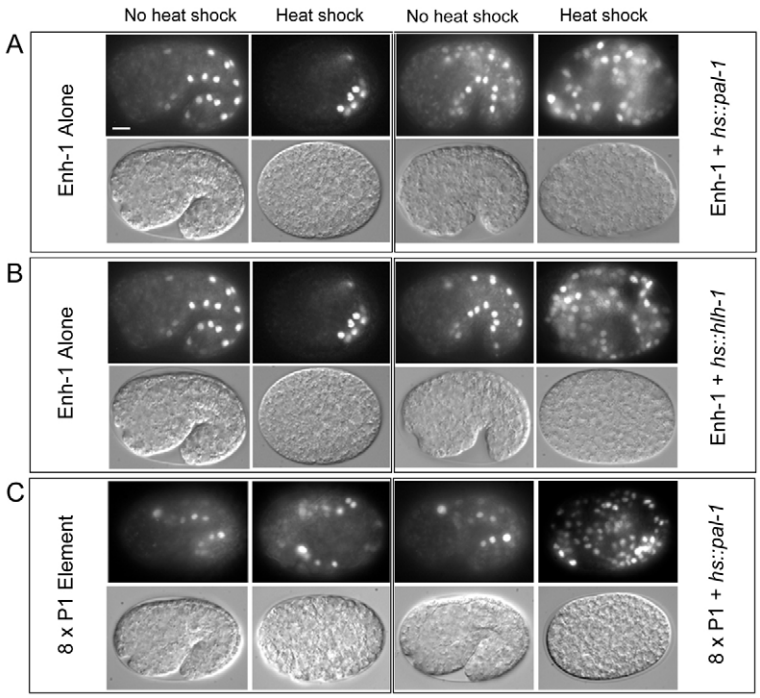


Fig. 3. *hlh-1* enh-1 responds to both PAL-1 and HLH-1 in vivo. (**A-C**) Integrated strains harboring a reporter gene (listed on the left) with either enh-1 (A,B) or eight copies of the enh-1 P1 element (C) were cloned upstream of the *myo-2* basal promoter driving *gfp* expression. On their own, these reporters resulted in GFP predominantly in embryonic D+C bodywall muscle lineages (leftmost panels) with a decrease in the number of GFP-positive cells upon heat shock owing to disrupted embryogenesis. These reporter strains were genetically crossed into strains that overproduced either PAL-1 (A,C) or HLH-1 (B) in response to heat-shock treatment. Overproduction of either PAL-1 or HLH-1 resulted in disrupted embryogenesis, widespread conversion of blastomeres to a bodywall muscle-like fate and widespread activation of reporters. All embryos shown are between 5 and 6 hours of development, and are orientated (when possible) with anterior towards the left and dorsal towards the top. Each embryo shown has a GFP (top) and a corresponding Nomarski image (bottom). Enh-1 alone images from A are duplicated in B. Scale bar: 10 μm.

Block 3. We concluded that site P1 was responsible for much of the muscle enhancer activity of JKL26, possibly serving as a direct PAL-1-binding site. E1 appeared to be an HLH-1 binding site that was responsible for JKL28 and JKL30 muscle enhancer activity, consistent with our *hlh-1(cc450)* results above.

PAL-1, but not POP-1, binds enh-1 sequences in vitro

Although there are no known in vivo PAL-1 binding sites, the conserved P1 site seemed a likely candidate. We used electrophoretic mobility shift assays (EMSA) with bacterially expressed recombinant PAL-1 protein and a 20-mer probe containing the conserved P1 sequence (TATTTATG). The P1 site probe was bound by PAL-1, whereas a probe containing a mutated version of the P1 site was not (Fig. 5A). PAL-1 also shifted in vitro the Block 3 site, but not derivatives in which the potential PAL-1-binding site was mutated (data not shown; see Table S1 in the supplementary material).

Several POP-1-binding sites have been previously identified in *C. elegans* (Arata et al., 2006; Korswagen et al., 2000; Lam et al., 2006; Maduro et al., 2005; Shetty et al., 2005). The conserved sequences within Block 2 of enh-1 match the eight-base POP-1 consensus (C/T)TTTG(A/T)(A/T)(A/G/C) with the exception of position 7 [Block 2 has a C, rather than (A/T) in this position]. Using a previously studied, truncated POP-1 bacterial expression clone kindly provided by D. Eisenmann, we assayed Block 2 by EMSA. Although we could easily demonstrate binding of POP-1 to control oligonucleotides that matched the consensus perfectly (CTTTGATC), POP-1 failed to bind a probe centered on the Block 2 site (TTTTGACG) (data not shown; see Table S1 in the supplementary material). Changing the seventh position C to match the consensus resulted in strong binding by POP-1 in vitro. Thus, we were unable to provide any positive evidence for Block 2 serving as a binding site for POP-1, suggesting one or more additional factors are needed to bind this sequence, either on their own or in combination with POP-1.

The putative PAL-1 and HLH-1 binding sites of enh-1 function in vivo

To test the ability of the P1 and E1 elements of enh-1 to respond to PAL-1 and HLH-1 in vivo, we made a series of site-directed mutations in P1, E1 or both sites within the *enh1::myo-2::gfp::lacZ* transgene and tested the ability of chromosomal integrants of each mutant version to respond to heat shock-induced PAL-1 or HLH-1. Mutation of the P1 site almost completely abolished the response of the reporter to PAL-1, but did not affect its response to HLH-1 (Fig. 6). Conversely, mutation of E1 eliminated the response of the reporter gene to HLH-1, but not its ability to respond to PAL-1. The E1 site mutation did diminish PAL-1 responsiveness of enh-1 (92.4% to 74.4%), presumably because this alteration eliminates the positive auto-regulatory HLH-1 feedback loop on this transgene. Finally, mutation of both the P1 and E1 sites of enh-1 severely reduced the responsiveness of the reporter to either PAL-1 or HLH-1. As a control, we tested each of these mutant reporters alone after heat shock and found very little expression in the absence of the PAL-1 or HLH-1 encoding transgenes; 4% (n=75) for the P1 mutant, 4% (n=78) for E1 mutant, and 3% (n=76) for the P1, E1 double mutant. These results demonstrate that the P1 and E1 sites are both necessary and sufficient for robust enh-1 responsiveness in vivo to PAL-1 and HLH-1, respectively.

PAL-1 binding site specificity

To determine whether additional sequence information was responsible for PAL-1 specificity for the P1 site, we focused on the potential contribution of flanking sequences that extended from the core. Each of the three transcription factors (HLH-1, UNC-120 and HND-1) that define the regulatory muscle module (Baugh et al., 2005a; Baugh et al., 2005b; Fukushige et al., 2006) have one or more perfect matches to the P1 core sequence (TTTATG). Two such sequences exist in the enh-1 region of *hlh-1* (P1 and Block 3), there is one matching sequence in the enh-2 region of *hlh-1*, one matching sequence at –2405 from the ATG of *unc-120*, and one matching sequence at –3589 from the ATG of *hnd-1*. We interrogated the

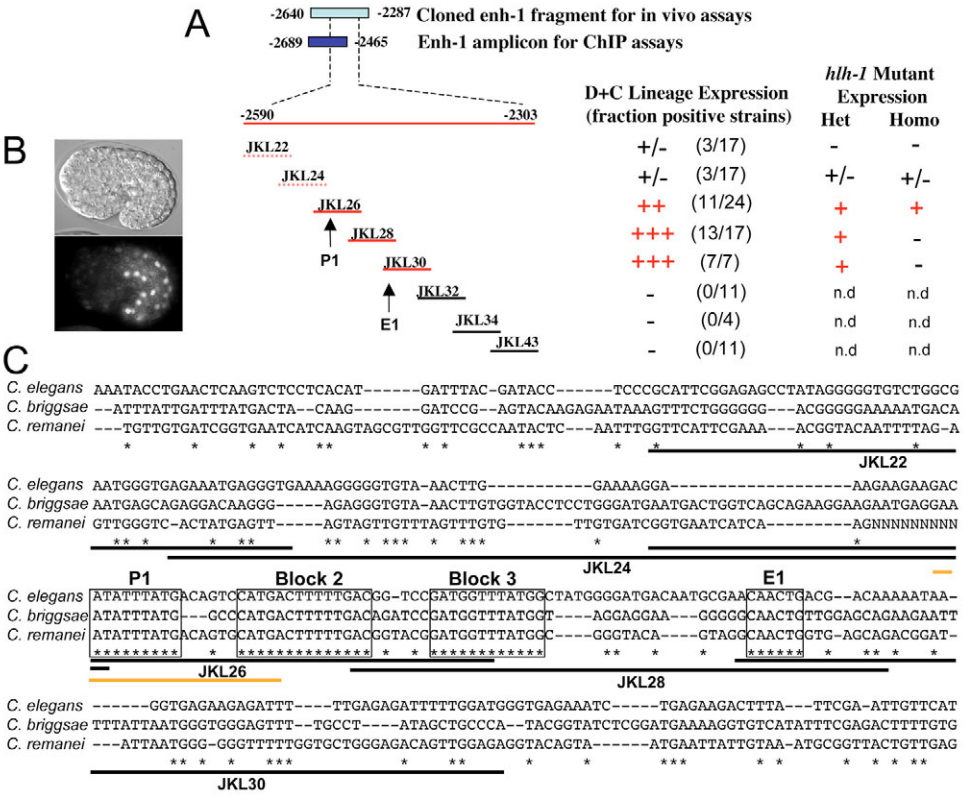


Fig. 4. Identification of *hlh-1* enh-1 subelements involved in embryonic bodywall muscle expression. (**A**) A series of partially overlapping oligonucleotides from the *hlh-1* enh-1 region were tested in transgenic strains for their ability to drive embryonic expression in the posterior PAL-1-dependent D+C lineages. Positive fragments are underlined in red, with broken underlines indicating weak or variable expression. Black underlines indicate fragments negative for expression. Scoring: no expression is indicated by a dash (–); body wall muscle (bwm) expression in <10% of strains (+/–), 10-25% of strains (+), 26%-50% of strains (++), >51% of strains (+++); n.d. is not tested. (**B**) Best case representation of expression for D+C-positive reporter genes (JKL28 shown). Normarski image at top with corresponding GFP pattern below. (**C**) An alignment of sequences upstream of the *hlh-1* homologs in *C. remanei* and *C. briggsae* to the *C. elegans* enh-1 region; dashes indicate spaces, Ns within *remanei* sequence reflect a break in contigs used for alignment. Four substantial blocks of conserved sequence are identified, as shown in the boxes. The positions for the putative PAL-1-binding site (P1) and HLH-1-binding site (E1) are indicated relative to enh-1 derivatives shown in A. Black lines below the sequence mark the extent of oligonucleotides used in enhancer assays as detailed in A; brown lines underscore the sequence multimerized for the eight copy P1 element.

ability of heat-shock-induced PAL-1 to bind to each of these sites in vivo by ChIP-qPCR. Only two sites were positive for PAL-1 binding, *hlh-1* enh-1 and the site upstream of *unc-120* (Fig. 5B), suggesting that *hlh-1* and *unc-120*, but not *hnd-1*, are direct targets of PAL-1.

A comparison of the PAL-1 ChIP-positive fragment sequences suggested that the PAL-1-binding site might be defined by the extended sequence ATTTATGAC as this sequence is present in all ChIP-positive genomic fragments tested. The functional significance of this extended sequence is also supported by our site-specific mutational analysis using oligonucleotides in bodywall muscle enhancer assays reported above (see JKL114 and JKL118, Tables S1 and S2 in the supplementary material). To test the validity of this extended sequence, we multimerized a 23-mer oligonucleotide centered on the P1 site (eight copies) and placed it upstream of the *myo-2::gfp::lacZ* basal promoter (see Fig. 4C). Integrated transgenes of this reporter were then crossed into strains containing the heat-shock-inducible PAL-1 transgene and the response assayed with and without heat-shock treatment. In the absence of heat shock, GFP was detected in embryos primarily in hypodermal and bodywall muscle derivatives of the C lineage that are all dependent on PAL-1 function (Fig. 3C). This expression pattern was mosaic within the population, although 53% (n=214) had some GFP-positive bodywall muscle precursors; heat-shock treatment of this reporter alone resulted in a small increase in this number (64%, n=163), although the number of positive cells per embryo remained about the same. Heat-shock induction of PAL-1 in embryos harboring the eight copy P1 element resulted in strong GFP throughout the embryo; 78% (n=105) of embryos were GFP-positive and there was nearly a threefold increase in the number of GFP-positive cells per embryo (Fig. 3C). These results demonstrated that the P1 sequence alone is sufficient to respond robustly to ectopic PAL-1.

The expression of the eight copy P1 reporter gene alone in hypodermis and bodywall muscle cells derived from the C blastomere was different from all other enh-1 fragments tested. We have shown previously that hypodermal and bodywall muscle fates in this lineage are specified by PAL-1 acting in concert with different levels of nuclear POP-1 (Fukushige et al., 2005). Comparing the sequence of the eight copy P1 oligonucleotide to a closely related enh-1 subfragment that was bodywall muscle specific (JKL26) suggested the putative TCF/POP-1 binding site in Block 2 might play a role in bodywall muscle specificity (see Fig. 4C). As noted above, we have been unable to demonstrate in vitro that POP-1 alone

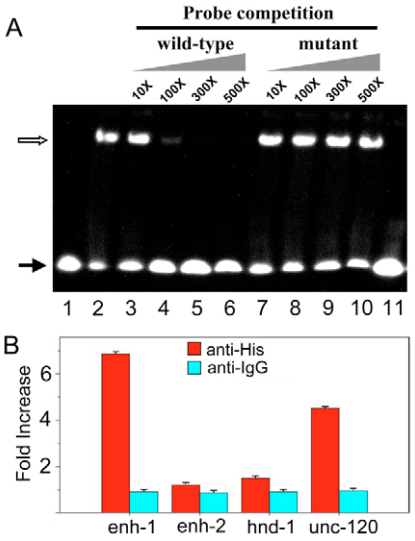


Fig. 5. PAL-1 binding. (**A**) Electromobility gel shifts were used to determine whether bacterially produced PAL-1 protein could specifically bind the P1 site of *hlh-1* enh-1. A biotin-labeled 20-mer oligonucleotide (60 fmol) centered on the wild-type P1 site sequence was annealed to a 100-fold excess of a non-labeled complementary oligonucleotide and run on the gel alone (lane 1) or after incubation with PAL-1 protein (lanes 2-10); free probe is indicated by a black arrow and PAL-1 bound probe is indicated by a white arrow. PAL-1 binding (lane 2) was competed with increasing amounts of either the wild-type probe (lanes 3-6) or a mutated sequence that eliminated the putative PAL-1-binding site (lanes 7-10). The mutant probe failed to compete for binding and did not bind PAL-1 on its own (lane 11). (**B**) Promoter regions upstream of *hlh-1*, *unc-120* and *hnd-1* that contained sequence elements resembling the core PAL-1-binding site (ATTTATG) from *hlh-1* enh-1 were interrogated by qPCR after overexpression of 6×His tagged PAL-1 and ChIP. Only *hlh-1* enh-1 and the region upstream of *unc-120* were reproducibly identified as being bound to PAL-1, as assayed by ChIP.

binds to Block 2 sequences, therefore, we tested the role of POP-1 in mediating the differences in expression seen with JKL26 and eight copy P1 in vivo using RNAi. Heat-shock induction of PAL-1 in embryos harboring the eight copy P1 reporter and depleted of POP-1 by RNAi resulted in a similar magnitude of effect on reporter gene expression as it did in embryos with wild-type levels of POP-1; there was an increase in both the number of GFP positive embryos (6-fold) and number of GFP positive cells per embryo (threefold). Although the disruption of embryogenesis due to POP-1 depletion prevents a direct comparison of numbers, this result suggests the absence of POP-1 has little effect on the activity of the eight copy P1 reporter. Similarly, RNAi of *lit-1*, which blocks POP-1 nuclear export, resulted in an ~2.5-fold increase in the number of GFP-positive cells per embryo when comparing heat shock-induced *pal-1* expression to the non-heat shocked controls. We concluded that the eight copy P1 element reporter responds to PAL-1 independently of the nuclear levels of POP-1, consistent with the expression of this reporter gene in both hypodermal and bodywall muscles derived from the C lineage.

We also tested a JKL26 reporter gene (four copies; strain PD4743) for its response to PAL-1 relative to POP-1 levels. Because our attempts to make genetic doubles between the *hs::pal-1* and JKL26 reporter strains were unsuccessful, we turned to RNAi. We used *mex-3* RNAi as a proxy for the induction of ectopic PAL-1 activity; loss of MEX-3 activity results in PAL-1 translation within anterior blastomeres that normally would keep this maternally supplied mRNA silent, converting these blastomeres to a C-like fate. PAL-1-positive blastomeres can then be converted to either a bodywall muscle or hypodermal fate by manipulating POP-1 levels. Knocking down nuclear levels of POP-1 by RNAi directs cells to the bodywall muscle fate, whereas maintaining high nuclear POP-1 levels (with *lit-1* RNAi) directs blastomeres to the hypodermal fate. PD4743 on its own has GFP expression primarily in C- and D-derived bodywall muscle cells in 79% (n=192) of the embryos with an average of nine GFP-positive cells per embryo. RNAi of *mex-3* in PD4743 has a relatively minor effect on these numbers (72% positive, n=129, 10 cell/embryo), as does double RNAi of *mex-3* and *pop-1* (66% positive, n=236, and nine cells/embryo). By contrast, double RNAi of *mex-3* and *lit-1* dramatically reduces both the number of GFP-positive embryos and the number of positive cells per embryo (18% positive, n=173, and three cells/embryo). We concluded that the response of JKL26 reporters to PAL-1 activity is POP-1 dependent, either directly or indirectly.

DISCUSSION

Our results provide the first direct molecular evidence for a downstream transcriptional target gene of the embryonic cell fate regulator PAL-1. This links directly the action of maternally supplied PAL-1 to the activation of the zygotic myogenic regulator *hlh-1* in bodywall muscle precursors, consistent with previous genetic and expression studies (Baugh et al., 2005a; Fukushige et al., 2006). We have limited evidence that *unc-120* may also be a direct PAL-1 target gene in the posterior muscle lineage. The molecular details for posterior bodywall muscle can now be modeled (Fig. 7). PAL-1 activation of *hlh-1* (and perhaps *unc-120*) initiates a self-sustaining positive-feedback loop through HLH-1 that results in an irreversible commitment of these cells to the bodywall muscle cell fate. Positive auto-regulation of HLH-1 is robust because of the presence of multiple HLH-1-binding sites in its own enhancer regions. We find no evidence that the third factor in the previously described muscle module, HND-1 (Baugh et al., 2005a; Baugh et al., 2005b; Fukushige et al., 2006), is directly controlled by PAL-1 (see Fig. 5). It is possible that HND-1 is controlled independently of PAL-1 to provide a transient and redundant function to HLH-1 in these lineages. Indeed, HND-1 is expressed in multiple embryonic lineages and has multiple roles during development in *C. elegans* and other organisms (Cserjesi et al., 1995; Firulli, 2003; Mathies et al., 2003; Srivastava et al., 1995). Therefore, the role of HND-1 is distinct from the main transcriptional activators HLH-1 and UNC-120 that persist through post-embryonic development and that represent the core transcriptional engine driving myogenesis and post-embryonic muscle cell growth in our model.

Our evidence demonstrates, both in vivo and in vitro, that PAL-1 can bind the sequence ATTTATGAC. We have begun to analyze the distribution of this sequence throughout the genome of *C. elegans* to determine whether other possible PAL-1 target genes can be identified. Previous studies of potential direct or indirect downstream targets of PAL-1 in both hypodermal and bodywall muscle cells identified and validated 21 genes (Baugh et al., 2005a). Six of those genes, including *hlh-1* and *unc-120*, have a single perfect match to this 9 bp consensus sequence within the 4 kb upstream of the translational start codon; the other positive genes in this group are *spp-10*, R02D3.1, *cwn-1* and *pal-1* itself. By comparison, a search of 71 randomly selected genes had no sequences matching this consensus within the 4 kb putative

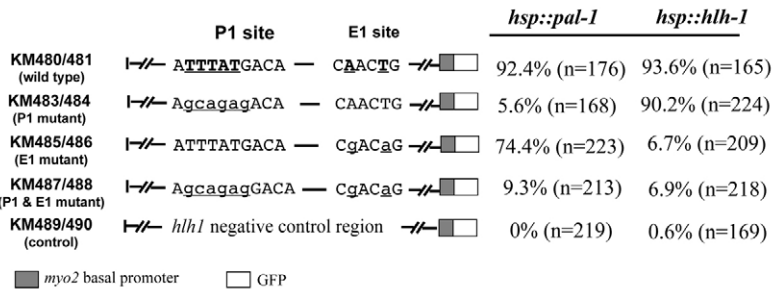


Fig. 6. PAL-1 and HLH-1 function through the P1 and E1 sites of *hlh-1* enh-1, respectively. The *hlh-1* enh-1 reporter gene constructs are diagrammed with the sequences of the P1 and E1 sites detailed. Wild-type (top) and mutant enh-1 sequences in which one or both of the sites have been altered were tested as integrated transgenes for their ability to respond to overproduction of either PAL-1 or HLH-1 in embryos. A region of the *hlh-1* promoter lacking muscle enhancer activity served as a negative control (see Fig. 1). The percentage of embryos with *gfp* reporter gene expression at 4.5 hours after heat-shock treatment is shown. Note that P1 is required for response to PAL-1, E-1 is required for response to HLH-1, and elimination of both P1 and E1 significantly drops the responsiveness of the reporter gene to either factor. Names of two independent strains tested for each integrated reporter gene construct are listed on the left.

promoter regions. Restricting the consensus sequence to 8 bp (ATTTATGA) hits nine of the previously identified 21 genes compared with 15/71 random genes; restricting to 7 bp (ATTTATG) hits 14 of the 21 genes compared with 25/71 random genes. We conclude that our results provide improved insight into potential direct PAL-1 target genes and specific sites that mediate PAL-1 binding, but that additional binding site specificity is probably provided by chromatin context and/or interacting proteins. This underscores the difficulty in assigning functional significance to genome-scale assays for transcription factor binding sites and highlights the need for experimental validation.

PAL-1 is necessary for both hypodermal (skin) and muscle development from the embryonic C founder blastomere lineage. Left unanswered is ‘what is the molecular basis for PAL-1 discriminating between these two disparate cell fates?’, a decision that depends (at least in part) on the relative levels of POP-1/TCF (Lin et al., 1995; Fukushige et al., 2006). An intriguing possibility is that the conserved sequence that resembles a *C. elegans* POP-1 binding site in the enh-1 region of *hlh-1* (Block 2) mediates part of this switch. Located next to the PAL-1-binding site, occupancy of this site by POP-1 in hypodermal lineages that have relatively high levels of nuclear POP-1 could interfere with PAL-1 binding, thus repressing *hlh-1* in these non-muscle lineages (Fig. 7). Conversely, the relatively low levels of POP-1 that are present in muscle cell precursors could allow PAL-1 to bind to enh-1 and activate *hlh-1*, as well as be permissive for HLH-1-positive auto-regulation that also occurs through enh-1. Consistent with this notion, our reporter gene studies demonstrate a role of POP-1 in determining enh-1-driven expression in response to PAL-1. However, our inability to demonstrate in vitro binding of POP-1 to Block 2 sequences suggests that another factor may mediate this response. A test of five additional genes (*hmg-1.2*, *hmg-3*, *hmg-4*, *hmg-12* and *sem-2*; see Fig. S6 in the supplementary material) encoding HMG-containing factors present in early embryogenesis failed to reveal an essential role for these POP-1-related factors in myogenesis. So, although our results reaffirm the crucial role for POP-1 in regulating gene expression, it leaves unanswered the molecular basis for its function. Clearly, understanding the molecular mechanisms underlying POP-1 function during embryogenesis will go a long way to revealing the logic of cell fate specification.

From a methodological perspective, our results demonstrate the feasibility of using ChIP in *C. elegans* to determine transcription factor binding sites in vivo. The lack of homogeneity among cells within the developing *C. elegans* embryo makes ChIP of cell type-specific transcription factors challenging. Indeed, we were unable to generate a ChIP signal from endogenous levels of HLH-1 when

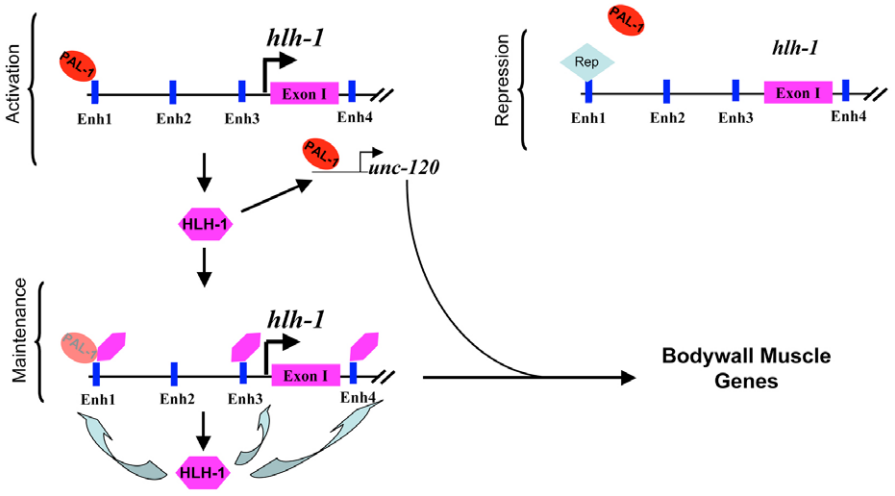


Fig. 7. A model of PAL-1 activation of the muscle module in posterior C and D lineages of the embryo. PAL-1 binds to enh-1 and activates *hlh-1* expression (and possibly *unc-120*) in C and D lineage descendents destined to become bodywall muscle. HLH-1 functions as a positive auto-regulatory factor for its own expression, acting through multiple enhancer elements, and also activates *unc-120*. Together, HLH-1 and UNC-120 activate downstream muscle-specific genes. In C lineage descendent blastomeres that are not fated to become muscle, an unidentified repressor blocks PAL-1 activation through enh-1, ensuring that *hlh-1* expression and its positive auto-regulation remains switched off.

we used multiple antibodies and strains that were transgenic for a multicopy *hlh-1* promoter-driven reporter gene that included the upstream enhancers that we have demonstrated here bind HLH-1 in vivo. This is a theme that will be common for many of the existing reagents for specific DNA-binding proteins. We have shown that these limitations can be overcome by either epitope-tagging and/or overexpressing the factor of interest. Although overexpression from heat-shock promoters far exceeds physiological levels when measurable, the technique remains a viable option. We were able to identify DNA-binding sites in vivo using ChIP after overexpression of either PAL-1 or HLH-1 and validate these sites by extensive mutational analysis, demonstrating that at least a subset of sites identified this way are biologically relevant. Thus, the use of epitope-tagged factors in transgenic animals offers the possibility to use ChIP to decipher DNA-binding elements in the *C. elegans* embryo and in larvae.

This work was supported by NIH grants R01-GM066953 to J.L. and R01-GM37706 to A.F., and by the Intramural Research Program of the NIH, NIDDK. Deposited in PMC for release after 12 months.

Supplementary material

Supplementary material for this article is available at
http://dev.biologists.org/cgi/content/full/136/8/1241/DC1

References

Arata, Y., Kouike, H., Zhang, Y., Herman, M. A., Okano, H. and Sawa, H. (2006). Wnt signaling and a Hox protein cooperatively regulate psa-3/Meis to determine daughter cell fate after asymmetric cell division in C. elegans. *Dev. Cell* **11**, 105-115.

Bianchi, M. E. and Agresti, A. (2005). HMG proteins: dynamic players in gene regulation and differentiation. *Curr. Opin. Genet. Dev.* **15**, 496-506.

Baugh, L. R. and Hunter, C. P. (2006). MyoD, modularity, and myogenesis: conservation of regulators and redundancy in C. elegans. *Genes Dev* **20**, 3342-3346.

Baugh, L. R., Hill, A. A., Claggett, J. M., Hill-Harfe, K., Wen, J. C., Slonim, D. K., Brown, E. L. and Hunter, C. P. (2005a). The homeodomain protein PAL-1 specifies a lineage-specific regulatory network in the C. elegans embryo. *Development* **132**, 1843-1854.

Baugh, L. R., Wen, J. C., Hill, A. A., Slonim, D. K., Brown, E. L. and Hunter, C. P. (2005b). Synthetic lethal analysis of Caenorhabditis elegans posterior embryonic patterning genes identifies conserved genetic interactions. *Genome Biol.* **6**, R45.

Blackwell, T. K. and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**, 1104-1110.

Buckingham, M. and Relaix, F. (2007). The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu. Rev. Cell Dev. Biol.* **23**, 645-673.

Charge, S. B. and Rudnicki, M. A. (2004). Cellular and molecular regulation of muscle regeneration. *Physiol. Rev.* **84**, 209-238.

Cserjesi, P., Brown, D., Lyons, G. E. and Olson, E. N. (1995). Expression of the novel basic helix-loop-helix gene eHAND in neural crest derivatives and extraembryonic membranes during mouse development. *Dev. Biol.* **170**, 664-678.

Dearolf, C. R., Topol, J. and Parker, C. S. (1989). The caudal gene product is a direct activator of fushi tarazu transcription during Drosophila embryogenesis. *Nature* **341**, 340-343.

Edgar, L. G., Carr, S., Wang, H. and Wood, W. B. (2001). Zygotic expression of the caudal homolog pal-1 is required for posterior patterning in Caenorhabditis elegans embryogenesis. *Dev. Biol.* **229**, 71-88.

Firulli, A. B. (2003). A HANDful of questions: the molecular biology of the heart and neural crest derivatives (HAND)-subclass of basic helix-loop-helix transcription factors. *Gene* **312**, 27-40.

Fukushige, T. and Krause, M. (2005). The myogenic potency of HLH-1 reveals wide-spread developmental plasticity in early C. elegans embryos. *Development* **132**, 1795-1805.

Fukushige, T., Brodigan, T. M., Schriefer, L. A., Waterston, R. H. and Krause, M. (2006). Defining the transcriptional redundancy of early bodywall muscle development in C. elegans: evidence for a unified theory of animal muscle development. *Genes Dev.* **20**, 3395-3406.

Hunter, C. P. and Kenyon, C. (1996). Spatial and temporal controls target pal-1 blastomere-specification activity to a single blastomere lineage in C. elegans embryos. *Cell* **87**, 217-226.

Kophengnavong, T., Michnowicz, J. E. and Blackwell, T. K. (2000). Establishment of distinct MyoD, E2A, and twist DNA binding specificities by different basic region-DNA conformations. *Mol. Cell. Biol.* **20**, 261-272.

Korswagen, H. C., Herman, M. A. and Clevers, H. C. (2000). Distinct beta-catenins mediate adhesion and signalling functions in C. elegans. *Nature* **406**, 527-532.

Krause, M., Fire, A., Harrison, S. W., Priess, J. and Weintraub, H. (1990). CeMyoD accumulation defines the body wall muscle cell fate during C. elegans embryogenesis. *Cell* **63**, 907-919.

Krause, M., Harrison, S. W., Xu, S. Q., Chen, L. and Fire, A. (1994). Elements regulating cell- and stage-specific expression of the C. elegans MyoD family homolog hlh-1. *Dev. Biol.* **166**, 133-148.

Krause, M., Park, M., Zhang, J. M., Yuan, J., Harfe, B., Xu, S. Q., Greenwald, I., Cole, M., Paterson, B. and Fire, A. (1997). A C. elegans E/Daughterless bHLH protein marks neuronal but not striated muscle development. *Development* **124**, 2179-2189.

Lam, N., Chesney, M. A. and Kimble, J. (2006). Wnt signaling and CEH-22/tinman/Nkx2.5 specify a stem cell niche in C. elegans. *Curr. Biol.* **16**, 287-295.

Lin, R., Thompson, S. and Priess, J. R. (1995). pop-1 encodes an HMG box protein required for the specification of a mesoderm precursor in early C. elegans embryos. *Cell* **83**, 599-609.

Maduro, M. F., Kasmir, J. J., Zhu, J. and Rothman, J. H. (2005). The Wnt effector POP-1 and the PAL-1/Caudal homeoprotein collaborate with SKN-1 to activate C. elegans endoderm development. *Dev. Biol.* **285**, 510-523.

Mathies, L. D., Henderson, S. T. and Kimble, J. (2003). The C. elegans Hand gene controls embryogenesis and early gonadogenesis. *Development* **130**, 2881-2892.

Maurer, C. W., Chiorazzi, M. and Shaham, S. (2007). Timing of the onset of a developmental cell death is controlled by transcriptional induction of the C. elegans ced-3 caspase-encoding gene. *Development* **134**, 1357-1368.

Okkema, P. G., Harrison, S. W., Plunger, V., Aryana, A. and Fire, A. (1993). Sequence requirements for myosin gene expression and regulation in Caenorhabditis elegans. *Genetics* **135**, 385-404.

Shetty, P., Lo, M. C., Robertson, S. M. and Lin, R. (2005). C. elegans TCF protein, POP-1, converts from repressor to activator as a result of Wnt-induced lowering of nuclear levels. *Dev. Biol.* **285**, 584-592.

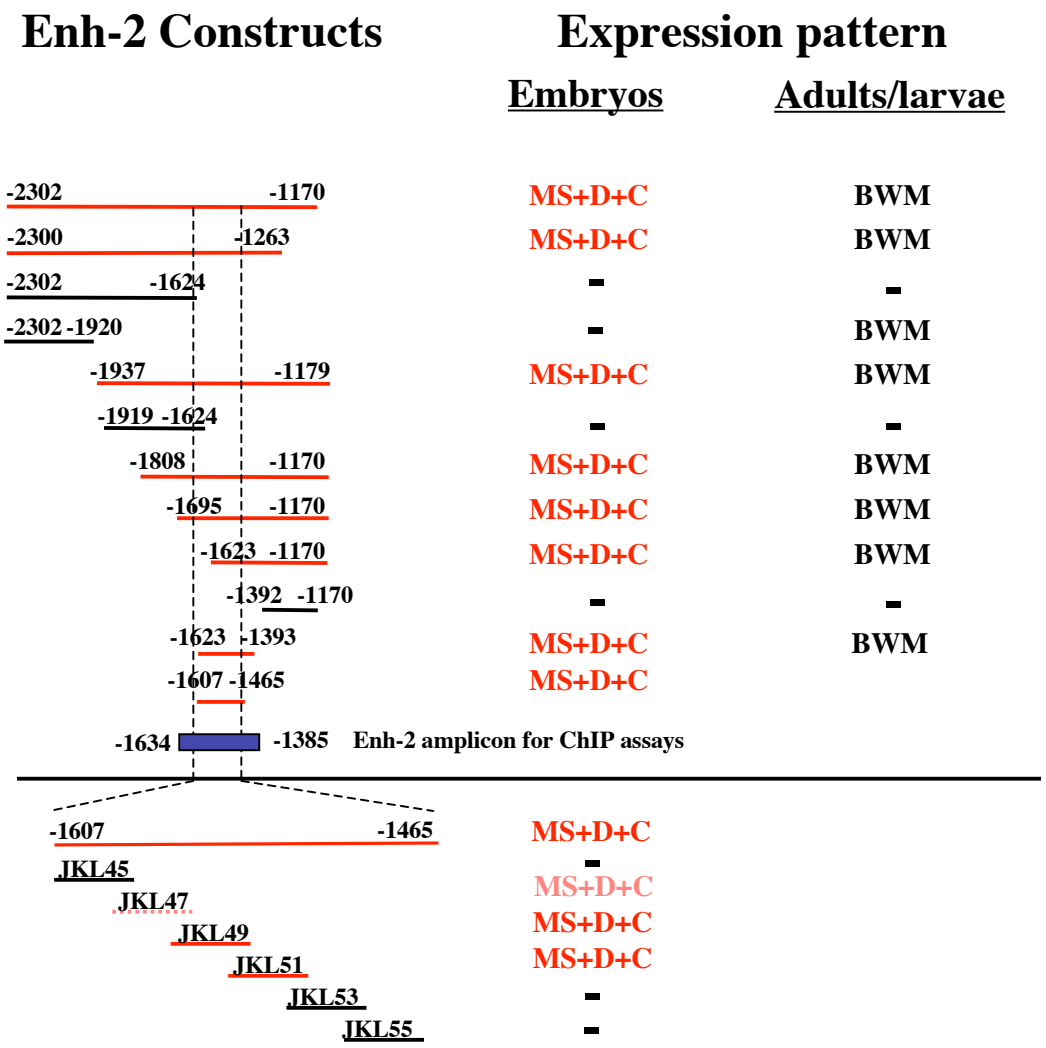
Srivastava, D., Cserjesi, P. and Olson, E. N. (1995). A subclass of bHLH proteins required for cardiac morphogenesis. *Science* **270**, 1995-1999.

Tapscott, S. J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132**, 2685-2695.

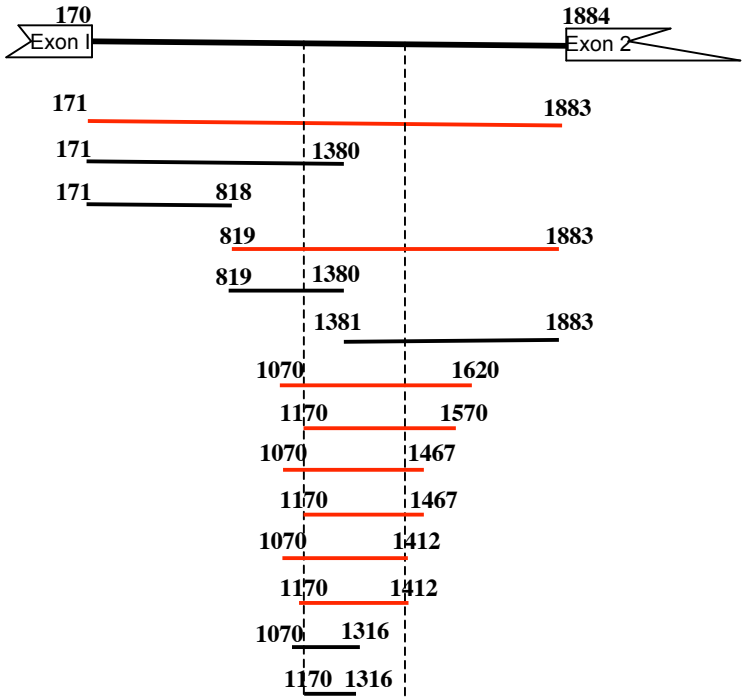
Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev* **5**, 880-894.

Waterman, M. L., Fischer, W. H. and Jones, K. A. (1991). A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. *Genes Dev* **5**, 656-669.

Yanai, I., Baugh, L. R., Smith, J. J., Roehrig, C., Shen-Orr, S. S., Claggett, J. M., Hill, A. A., Slonim, D. K. and Hunter, C. P. (2008). Pairing of competitive and topologically distinct regulatory modules enhances patterned gene expression. *Mol. Syst. Biol.* **4**, 163.

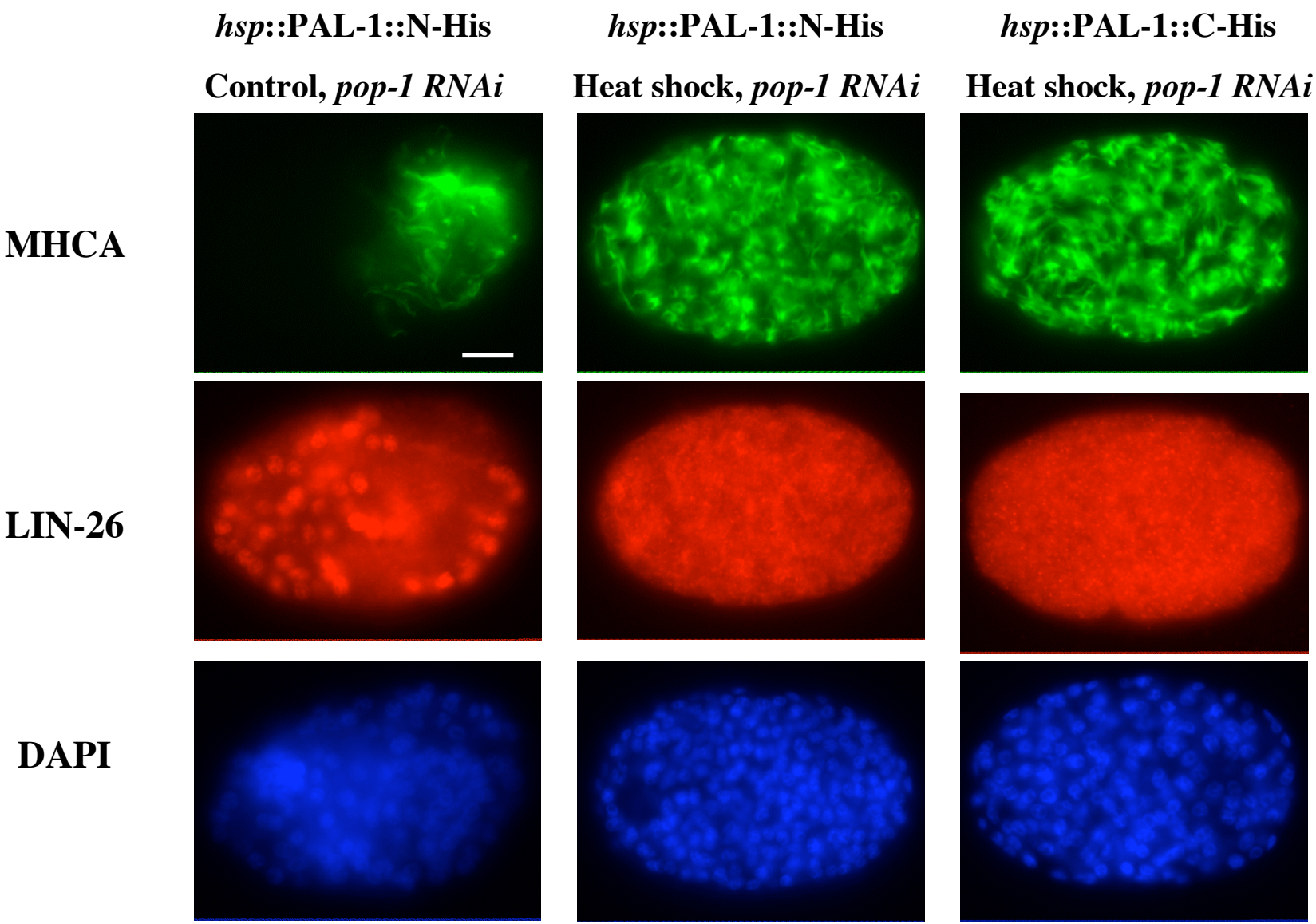

Enh-2 Constructs
Expression pattern
Embryos
Adults/larvae
-2302 -1170 MS+D+C BWM
-2300 -1263 MS+D+C BWM
-2302 -1624 - -
-2302 -1920 - BWM
-1937 -1179 MS+D+C BWM
-1919 -1624 - -
-1808 -1170 MS+D+C BWM
-1695 -1170 MS+D+C BWM
-1623 -1170 MS+D+C BWM
-1392 -1170 - -
-1623 -1393 MS+D+C BWM
-1607 -1465 MS+D+C
-1634 -1385 Enh-2 amplicon for ChIP assays
-1607 -1465 MS+D+C
JKL45 -
JKL47 MS+D+C
JKL49 MS+D+C
JKL51 MS+D+C
JKL53 -
JKL55 -

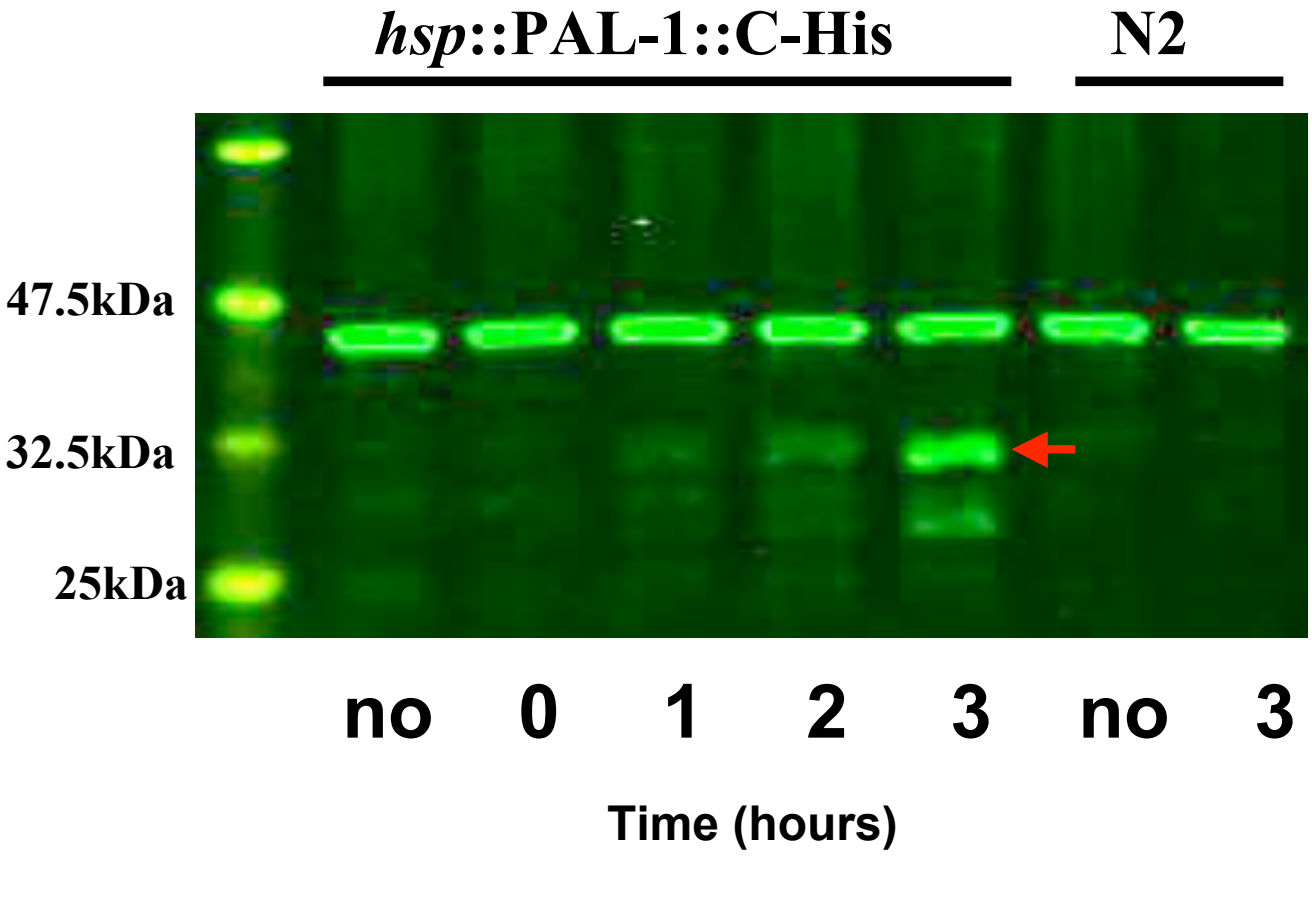
Enh-4 (Intron) Constructs

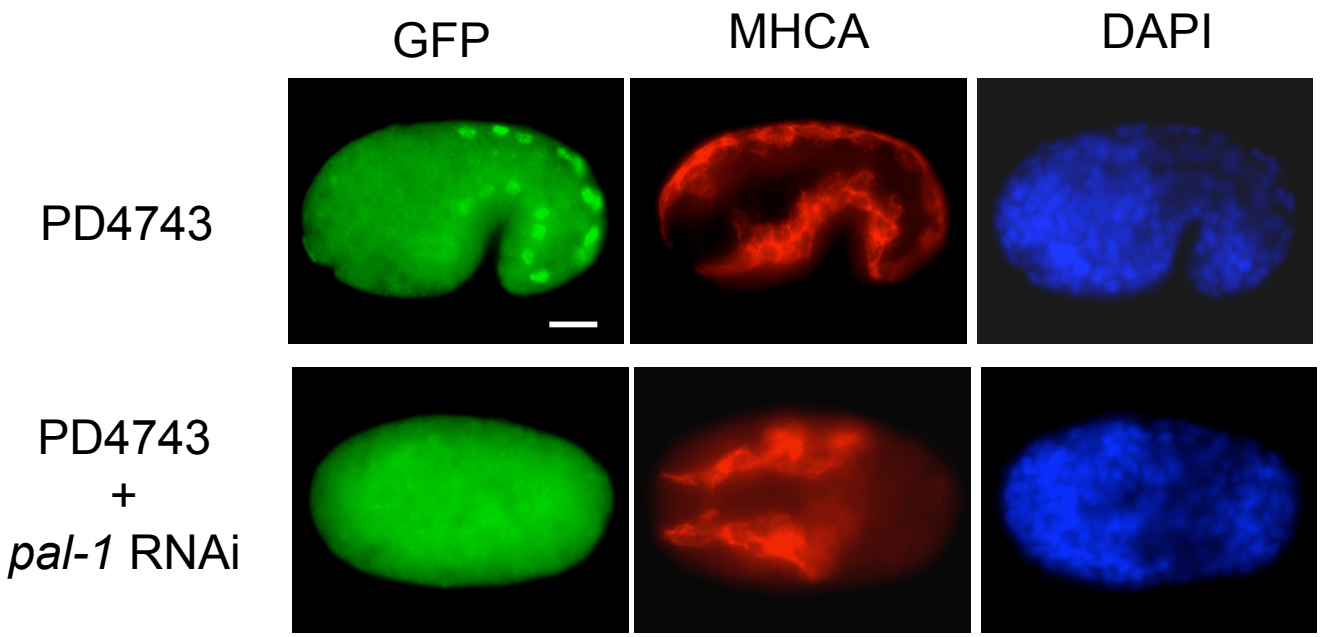


Expression pattern

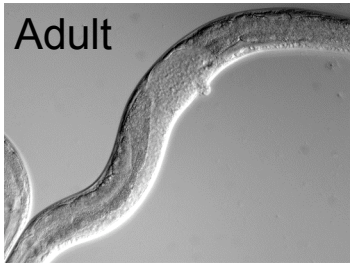
Embryos	Adults/larvae
MS+D+C	BWM, GLR
GLR	GLR
GLR	GLR
MS+D+C	BWM
-	-
-	-
MS+D+C	BWM
MS+D+C	BWM
MS+D+C	BWM
MS+D+C	BWM
MS+D+C	BWM
MS+D+C	BWM
-	-
-	-


hsp::PAL-1::N-His
Control, *pop-1 RNAi*
hsp::PAL-1::N-His
Heat shock, *pop-1 RNAi*
hsp::PAL-1::C-His
Heat shock, *pop-1 RNAi*
MHCA
LIN-26
DAPI

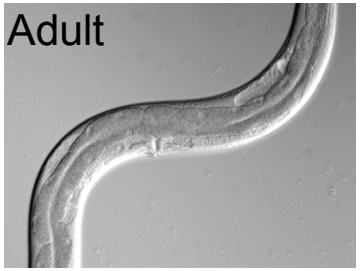

hsp::PAL-1::C-His
N2
47.5kDa
32.5kDa
25kDa
no
0
1
2
3
no
3
Time (hours)


GFP
MHCA
DAPI
PD4743
PD4743
+
pal-1 RNAi

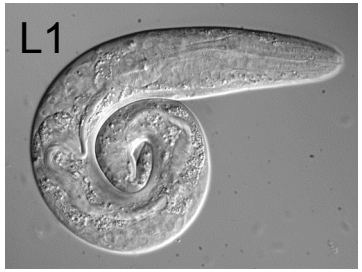
hmg-1.2 RNAi
~95% Sterile

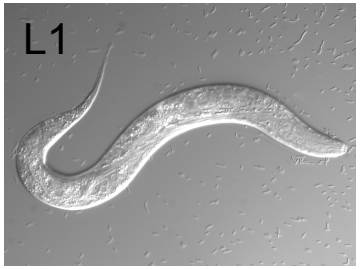


hmg-3 RNAi
~95% Sterile



hmg-4 RNAi
100% L1 Let





sem-2 RNAi
~30% emb/L1 Let
~70% Egl

Table S1. Enhancer 1 oligo concatenate reporter gene constructs and expression patterns
Bolded oligos are wild type sequence shown in figure; oligos shown below bold pair are mutant derivatives
Expression patterns in bodywall muscle cells or their precursors (bwm) are highlighted in red; n.d = not detected (wild type) or not tested (*hlh-1* mutant).
Expression result indicate number of lines showing the patten of expression described out total number of lines generated

			Wild Type Expression		hlh-1 Mutant Embryonic Expression			
Oligo Pair	**Copy #**	**# of Lines**	**Embryos**	**Adults**	**heterozygous**	**homozygous**	**[DNA] for injections**	**Array Type**
JKL 22/23	1	1	background	many cells	n.d.	n.d.	8ng/ul	complex
	2	3	1 of 3 D+C at comma	2 of 3 faint bwm	n.d.	n.d.	5ng/ul	complex
	4	1	faint to none	dense body signal	n.d.	n.d.	8ng/ul	complex
	4	1	n.d.	n.d.	none	none	8ng/ul	complex
	8	2	faint to none	gut, 1 of 2 bwm	n.d.	n.d.	5ng/ul	complex
	4	6	none	none	n.d.	n.d.	60ng/ul	simple
	8	4	2 of 4 few random cells	1 of 4 gut	n.d.	n.d.	60ng/ul	simple
JKL 24/25	3	2	background	pharyngeal	n.d.	n.d.	8ng/ul	complex
	4	2	background + D+C	head +tail neurons	n.d.	n.d.	8ng/ul	complex
	4	3	n.d.	n.d.	faint D+C	faint D+C	8ng/ul	complex
	6	2	1 of 2 faint D+C	neurons and gut	n.d.	n.d.	8ng/ul	complex
	4	3	ccaisional and multiple cel	head, tail, gut cells	n.d.	n.d.	60ng/ul	simple
	8	5	none	faint head, gut cells	n.d.	n.d.	60ng/ul	simple
JKL 26/27	2	3	faint to none	faint to none	n.d.	n.d.	4ng/ul	complex
	4	1	n.d.	n.d.	D+C bwm	D+C bwm	8ng/ul	complex
	4	5	D+C bwm	1 of 5 bwm	n.d.	n.d.	8ng/ul	complex
	8	6	D+C bwm	5 of 6 bwm, gut	n.d.	n.d.	4ng/ul	complex
	4	1	none	none	n.d.	n.d.	60ng/ul	simple
	8	8	none	none	n.d.	n.d.	60ng/ul	simple
JKL 58/59	2	4	none	faint pharynx	n.d.	n.d.	7ng/ul	complex
	4	6	strong anterior cells	strong pharynx	n.d.	n.d.	5ng/ul	complex
								complex
JKL 60/61	3	2	D+C bwm	faint bwm	n.d.	n.d.	6ng/ul	complex
	6	6	none	none	n.d.	n.d.	8ng/ul	complex
JKL 62/63	2	5	none	faint pharyngeal	n.d.	n.d.	6ng/ul	complex
	4	5	D+C bwm	bwm, faint gut	n.d.	n.d.	5ng/ul	complex
JKL 64/65	3	2	faint signal	pharyngeal, gut cells	n.d.	n.d.	6ng/ul	complex
	6	2	D+C bwm	bwm, faint gut	n.d.	n.d.	5ng/ul	complex
JKL 114/115	4	3	none	none, faint hyp	n.d.	n.d.	7ng/ul	complex
JKL 116/117	4	7	none	4 of 7 bwm, gut	n.d.	n.d.	7ng/ul	complex
JKL 118/119	4	1	weak D+C	pharyngeal, gut cells	n.d.	n.d.	7ng/ul	complex
JKL 120/121	4	4	2 of 4 weak D+C	bwm, gut cells	n.d.	n.d.	7ng/ul	complex
JKL 28/29	3	7	5 of 7 D+C bwm	faint,mosaic bwm	D+C bwm	none	5ng/ul	complex
	6	10	8 of 10 D+C bwm	mosaic bwm	n.d.	n.d.	8ng/ul	complex
JKL 30/31	4	2	D+C bwm, anterior cells	strong bwm	n.d.	n.d.	8ng/ul	complex
	8	1	D+C bwm, anterior cells	strong bwm	n.d.	n.d.	4ng/ul	complex
	8	1	n.d.	n.d.	D+C bwm	none	4ng/ul	complex
	4	1	D+C in older emb	mosaic bwm, gut	n.d.	n.d.	60ng/ul	simple
	8	3	D+C bwm, anterior cells	mosaic bwm	n.d.	n.d.	60ng/ul	simple
JKL 66/67	2	1	none	few gut cells	n.d.	n.d.	6ng/ul	complex
	4	2	anterior cells and C?	pharyngeal, gut cells	n.d.	n.d.	6ng/ul	complex
JKL 68/69	10	2	1 of 2 faint anterior	bwm	n.d.	n.d.	6ng/ul	complex
JKL 70/71	2	6	none	faint pharyngeal	n.d.	n.d.	7ng/ul	complex
	4	2	none	faint bwm	n.d.	n.d.	5ng/ul	complex
JKL 72/73	2	3	faint random	strong pharyngeal	n.d.	n.d.	7ng/ul	complex
	4	1	faint D+C	gut cells	n.d.	n.d.	8ng/ul	complex
JKL 74/75	4	3	strong, mosaic D+C	bwm, pharyngeal	n.d.	n.d.	5ng/ul	complex
JKL 32/33	2	3	anterior cells	pharyngeal, gut cells	n.d.	n.d.	8ng/ul	complex
	4	3	none	pharyngeal cells	n.d.	n.d.	5ng/ul	complex
	8	5	random	3 of 5 bwm	n.d.	n.d.	8ng/ul	complex
JKL 34/35	4	1	anterior cells	gut and head cells	n.d.	n.d.	7ng/ul	complex
	8	3	anterior cells	gut and head cells	n.d.	n.d.	5ng/ul	complex
JKL 43/44	2	3	faint background	gut, pharyngeal cells	n.d.	n.d.	5ng/ul	complex
	4	4	none	pharyngeal, hyp cells	n.d.	n.d.	5ng/ul	complex
	8	4	none	pharyngeal, faint gut	n.d.	n.d.	5ng/ul	complex

Table S2. Enhancer 2 oligo concatentate reporter gene constructs and expression patterns
Bolded oligos are wild type sequence shown in figure; oligos shown below bold pair are mutant derivatives
Expression patterns in bodywall muscle cells or their precursors (bwm) are highlighted in red; n.d = not dete
Expression result indicate number of lines showing the patten of expression described out total number of lines generated

Oligo Pair	Copy #	# of Lines	Wild Type Expression		[DNA] for injections	Array Type
			Embryos	Adults		
JKL 45/46	2	3	1 of 3 anterior cells	head cells	5ng/ul	complex
	4	4	anterior cells	pharyngeal cells	7ng/ul	complex
JKL 47/47	2	2	1 of 2 faint D+C	bwm, gut, pharyngeal	4ng/ul	complex
	4	3	2 of 3 faint D+C	2 of 3 bwm, gut	5ng/ul	complex
JKL 92/93	2	3	1 of 3 anterior cells	head cells	8ng/ul	complex
	5	3	background	head cells	5ng/ul	complex
JKL 94/95	4	8	none	3 of 8 faint head	5ng/ul	complex
JKL 96/97	3	2	none	faint head cells	8ng/ul	complex
JKL 98/99	4	4	none	none	7ng/ul	complex
JKL 132/133	4	6	MS+C+D bwm, others	none	5ng/ul	complex
JKL 134/135	3	7	none	none	8ng/ul	complex
JKL 136/137	4	1	anterior cells	none	5ng/ul	complex
JKL 138/139	3	4	none	none	8ng/ul	complex
	7	2	anterior cells	1 of 2 bwm	8ng/ul	complex
JKL 49/50	2	4	mosaic MS+D+C bwm	1 pf 2 bwm	5ng/ul	complex
	4	6	5 of 6 MS+D+C bwm	bwm	6ng/ul	complex
JKL 100/101	4	2	MS+D+C bwm	bwm, head cells	8ng/ul	complex
JKL 102/103	4	7	none	3 of 7 faint bwm	5ng/ul	complex
JKL 104/105	4	2	1 of 2 MS+D+C bwm	faint gut	7ng/ul	complex
JKL 140/141	3	2	none	none	5ng/ul	complex
	8	6	4 of 6 MS+D+C bwm	5 of 6 bwm	5ng/ul	complex
JKL 142/143	4	1	MS+D+C bwm	none	5ng/ul	complex
JKL 51/52	2	3	none	2 of 3 bwm	8ng/ul	complex
	4	2	MS+D+C bwm	strong bwm	6ng/ul	complex
JKL 106/107	5	6	MS+D+C bwm	bwm, head cells	10ng/ul	complex
JKL 108/109	4	3	1 of 3 MS+D+C bwm	faint head, tail cells	5ng/ul	complex
JKL 110/111	4	4	none	2 of 4 faint bwm	5ng/ul	complex
JKL 126/127	4	7	3 of 7 faint MS+D+C bwm	3 of 7 some bwm	7ng/ul	complex
JKL 144/145	9	5	MS+D+C bwm	bwm	5ng/ul	complex
	21	5	4 of 5 MS+D+C bwm	4 of 5 bwm	5ng/ul	complex
JKL 146/147	8	1	MS+D+C bwm, others	strong bwm, others	5ng/ul	complex
JKL 160/161	8	2	unknown cells	faint bwm, pharyngeal, gut	8ng/ul	complex
JKL 162/163	7	2	unknown cells	mosiac bwm	8ng/ul	complex
JKL 164/165	11	2	pharyngeal, other cells	pharyngeal, gut, vuval	8ng/ul	complex
JKL 166/167	5	2	unknown cells	mosaic bwm, pharyngeal, gut	8ng/ul	complex
	8	2	unknown cells	pharyngeal, gut cells	8ng/ul	
JKL 53/54	2	3	none	pharyngeal cells	6ng/ul	complex
	4	4	faint unknown cells	3 of 4 faint bwm, gut	5ng/ul	complex
JKL 55/56	3	3	unknown cells	head and tail cells	8ng/ul	complex
	6	1	background	pharyngeal cells	5ng/ul	complex