Published in final edited form as:

Dev Biol. 2017 September 01; 429(1): 335–342. doi:10.1016/j.ydbio.2017.06.011.

The *C. elegans* Spalt-like protein SEM-4 functions through the SoxC transcription factor SEM-2 to promote a proliferative blast cell fate in the postembryonic mesoderm

Qinfang Shen, Herong Shi, Chenxi Tian, Vikas Ghai, and Jun Liu*

Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853

Abstract

Proper development of a multicellular organism relies on well-coordinated regulation of cell fate specification, cell proliferation and cell differentiation. The *C. elegans* postembryonic mesoderm provides a useful system for uncovering factors involved in these processes and for further dissecting their regulatory relationships. The single Spalt-like zinc finger containing protein SEM-4/SALL is known to be involved in specifying the proliferative sex myoblast (SM) fate. We have found that SEM-4/SALL is sufficient to promote the SM fate and that it does so in a cell autonomous manner. We further showed that SEM-4/SALL acts through the SoxC transcription factor SEM-2 to promote the SM fate. SEM-2 is known to promote the SM fate by inhibiting the expression of two BWM-specifying transcription factors. In light of recent findings in mammals showing that Sall4, one of the mammalian homologs of SEM-4, contributes to pluripotency regulation by inhibiting differentiation, our work suggests that the function of SEM-4/SALL proteins in regulating pluripotency versus differentiation appears to be evolutionarily conserved.

Keywords

SEM-4; SALL; SEM-2; SoxC; pluripotent; mesoderm; cell fate specification; M lineage; sex myoblast; coelomocyte; bodywall muscle; proliferation; differentiation

INTRODUCTION

During the development of a multicellular organism, cell fate specification, cell proliferation and cell differentiation need to be tightly regulated and precisely coordinated. The *C. elegans* postembryonic mesoderm provides a unique system for dissecting the regulatory network underlying these developmental processes. In *C. elegans*, all the postembryonic non-gonadal mesoderm cells arise from the M lineage. The M lineage is derived from a single pluripotent precursor cell, the M mesoblast. In a hermaphrodite animal, the M cell undergoes two waves of proliferation to produce thirty-two differentiated cells, which

*Corresponding author: Jun Liu, 439 Biotechnology Building, Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853. Tel: 607-254-8942. Fax: 607-255-6249. JL53@cornell.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

include fourteen striated body wall muscles (BWMs), two non-muscle coelomocytes (CCs), and sixteen non-striated egg-laying muscles of four different types (four type I and four type II vulval muscles and four type I and four type II uterine muscles) (Sulston and Horvitz, 1977). The sixteen egg-laying muscles are derived from two pluripotent cells, called the sex myoblasts (SMs). The BWMs, CCs and SMs are products of the first wave of M cell proliferation during the L1 larval stage, while all the egg-laying muscles are products of the SM cells during the second wave of M lineage proliferation. Importantly, the sister cells of the two SMs are differentiated BWMs (Figure 1A).

Multiple transcription factors and signaling pathways are known to play a role in the proper proliferation and subsequent specification and differentiation of the various cell types that are derived from the M lineage (Krause and Liu, 2012). In particular, we have previously shown that the SoxC transcription factor SEM-2 is required for the binary fate decision between the SMs and BWMs, and that SEM-2 is both necessary and sufficient to promote the proliferative SM fate and inhibit the differentiated BWM fate (Tian et al., 2011). In *sem-2* mutants, two redundant transcription factors required for specifying the BWM fate, HLH-1/MRF and FOZI-1 (Amin et al., 2007; Harfe et al., 1998), are de-repressed in the two presumptive SMs, and these two cells adopt the fate of their sister cells, the differentiated BWMs (Tian et al., 2011).

Another transcription factor, a zinc finger containing protein SEM-4 has also been shown to be required for specifying the SM fate (Basson and Horvitz, 1996). SEM-4 contains seven C2H2-type zinc finger motifs and is homologous to the Spalt-like (SALL) proteins in flies and mammals. SEM-4 is known to function in multiple cell types in *C. elegans*. It acts to restrict the touch neuron fate by directly binding to the promoters of the Hox factor EGL-5 and the LIM homeodomain protein MEC-3 to repress their expression (Toker et al., 2003). In the developing vulva, SEM-4 regulates the expression of the Hox factor LIN-39 to promote vulval cell fate determination (Grant et al., 2000). SEM-4 is also required for Y-to-PDA trans-differentiation, a natural epithelial-neuronal reprogramming event in *C. elegans* (Jarriault et al., 2008). It has been shown that the function of SEM-4 in this reprogramming process involves its association with the transcriptional repressor EGL-27/MTA1 in a NODE (Nanog and Oct4-associated deacetylase)-like complex (Jarriault et al., 2008; Kagias et al., 2012).

Despite the initial finding by Basson and Horvitz (Basson and Horvitz, 1996) that SEM-4 is required for specifying the SM fate, how SEM-4 functions in this process is not known. In this study, we dissected the role of SEM-4 and its relationship with SEM-2 in SM fate specification. We found that SEM-4 is both necessary and sufficient to promote the SM fate, and that it does so by acting through SEM-2.

MATERIALS AND METHODS

C. elegans strains

All strains were maintained at 20°C unless otherwise noted. The following strains were used in this study: *sem-2(n1343)*, *sem-4(n1971)/hT2[qIs48]*, *sem-4(n2477)/hT2[qIs48]*, *sem-4(n1378)/hT2[qIs48]*, *sem-4(h769)/hT2[qIs48]*, *sem-4(tm547)/hT2[qIs48]*,

sem-4(n2087)/hT2[qIs48]. Additional integrated transgenic lines used are: LW0001: *hT2[qIs48]/qC1; ayIs2(egl-15::GFP) IV; arIs39(secreted CC::GFP) X*. LW0081: *ccIs4438(intrinsic CC::GFP) III; ayIs2(egl-15::GFP) IV; ayIs6(hlh-8::GFP) X*. LW1380: *jjIs1380(egl-15::mRFP + intrinsic CC::mRFP + unc-119(+))*. LW1479: *jjIs1479(myo-3::NLS::mRFP::lacZ) + unc-119(+)* V. LW3900: *jjIs3900(hlh-8p::NLS::mCherry::lacZ + myo-2p::mCherry) IV*. LW4191: *jjIs4191(hlh-8p::sem-4 cDNA::sem-4 3'UTR + myo-2p::mCherry); ccIs4438(intrinsic CC::GFP) III; ayIs2(egl-15::GFP) IV; ayIs6(hlh-8::GFP) X*. PD4251: *ccIs4251(myo-3p::NLS::GFP::lacZ + myo-3p::mit::GFP + dpy-20(+))*. MH1337: *kuIs34(sem-4::GFP + unc-119(+))* and MH1346: *kuIs35(sem-4::GFP + unc-119(+))* (Grant et al., 2000).

Plasmid constructs, transgenic lines and RNAi

The *sem-4* cDNA clone *yk1509a11* (gift from Yuji Kohara, National Institute of Genetics, Japan), which contains the coding region and the 3'UTR of *sem-4*, was used to generate pSQF2 (*hlh-8p::sem-4cDNA::sem-4 3'UTR*).

The following *sem-4* transcriptional fusion reporter plasmids are gifts from Michael Koelle (Yale University): pMK326 (*SphI-SacII 11kb sem-4 genomic DNA including 25bp sem-4 ORF::GFP*), pMK329 (*MluI-SacII 4.5kb sem-4 genomic DNA::GFP*), pMK330 (*StuI-SacII 3.7kb sem-4 genomic DNA::GFP*), pMK331 (*SmaI-SacII 2.1kb sem-4 genomic DNA::GFP*), pMK332 (*NsiI-SacII, 0.4kb sem-4 genomic DNA::GFP*). pSQF14 (*4.7kb first intron::pes-10p::4xNLS::GFP::unc-54 3'UTR*) was used to assay for enhancer activity in the first intron of *sem-4*.

All *sem-2* containing constructs are described in (Tian et al., 2011), which include pJKL776 (*hlh-8p::sem-2cDNA::unc-54 3'UTR*), the *GFP::sem-2* translational fusion-containing fosmid, and pCXT33 (*598bp sem-2 M lineage enhancer::pes-10p::NLS::GFP::lacZ::unc-54 3'UTR*).

Transgenic lines were generated using the plasmid pCFJ90 (*myo-2p::mCherry::unc-54 3'UTR*) (Frokjaer-Jensen et al., 2008) and pJKL449 (*myo-2p::GFP::unc-54 3'UTR*) as markers.

sem-2 dsRNA was prepared using *yk404e6* as a template, and JKL-1585 and JKL-1586 as primers. In vitro transcription and RNAi by soaking were carried out as described in (Tian et al., 2011). Water was used as a soaking control.

Microscopy

Animals were visualized under a Leica DMRA2 compound microscope equipped with a Hamamatsu Orca-ER camera using the iVision software (Biovision technology, Inc.). Subsequent image processing was performed using Fiji (Schindelin et al., 2012).

RESULTS

***sem-4* mutants exhibit a fate transformation of the proliferative SMs to differentiated BWMs**

sem-4 mutations have been previously reported to affect the specification of M lineage-derived SMs and CCs (Basson and Horvitz, 1996), in which the SMs are transformed to their BWM siblings, and the CCs are transformed to their ventral counterparts, the SM mothers, which subsequently give rise to two BWMs each (Figure 1B, 1D). To further dissect the function of SEM-4 in M lineage development, we examined the M lineage phenotypes of six *sem-4* mutant alleles (Figure 1E) by introducing into each mutant three cell type specific GFP markers: *hlh-8p::GFP* that labels all undifferentiated cells in the M lineage, *egl-15p::GFP* that labels the SM-derived type I vulval muscles (VM1s), and a *secreted CC::GFP* that labels all CCs, including the two M-derived CCs (see materials and methods). We also introduced into *n1378* and *n1971* the *myo-3p::GFP* marker to label the BWMs. Examining the M lineage phenotypes of the *sem-4* mutants using these markers showed that different alleles exhibited different penetrance in their SM or CC mis-specification defects (Figure 1F). While all six alleles exhibited a high penetrance of the SM-to-BWM phenotype, the penetrance of the CC mis-specification phenotype in *n1971*, *tm547* and *n1378* are 75%, 69% and 59%, respectively. Previous work has shown that different *sem-4* mutant alleles exhibit different severity of their neuronal versus mesoderm defect (Basson and Horvitz, 1996; Toker et al., 2003). Our observations showed that even in the mesoderm, different *sem-4* mutations affect the CCs and the SMs differently. *sem-4* encodes a predicted protein that contains seven C2H2 type zinc finger (ZF) domains (Figure 1E). *tm547* contains an in-frame deletion that deletes ZF2 and ZF3, while *n1378* is a nonsense mutation that results in a truncated protein lacking ZF5, ZF6 and ZF7. The amino acid sequences for ZF2 and ZF3 closely resemble those of ZF5 and ZF6, particularly between ZF3 and ZF6 ((Basson and Horvitz, 1996); Figure 1G). It is possible that ZF2-3 and ZF5-6 share redundant functions in specifying M-derived CCs. The *n1971* allele is a splice site mutation that would result in a truncation of the SEM-4 protein before the first ZF, and it has been previously considered to be a null allele of *sem-4* (Basson and Horvitz, 1996). The incomplete penetrance of the CC defect of *n1971* mutants suggests that it may not be a null allele. It is possible that cryptic splice site(s) are used in *n1971* mutants to generate partially functional SEM-4 proteins for CC specification.

The high penetrance of the SM-to-BWM fate transformation phenotype shared by all six *sem-4* alleles suggests that SEM-4 is required for specifying the proliferative SM fate. We therefore focused our subsequent analysis on the role of SEM-4 in SM fate specification. We used the allele *h769* for our studies because it has a nonsense mutation in residue 275 (W275Stop) and exhibits 100% penetrance for both the SM-to-BWM and the CC-to-BWM phenotypes (Figure 1D, 1F).

***sem-4* is expressed throughout the M lineage**

To determine how SEM-4 functions to specify M-derived SMs, we examined the expression pattern of a *sem-4::GFP* partial translational fusion reporter (*kuIs34* and *kuIs35* [*sem-4::GFP* +*unc119(+)*] (Grant et al., 2000)) relative to M lineage development. As shown in Figure 2A–I, *sem-4::GFP* is expressed throughout the M lineage from the 1-M stage to the 16-SM

stage, and in the SM sub-lineage from the 2-SM stage to the 16-SM stage, before the cells differentiate. When these cells terminally differentiate, *sem-4::GFP* becomes undetectable in the differentiated CCs and vulval muscles (Figure 2JO), but remains expressed in the differentiated BWMs (Figure 2G–I).

We attempted to identify the cis-regulatory element(s) involved in regulating the expression of *sem-4* in the M lineage. As described above, the integrated *sem-4::GFP* partial translational fusion reporter (*kuIs34* and *kuIs35*) contains about 11kb upstream and 7.5kb exon/intron sequences of *sem-4*, and is expressed throughout the M lineage (Figure 3A). Using a series of promoter deletion constructs (the pMK plasmid series, Figure 3A) and an enhancer assay construct (pQFS14), we found that *sem-4* has at least two cis-regulatory elements mediating its expression in the early M lineage. One element is located within the 402bp region immediately upstream of the ATG (pMK332 in Figure 3A), and is sufficient to drive GFP expression from the 1-M stage till the stage when the 2 SMs are migrating (we have referred to as the 18-M/2-SM stage) (Figure 3A–G). Another element is located in the 4.7kb first intron (pSQF14 in Figure 3A), and is sufficient to drive GFP expression from the 1-M stage till the 16-M stage. Once the 2-SMs have migrated to the presumptive vulva region, none of these reporters showed GFP expression in the SM lineage (Figure 3H–J). While the lack of reporter expression in the SM lineage is simply due to the drop of GFP expression level below the threshold for detection, it is also possible that additional cis-regulatory element(s) is required for *sem-4* expression in the SM lineage.

***sem-4* functions within the M lineage to specify the SMs**

The expression pattern of *sem-4::GFP* in the M lineage suggests that it may function autonomously within the M lineage to specify M-derived SMs. We tested this hypothesis by expressing *sem-4* cDNA exclusively in the M lineage in the null *sem-4(h769)* mutants using the M lineage-specific *hlh-8* promoter. As shown in Table 1, forced expression of *sem-4* in the M lineage can partially rescue the M lineage phenotype of *sem-4(h769)* mutants: 12% of the transgenic animals (n=102) re-gained the normal number of SMs. Moreover, 11% of the transgenic animals had extra SM-like cells. Taken together, forced expression of *sem-4* cDNA exclusively in the M lineage could restore the SM fate in 23% of *sem-4(h769)* transgenic animals, indicating that SEM-4 functions autonomously within the M lineage to specify M-derived SMs. The low penetrance of rescue (23%) might be due to the mosaic nature of the extra chromosomal array carrying the *hlh-8p::sem-4 cDNA* transgene. The observed 11% penetrance of the extra SM-phenotype could be a result of *sem-4* overexpression (see below).

Overexpression of *sem-4* in the M lineage is sufficient to convert most M-derived cells to SM-like cells

To directly determine the consequences of *sem-4* overexpression in the M lineage, we introduced the *hlh-8p::sem-4 cDNA* transgene into wild-type worms that has normal endogenous *sem-4* expression and generated stable lines with the transgene integrated into the genome. As shown in Figure 4 and Table 1, 88% of these transgenic worms (N=130) had extra number of SM-like cells and a concurrent loss of M-derived CCs. When we directly followed M lineage development in 30 of these transgenic animals from the 1-M stage to the

18-M stage, we found that in 26 of them both M-derived CCs and variable number of M-derived BWMs were transformed to SM-like cells. The loss of M-derived CCs and some M-derived BWMs in these animals were confirmed using the *CC::GFP* reporter and the *myo-3p::NLS::mRFP::lacZ* reporter, which labels the BWMs (Figure 4). Taken together, the above results indicate that high levels of SEM-4 in the M lineage is sufficient to drive most M lineage cells to the proliferative SM-like fate.

***sem-4* is required for the M lineage specific expression of the SoxC transcription factor SEM-2**

We have previously shown that SEM-2, the sole SoxC protein in *C. elegans*, is both necessary and sufficient to promote the proliferating SM fate (Tian et al., 2011). Because SEM-4 is also necessary and sufficient for specifying the SM fate, we dissected the regulatory relationship between SEM-2 and SEM-4. We first examined the expression of *sem-4::GFP* in the M lineage in *sem-2(n1343)* mutants. *sem-2(n1343)* is a transposon-induced mutation that specifically disrupts *sem-2* expression and function in the M lineage (Tian et al., 2011). As shown in Figure 5A–C, *sem-4::GFP* expression in the M lineage is unaltered in *sem-2(n1343)* mutants compared to wild-type animals. In contrast, the M lineage expression of a functionally tagged GFP::SEM-2 is undetectable in *sem-4(h769)* mutants (Figure 5D–K). This loss of GFP::SEM-2 is specific to the M lineage because its expression outside of the M lineage remains unchanged (data not shown). These observations suggest that *sem-4* is specifically required for the expression of *sem-2* in the M lineage.

Expression of *sem-2* in the M lineage is mediated by an M lineage enhancer located in the first large intron of *sem-2* (Tian et al., 2011). We tested if SEM-4 acts through this enhancer element to regulate the M lineage expression of *sem-2*. We introduced a *GFP* reporter driven by this enhancer, pCXT33 (Tian et al., 2011), into the *sem-4(h769)* mutants. The level of M lineage expression of pCXT33 was not altered in *h769* animals compared to their heterozygous *h769/hT2[qIs48]* siblings (data not shown), but the percentage of transgenic worms showing M lineage expression of the pCXT33 reporter is slightly reduced in *h769* animals compared to their heterozygous *h769/hT2[qIs48]* siblings (Table 2). These observations suggest that this specific enhancer element contributes to, but is not essential for, the regulation of *sem-2* expression by SEM-4. At present, it is unclear how SEM-4 may regulate *sem-2* expression through this element, because we did not find any potential SEM-4 binding site(s) (Toker et al., 2003) within this element.

SEM-4 functions through SEM-2 to specify M-derived SMs

The loss of *sem-2* expression in the M lineage in *sem-4(h769)* mutants suggested that the role of SEM-4 in promoting SM fate specification may be mediated by SEM-2. To test this hypothesis, we first asked if forced expression of *sem-2* in *sem-4(h769)* mutants can restore the formation of M-derived SMs. As shown in Table 1, forced expression of *sem-2* specifically in the M lineage using the *hlh-8p::sem-2* cDNA construct can partially restore the formation of M-derived SMs in *sem-4(h769)* null mutants (14%, n=56; Table 1). We further found that the formation of SMs and the ectopic SM-like cells in transgenic worms overexpressing *sem-4* requires the presence of *sem-2*, because depleting *sem-2* by RNAi

soaking (see Materials and Methods) in transgenic animals carrying the *hlh-8p::sem-4* *cDNA* transgene led to the loss of both the normal SMs and all the ectopic SM-like cells (n=1037, Table 1). These findings indicate that SEM-4 functions through SEM-2 to promote the proliferating SM fate in the M lineage.

DISCUSSION

In this study, we dissected the function of the SALL factor SEM-4 in specifying the SM fate in the *C. elegans* postembryonic mesoderm. The SM cells are pluripotent precursor cells that can further divide to produce four types of non-striated egg-laying muscles in *C. elegans* hermaphrodites (Figure 1). We have shown that in the binary cell fate decision between this pluripotent SM fate and a differentiated BWM fate, SEM-4 is both necessary and sufficient to promote the pluripotent SM fate. Furthermore, SEM-4 acts through the SoxC transcription factor SEM-2 to promote the SM fate. SEM-2 is known to promote the SM fate by inhibiting the expression of two redundant BWM-specifying transcription factors, FOZI-1 and HLH-1 (Tian et al., 2011). Together, these findings led us to the following model: in the cells fated to become the SMs, SEM-4, either directly or indirectly, activates the expression of SEM-2, which inhibits the expression of FOZI-1 and HLH-1, leading to the specification of SMs (Figure 6). Notably, the partial translational SEM-4::GFP fusion is also detectable in M-lineage cells that are fated to become BWMs (Figure 2). This could be due to perdurance of the GFP protein. Alternatively, it is possible that in cells fated to become BWMs, there is either an inhibitory factor Y that prevents SEM-4 from activating *sem-2* expression, or there is a lack of a factor X that promotes *sem-2* expression (Figure 6). The identification of additional factors involved in SM fate specification will help distinguish between these possibilities. Nevertheless, our work adds another missing piece to the regulatory network involved in the diversification and fate specification of the *C. elegans* postembryonic mesoderm (Krause and Liu, 2012). Photos and colleagues (Photos et al., 2006) showed that SEM-4 is required for specifying the SM fate in the nematode *Pristionchus pacificus*, which diverged from *C. elegans* about 200–300 million years ago. Thus the role of SEM-4 in specifying the pluripotent SM fate is conserved in other nematodes as well.

SALL transcription factors are essential for normal development from humans to worms. In *C. elegans*, SEM-4 controls the development of multiple cell types, including neurons (Basson and Horvitz, 1996; Toker et al., 2003; Weinberg et al., 2013), the vulva (Eisenmann and Kim, 2000; Grant et al., 2000), as well as the postembryonic mesoderm (Basson and Horvitz, 1996, and this work). In particular, SEM-4 has been previously shown to act as a transcriptional repressor in the touch neurons in *C. elegans* (Toker et al., 2003). It has also been shown to associate with the transcriptional repressor EGL-27/MTA1 in a NODE (Nanog and Oct4-associated deacetylase)-like complex to regulate the Y-to-PDA natural reprogramming event (Jarriault et al., 2008; Kagias et al., 2012). Whether SEM-4 also functions as a repressor in the M lineage is not known. We have found that *egl-27(0)* mutants have extra SM-like cells (unpublished), a phenotype opposite to that of *sem-4(0)* mutants, suggesting that it is unlikely that SEM-4 acts together with EGL-27/MTA1 to specify the SM fate. This is consistent with the proposal that SEM-4 may exert its function differentially in different cell types, possibly through interacting with different cofactors.

The function of SEM-4 in specifying M-derived SMs and CCs appears paradoxical. On the one hand, the expression pattern and mutant phenotype of *sem-4* suggest a permissive role for SEM-4 in SM specification. In this model, SEM-4 functions to maintain those cells that are fated to become SMs in a pluripotent-like state, permitting the expression of the SM-specifying factor SEM-2, which acts to repress HLH-1 and FOZI-1 expression, thus preventing these cells from adopting the differentiated BWM fate. Such a role for SEM-4 would be consistent with the role of Sall4, one of the mammalian homologs of SEM-4, in pluripotency regulation. Sall4 is known to be essential for early mammalian development (Elling et al., 2006; Sakaki-Yumoto et al., 2006; Warren et al., 2007) and is frequently overexpressed in various cancers (Zhang et al., 2015). Recent studies have shown that Sall4 contributes to pluripotency regulation by inhibiting differentiation (Miller et al., 2016; Yamaguchi et al., 2015). However, our ectopic expression results argue against a permissive role of SEM-4 in SM specification. We have shown that overexpression of *sem-4* throughout the M lineage can convert most, if not all, M-derived cells (including the CCs and BWMs) into SM-like cells (Table 1 and Figure 4), suggesting that SEM-4 plays a positive role in promoting SM fate specification. Additionally, both loss of *sem-4* (in *sem-4* null mutants) and overexpression of *sem-4* resulted in the loss of M-derived CCs. However, while loss of *sem-4* caused a CC-to-BWM fate transformation, overexpression of *sem-4* led to a CC-to-SM fate transformation. These observations suggest that the binary cell fate choices among BWMs, CCs and SMs in the M lineage can be influenced by different amount of SEM-4. Identifying additional factors that SEM-4 functionally interacts with will help uncover the molecular basis underlying these binary fate decisions.

Acknowledgments

We thank Min Han, Wendy Hanna-Rose, Bob Horvitz, Michael Koelle, Yuji Kohara, Shohei Mitani for strains or plasmids, Melisa DeGroot, Zhiyu Liu and two anonymous reviewers for critical comments on the manuscript. Some strains were obtained from the *C. elegans* Genetics Center, which is funded by NIH Office of 27 Research Infrastructure Programs (P40 OD010440). This work was supported by NYSTEM contract #C028110 and National Institutes of Health R01 GM066953 to J.L.

References

- Amin NM, Hu K, Pruyne D, Terzic D, Bretscher A, Liu J. A Zn-finger/FH2-domain containing protein, FOZI-1, acts redundantly with CeMyoD to specify striated body wall muscle fates in the *Caenorhabditis elegans* postembryonic mesoderm. *Development* (Cambridge, England). 2007; 134:19–29.
- Basson M, Horvitz HR. The *Caenorhabditis elegans* gene *sem-4* controls neuronal and mesodermal cell development and encodes a zinc finger protein. *Genes & development*. 1996; 10:1953–1965. [PubMed: 8756352]
- Eisenmann DM, Kim SK. Protruding vulva mutants identify novel loci and Wnt signaling factors that function during *Caenorhabditis elegans* vulva development. *Genetics*. 2000; 156:1097–1116. [PubMed: 11063687]
- Elling U, Klasen C, Eisenberger T, Anlag K, Treier M. Murine inner cell mass-derived lineages depend on Sall4 function. *Proc Natl Acad Sci U S A*. 2006; 103:16319–16324. [PubMed: 17060609]
- Frokjaer-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, Olesen SP, Grunnet M, Jorgensen EM. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nature genetics*. 2008; 40:1375–1383. [PubMed: 18953339]

- Grant K, Hanna-Rose W, Han M. sem-4 promotes vulval cell-fate determination in *Caenorhabditis elegans* through regulation of lin-39 Hox. *Developmental biology*. 2000; 224:496–506. [PubMed: 10926783]
- Harfe BD, Branda CS, Krause M, Stern MJ, Fire A. MyoD and the specification of muscle and non-muscle fates during postembryonic development of the *elegans* mesoderm. *Development* (Cambridge, England). 1998; 125:2479–2488.
- Jarriault S, Schwab Y, Greenwald I. A *Caenorhabditis elegans* model for epithelial-neuronal transdifferentiation. *Proc Natl Acad Sci U S A*. 2008; 105:3790–3795. [PubMed: 18308937]
- Kagias K, Ahier A, Fischer N, Jarriault S. Members of the NODE (Nanog and Oct4-associated deacetylase) complex and SOX-2 promote the initiation of a natural cellular reprogramming event in vivo. *Proc Natl Acad Sci U S A*. 2012; 109:6596–6601. [PubMed: 22493276]
- Krause M, Liu J. Somatic muscle specification during embryonic and post-embryonic development in the nematode *C. elegans*. *Wiley Interdiscip Rev Dev Biol*. 2012; 1:203–214. [PubMed: 23801436]
- Miller A, Ralser M, Kloet SL, Loos R, Nishinakamura R, Bertone P, Vermeulen M, Hendrich B. Sall4 controls differentiation of pluripotent cells independently of the Nucleosome Remodelling and Deacetylation (NuRD) complex. *Development*. 2016; 143:3074–3084. [PubMed: 27471257]
- Photos A, Gutierrez A, Sommer RJ. sem-4/spalt and egl-17/FGF have a conserved role in sex myoblast specification and migration in *P. pacificus* and *C. elegans*. *Developmental biology*. 2006; 293:142–153. [PubMed: 16515780]
- Sakaki-Yumoto M, Kobayashi C, Sato A, Fujimura S, Matsumoto Y, Takasato M, Kodama T, Aburatani H, Asashima M, Yoshida N, Nishinakamura R. The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* (Cambridge, England). 2006; 133:3005–3013.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012; 9:676–682. [PubMed: 22743772]
- Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Developmental biology*. 1977; 56:110–156. [PubMed: 838129]
- Tian C, Shi H, Colledge C, Stern M, Waterston R, Liu J. The *C. elegans* SoxC protein SEM-2 opposes differentiation factors to promote a proliferative blast cell fate in the postembryonic mesoderm. *Development* (Cambridge, England). 2011; 138:1033–1043.
- Toker AS, Teng Y, Ferreira HB, Emmons SW, Chalfie M. The *Caenorhabditis elegans* spalt-like gene sem-4 restricts touch cell fate by repressing the selector Hox gene egl-5 and the effector gene mec-3. *Development* (Cambridge, England). 2003; 130:3831–3840.
- Warren M, Wang W, Spiden S, Chen-Murichie D, Tannahill D, Steel KP, Bradley A. A Sall4 mutant mouse model useful for studying the role of Sall4 in early embryonic development and organogenesis. *Genesis*. 2007; 45:51–58. [PubMed: 17216607]
- Weinberg P, Flames N, Sawa H, Garriga G, Hobert O. The SWI/SNF chromatin remodeling complex selectively affects multiple aspects of serotonergic neuron differentiation. *Genetics*. 2013; 194:189–198. [PubMed: 23457234]
- Yamaguchi YL, Tanaka SS, Kumagai M, Fujimoto Y, Terabayashi T, Matsui Y, Nishinakamura R. Sall4 is essential for mouse primordial germ cell specification by suppressing somatic cell program genes. *Stem Cells*. 2015; 33:289–300. [PubMed: 25263278]
- Zhang X, Yuan X, Zhu W, Qian H, Xu W. SALL4: an emerging cancer biomarker and target. *Cancer Lett*. 2015; 357:55–62. [PubMed: 25444934]

HIGHLIGHTS

1. SEM-4 is necessary and sufficient to specify the multipotent sex myoblast (SM) fate.
2. *sem-4* functions autonomously to specify the SM fate.
3. SEM-4 functions through the SoxC transcription factor SEM-2 to specify the SM fate.

Figure 1. SEM-4 is required for specifying M lineage-derived coelomocytes (CCs) and sex myoblasts (SMs)

(A, B) Diagrams showing the postembryonic M lineage in wild-type (A) and *sem-4(0)* mutant (B) animals. a, anterior; p, posterior; d, dorsal; v, ventral; l, left; r, right. (C–D) Composite GFP/DIC images of *sem-4(h769)/+* (C) and *sem-4(h769)* (D) worms with their CCs labeled with a *secreted CC::GFP* and type I vulval muscles (VM1s) labeled with *egl-15::GFP*. The *sem-4(h769)/+* worm in panel C expresses *myo-2p::GFP* from the “+” balancer chromosome *hT2[qIs48]*, which is absent in *sem-4(h769)* homozygous worms. * points to the vulva region, and arrows point to M lineage-derived CCs. (E) Diagram of the

SEM-4 protein structure marked with locations of the different *sem-4* mutations. ZF: zinc finger motif. (F) Summary of the M lineage phenotypes of various *sem-4* alleles. (G) Alignment between ZF2-4 and ZF5-7 in the SEM-4 protein, showing high degree of similarity between ZF2-3 and ZF5-6.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2. SEM-4 is expressed in all the undifferentiated cells in the M lineage
 (A–I) *sem-4::GFP* (A, D, G) expression in the undifferentiated M lineage cells marked by *hlh-8::NLS::mCherry::lacZ* (B, E, H) at the 1-M stage (A–C), 8-M stage (D–F) and the 16-SM stage (G–I). C, F, I show the corresponding merged images. Images are shown with anterior to the left and dorsal up. Only images of the left side of the animal are shown in panels D–F (showing 4 M-derived cells) and in panels G–I (showing 8 SM-derived cells). Expression is also seen in the equivalent cells on the right side (not shown). Open arrowheads in panels A, D and G point to M lineage cells expressing *sem-4::GFP*. Arrows in

panel I point to the *sem-4::GFP*-expressing BWMs derived from the M lineage, which also express the *hlh-8::NLS::mCherry::lacZ* reporter due to its perdurance. (J–O) *sem-4::GFP* expression (J, M) is not detected in the VM1s labeled by *egl-15::mRFP* (K) or M-derived CCs labeled by an *intrinsic CC::mRFP* (N). L and O are the corresponding merged images. (P) Summary of *sem-4::GFP* expression in the M lineage. *sem-4::GFP*-expressing cells are in green.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. *sem-4* expression in the M lineage requires multiple cis-regulatory elements
(A) Diagrams of tested *sem-4::GFP* reporter constructs and their M lineage expression patterns. 18-M/2-SM refers to the stage when 18 M lineage descendants are born and the 2 SMs are migrating. At least two independent transgenic lines were examined for each construct. (B–J) Examples of *sem-4::GFP* reporter expression in the M lineage using transgenic animals carrying pMK332 as an example. Images are shown with anterior to the left and dorsal up. Only images of the left side of the animal are shown. Arrowhead in panel

E points to the migrating SM cell that expresses the *sem-4::GFP* reporter. Asterisk denotes the vulva.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4. Overexpression of *sem-4* is sufficient to convert M-derived CCs and BWMs to SM-like cells

(A–B) Wild-type (A) or transgenic animals carrying *hlh-8p::sem-4 cDNA* that over-expresses *sem-4* in the M lineage (B) expressing the VM1 marker *egl-15::GFP* and the *CC::GFP* marker. Note the extra VM1-like cells and the loss of M-derived CCs in the *sem-4* over-expressing worm. (C–D) Wild-type (C) or *sem-4* over-expressing (D) animals carrying the *myo-3p::NLS::mRFP::lacZ* that marks the BWMs. The *sem-4* over-expressing worm, marked by the *myo-2p::mRFP* transgene, has fewer BWMs in the region posterior to the vulva, which is marked by an asterisk.

Figure 5. SEM-4 is required for *sem-2* expression in the M lineage

(A–C) A *sem-2(n1343)* mutant carrying *sem-4::GFP*, showing GFP expression in the M lineage (marked by *hhh-8p::NLS::mCherry*) at the 18-M stage. (D–K) GFP::SEM-2 expression in wild-type (D–G) and in *sem-4(h769)* (H–K) animals. Notice the loss of GFP::SEM-2 expression in the SM mother cell (M.vlpa, arrowhead) in *sem-4(h769)* mutant. Images are shown with anterior to the left and dorsal up. Only images of the left side of the animal are shown in panels A–F and H–J. (G, K) Summary of GFP::SEM-2 expression in the M lineage of wild-type (G) and *sem-4(h769)* mutant (K). GFP::SEM-2-expressing cells are in green.

Figure 6. A proposed model for SEM-4 and SEM-2 function in SM fate specification
Solid lines do not distinguish between direct and indirect regulatory relationships. Dashed lines represent hypothetical scenarios. Gray text indicates a lack of expression in the indicated cell.

Table 1

SEM-4 functions through SEM-2 to specify the SM fate.

Genotype	% with SMs				% with 0 SMs	# of worms scored
	1-2 M-CCs	0 M-CCs	1-2 M-CCs	0 M-CCs		
Wild-type	100%	0	0	0	0	133
<i>sem-4(h769)</i>	0	0	0	0	100%	154
<i>sem-4(h769); jfEx3385[hhlh-8p::sem-4 cDNA]*</i>	12%	0	0	11%	0	102
<i>sem-4(h769); jfEx4690[hhlh-8p::sem-2 cDNA]*</i>	0	14%	0	0	0	56
<i>N2; jfIs4191[hhlh-8p::sem-4 cDNA]</i>	10%	0	0	88%	1%	130
<i>sem-2(RNAi)</i>	1%	0	0	0	99%	1167
<i>sem-2(RNAi); jfIs4191[hhlh-8p::sem-4 cDNA]</i>	1%	0	0	0	99%	1037

The SMs and their descendants were scored by using a combination of *hhlh-8p::gfp* (labeling the SMs) and *egl-15p::gfp* (labeling the VM1s). The CCs were scored using either the *secreted CC::gfp(arls39)* or the *intrinsic CC::gfp(ccIs4438)*.

* Two transgenic lines were examined. Data from one transgenic line are shown.

Table 2

SEM-4 may partially act through the *sem-2* M lineage enhancer to regulate the M lineage expression of *sem-2*.

Genotype	% of transgenic animals expressing pCXT33* in the M lineage	
	Transgenic line 1	Transgenic line 2
<i>sem-4(h769)/hT2[qIs48]</i>	37% (N=267)	32% (N=508)
<i>sem-4(h769)</i>	23% (N=137)	21% (N=197)

* pCXT33(598bp *sem-2* M lineage enhancer::pes-10p::NLS::gfp::lacZ::unc-54 3'UTR) plasmid was injected into *sem-4(h769)/hT2[qIs48]* worms. Transgenic *sem-4(h769)/hT2[qIs48]* and *sem-4(h769)* siblings from stable transgenic lines were scored for M lineage expression of the transgene.