Development 138, 1033-1043 (2011) doi:10.1242/dev.062240

The *C. elegans* SoxC protein SEM-2 opposes differentiation factors to promote a proliferative blast cell fate in the postembryonic mesoderm

Chenxi Tian[1], Herong Shi[1], Clark Colledge[2], Michael Stern[3], Robert Waterston[2] and Jun Liu[1,*]

SUMMARY

The proper development of multicellular organisms requires precise regulation and coordination of cell fate specification, cell proliferation and differentiation. Abnormal regulation and coordination of these processes could lead to disease, including cancer. We have examined the function of the sole *C. elegans* SoxC protein, SEM-2, in the M lineage, which produces the postembryonic mesoderm. We found that SEM-2/SoxC is both necessary and sufficient to promote a proliferating blast cell fate, the sex myoblast fate, over a differentiated striated bodywall muscle fate. A number of factors control the specific expression of *sem-2* in the sex myoblast precursors and their descendants. This includes direct control of *sem-2* expression by a Hox-PBC complex. The crucial nature of the HOX/PBC factors in directly enhancing expression of this proliferative factor in the *C. elegans* M lineage suggests a possible more general link between Hox-PBC factors and SoxC proteins in regulating cell proliferation.

KEY WORDS: Mesoderm, M lineage, Sex myoblast, Bodywall muscle, SoxC, SEM-2, Hox, PBC, Proliferation, Differentiation

INTRODUCTION

The group C Sox proteins are Sry-related HMG box (Sox)-containing transcription factors (Wegner, 2010). Vertebrates contain three highly conserved SoxC genes, Sox4, Sox11 and Sox12, which play important roles in development, including cell differentiation, proliferation and survival (Penzo-Mendez, 2009). Increasing evidence has also shown that many tumor types in humans are associated with significantly elevated level of SoxC gene expression, suggesting that mis-regulation of the SoxC genes may contribute to tumor formation (Penzo-Mendez, 2009; Moreno, 2010). However, the molecular mechanisms underlying these multiple functions of the SoxC genes are not fully understood, nor are the molecular events that regulate the expression of SoxC genes.

The nematode *C. elegans* contains a single SoxC gene, C32E12.5 (Phochanukul and Russell, 2010), providing an opportunity to determine the functions of SoxC genes at single cell resolution. In this study, we show that C32E12.5 is the *sem-2* gene that is widely expressed and essential for *C. elegans* embryonic development. Furthermore, *sem-2* plays a crucial role in a binary fate decision in the postembryonic mesoderm, the M lineage. The M mesoblast is an embryonically born pluripotent precursor cell (Fig. 1A). During hermaphrodite larval development, the M cell first divides dorsoventrally, then left-right and then twice anterioposteriorly to produce 16 cells (Sulston and Horvitz, 1977). Two of them, M.vlpa and M.vrpa, divide one more time. The anterior cells from these final divisions become sex myoblasts (SMs), the precursors to all the non-striated egg-laying vulval and uterine muscles. The two posterior cells differentiate into striated body wall muscles (BWMs). The SMs then migrate to the future vulval region and further proliferate to produce eight vulval muscles (vms) and eight uterine muscles (ums). We show here that *sem-2* is required for the binary fate decision between the SMs and BWMs, and is both necessary and sufficient to promote the proliferative SM fate as opposed to the differentiated BWM fate. This specific function and expression of *sem-2* in the M lineage is under the direct regulation of Hox/PBC proteins, MAB-5, LIN-39 and CEH-20. This finding is intriguing in light of the oncogenic roles of Hox and PBC factors (Shah and Sukumar, 2010), suggesting the possibility of Hox regulation of SoxC genes during tumorigenesis.

MATERIALS AND METHODS

C. elegans strains

The following strains were used in this study. LG I: *sem-2(n1343); sem-2(ok2422)/hT2[qIs48]; sys-1(q544)/hT2[qIs48]*. LG II: *hlh-1(cc561ts)*. LG III: *fozi-1(cc609); mab-5(e1239) lin-39(n1760)/hT2[qIs48]; ceh-20(os39)/hT2[qIs48]; unc-32(e189) lin-12(n676n930ts)*. LG X: *sma-9(cc604)*. Analyses were performed at 20°C, unless otherwise noted.

Integrated lines were as follows. LW0081: *ccIs4438 (intrinsic CC::gfp) III; ayIs2 (egl-15::gfp) IV; ayIs6 (hlh-8::gfp) X*. LW1066: *jjIs1066[pJKL705.1(hlh-8p::mRFP+unc-119(+)]?; unc-119(ed4) III*. LW1639: *mab-5(e1239) lin-39(n1760) III/hT2[qIs48] (I;III); ayIs6(hlh-8p::gfp) X*. LW2466: *jjIs1475(myo-3::rfp) I; ccIs4438 (intrinsic CC::gfp) III; ayIs2 (egl-15::gfp) IV; ayIs6 (hlh-8::gfp) X*.

Isolation, genetic and molecular analysis of *sem-2* alleles

sem-2(n1343) was identified in a Tc1 mutagenesis screen (Desai et al., 1988). It was mapped using three factor crosses into a small region between *unc-63* and *spe-11*, 0.2 map units to the right of *unc-63*. Cosmids spanning that region were tested for rescuing activity. Two cosmids, F47D3 and C32E12, both rescued the egg-laying defect of *n1343*. By further truncating the overlapping region of the cosmids, the rescuing activity was mapped to C32E12.5. To identify the molecular lesion of *n1343*, PCR reactions were performed using primers designed to span the genomic region of C32E12.5. Primer pairs JKL-620 and JKL-621 amplified a fragment bigger than expected. Further PCR and sequencing identified a Tc1 insertion in the first intron of C32E12.5 (between –9186 and –9185, Fig. 2A).

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA. [2]Department of Genome Sciences, University of Washington School of Medicine, 1705 NE Pacific Street, HSC, Rm357, Health Sciences K-357B, Box 357730, Seattle, WA 98195-7730, USA. [3]Department of Biology, University of Central Florida, 4000 Central Florida Boulevard, Orlando, FL 32816-2368, USA.

*Author for correspondence (jl53@cornell.edu)

Accepted 30 December 2010

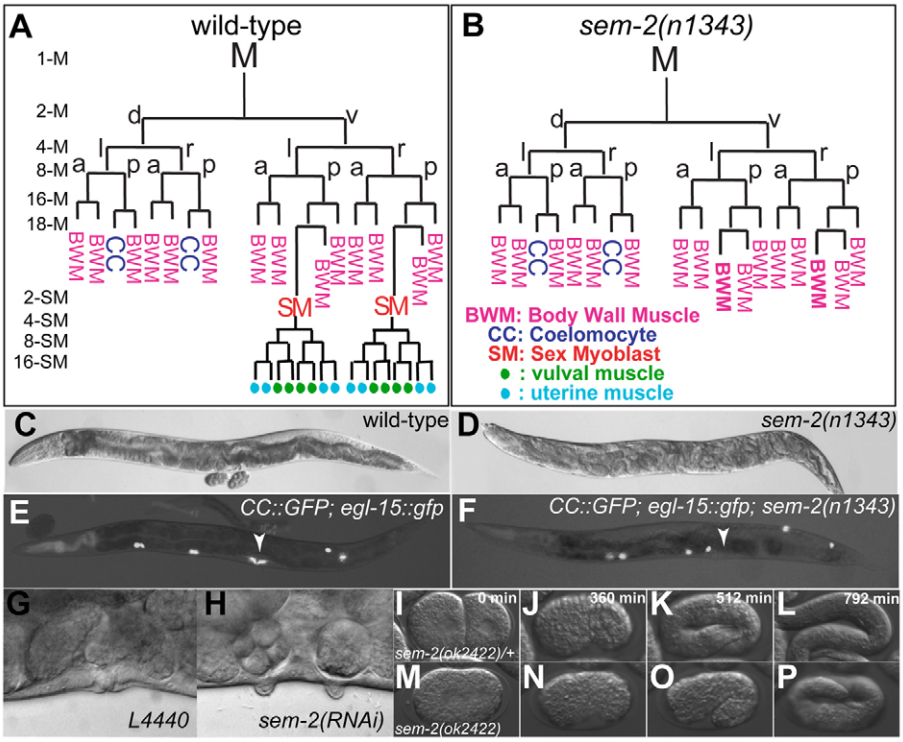

Fig. 1. *sem-2* is required for proper embryonic and postembryonic development. (**A**) The wild-type postembryonic M lineage showing all differentiated cell types that arise from M. (**B**) The M lineage of *sem-2(n1343)* mutants. (**C**,**D**) DIC images of a wild-type (C) and a *sem-2(n1343)* (D) adult hermaphrodites. All images presented hereafter are oriented with dorsal side upwards and anterior towards the left, unless noted otherwise. (**E**,**F**) A wild-type (E) and a *sem-2(n1343)* animal (F) with *CC::gfp* and *egl-15::gfp*. Arrowheads indicate the vulva. Both wild-type and *n1343* animals have four embryonically derived and two M-derived CCs. (**G**,**H**) Vulval phenotypes of *sem-2(RNAi)* animals (H) compared with *L4440* control RNAi animals (G). (**I-P**) Frames from time-lapse movies of *sem-2(ok2422)*/+ (I-L) and *sem-2(ok2422)* (M-P) worms. Time 0 is when the movie was started. Four time points were selected to represent different embryonic stages. M, mesoblast; d, dorsal; v, ventral; a, anterior; p, posterior.

Plasmid constructs and transgenic lines

Fosmid WRM0623cE02 (Geneservice) was used to generate the *gfp::sem-2* translational fusion construct via recombineering (Warming et al., 2005). *gfp* sequences from the Fire lab vector pPD95.75 were inserted immediately after the ATG start codon. Recombineering was also used to insert Tc1 into the *gfp::sem-2*-containing fosmid generated above between –9186 and –9185. The insertions were verified by PCR and sequencing.

sem-2 cDNA from *yk657g12* was used to generate *pJKL776 (hlh-8p::sem-2 cDNA::unc-54* 3′UTR).

A 4554 bp fragment of the *sem-2* 1st intron (–11650 to –7097) was cloned into L3135 of the Fire lab vector kit to generate pCXT12, which was subsequently used to generate deletion constructs pCXT18-22 and pCXT26-33. pCXT33 was used to generate pCXT97-99 and pCXT173, carrying mutations in the PBC/Hox-binding site. *sys-1*(T23D8.9) and *let-381*(F26B1.7) RNAi constructs were obtained from the Ahringer RNAi library provided by Geneservice (Kamath et al., 2003). *sem-2(RNAi)* construct pCXT9 was made by subcloning the *sem-2* cDNA fragment from *yk657g12* into L4440 (Timmons and Fire, 1998). pNMA49 and pNMA50 were used for knocking down *fozi-1* and *mab-5*, respectively (Amin et al., 2009).

Transgenic lines were generated using the plasmids pRF4 (Mello et al., 1991), pJKL449 (*myo-2p::gfp::unc-54 3′ UTR*) (Jiang et al., 2009) or LiuFD61 (*mec-7p::mRFP)* (Amin et al., 2009) as markers.

RNAi

The T7 Ribomax RNA Production System (Promega) was used to generate *sem-2* dsRNA using *yk404e6* as a template. Synchronized L1 animals expressing various M lineage GFP markers were soaked in the dsRNA solution at 20°C for 24-48 hours following the protocol of Maeda and colleagues (Maeda et al., 2001). Animals were allowed to recover at 20°C and adult worms were scored for M lineage phenotypes. Water was used as a soaking control. For *fozi-1(RNAi)* (Amin et al., 2007), *let-381(RNAi)* (Amin et al., 2010), *mab-5(RNAi)* and *sys-1(RNAi)* (Amin et al., 2009) synchronized L1 animals expressing different M lineage markers were plated on HT115(DE3) bacteria expressing dsRNA for the gene of interest. Bacteria for ingestion were prepared as described by Kamath and Ahringer (Kamath and Ahringer, 2003). RNAi by ingestion was performed at 25°C and animals were scored for M lineage phenotypes 24-48 hours after plating.

Immunofluorescence staining

Animal fixation, immunostaining, microscopy and image analysis were performed as described previously (Amin et al., 2007). Guinea pig anti-FOZI-1 (Amin et al., 2007) (1:200) and goat anti-GFP (Rockland Immunochemicals; 1:1000) were used.

Electrophoretic mobility shift assays (EMSA)

6xHis-tagged LIN-39 and CEH-20 fusion proteins were purified as previously described (Liu and Fire, 2000). Complementary single-stranded DNA oligos were 3′-end labeled with biotin using the Biotin 3′ End DNA Labeling Kit (Pierce) and annealed at room temperature for 1 hour. Gel shift reactions and detection were performed using the LightShift Chemiluminescent EMSA Kit (Pierce). Oligonucleotides used are: wt, CXT216/217; canonical, CXT218/219; mut 1, CXT220/221; mut 2, CXT222/223; mut 3, CXT224/225.

RESULTS

sem-2(n1343) mutants exhibit a fate transformation of the proliferating SMs to differentiated BWMs

sem-2(n1343) was identified in a Tc1 insertion mutagenesis screen for egg-laying defective (Egl) mutants in the *mut-2(r459)* background (Desai et al., 1988). Hermaphrodites homozygous for *n1343* are 100% Egl (*n*>200), lack all vulval and uterine muscles required for egg laying, as monitored by DIC and polarized light microscopy, and fail to express egg-laying muscle specific reporters, such as *egl-15::gfp* and *ceh-24::gfp* [expressed in type I vulval muscles, VM1s (Harfe et al., 1998b), or in all the egg-laying muscles (Harfe and Fire, 1998)] (Fig. 1C-F, data not shown). To determine the basis for the missing egg-laying muscles, we followed the M lineage in *n1343* hermaphrodites using both Normaski microscopy and the *hlh-8::gfp* reporter that marks all undifferentiated cells of the M lineage (Harfe et al., 1998b). In *n1343* hermaphrodites, the SM mother cells, M.v(l/r)pa as in wild type give rise to two daughter cells, M.v(l/r)paa and M.v(l/r)pap. However, the anterior daughters, normally the SMs, exhibit the

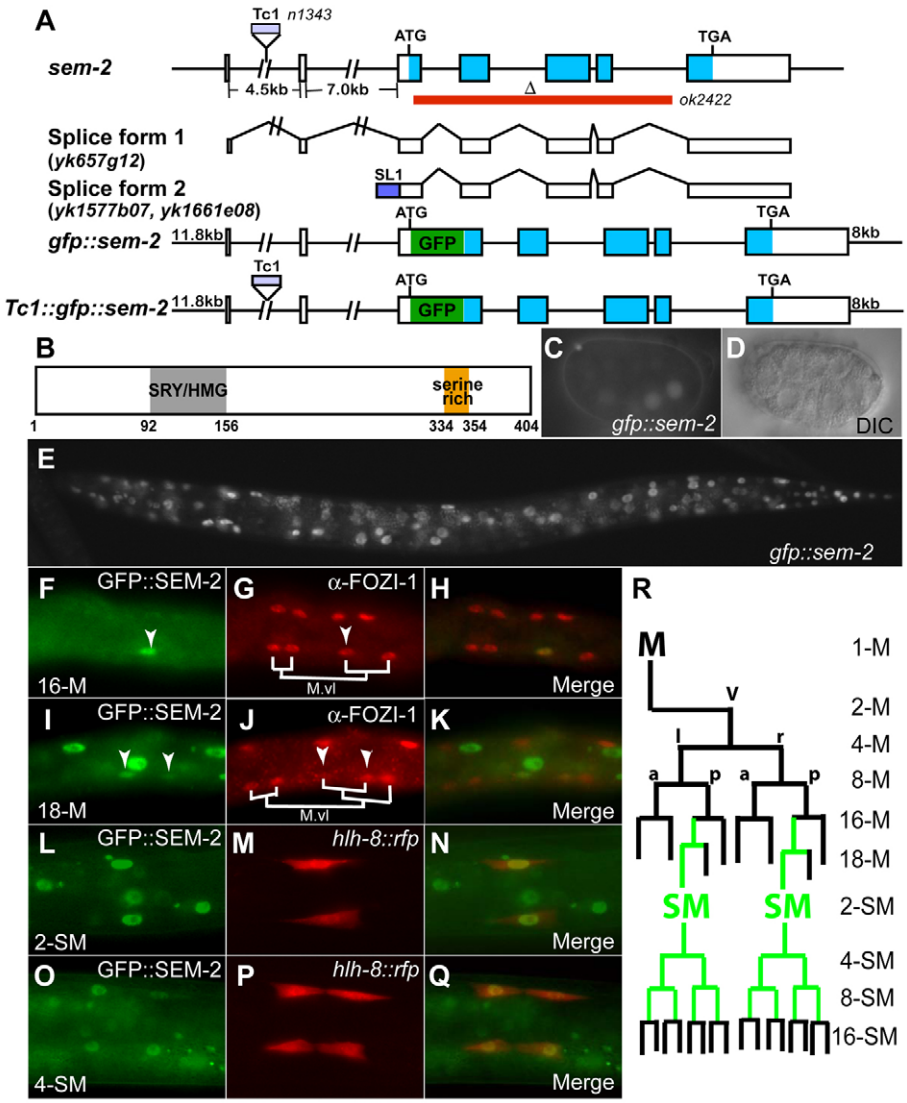

Fig. 2. *sem-2* gene structure and expression pattern. (**A**) *sem-2* gene structure (not drawn to scale) showing the molecular lesions of the *sem-2(n1343)* and *sem-2(ok2422)* mutations, two splice forms identified by cDNAs, and two translational reporters: *gfp::sem-2* and *Tc1::gfp::sem-2*. Exons are in boxes. The coding region is in blue. (**B**) A schematic of the predicted SEM-2 protein. SEM-2 contains a SRY/HMG box and a serine-rich region. (**C**,**D**) *gfp::sem-2* expression [C, and the corresponding DIC image (D)] is first detectable in embryos at the beginning of gastrulation. (**E**) Expression of *gfp::sem-2* in a transgenic larva. (**F-Q**) Expression of *gfp::sem-2* in the M lineage. (F-K) Double labeling of wild-type animals with anti-GFP antibodies (F,I) and anti-FOZI-1 antibodies (G,J), and the corresponding merged images (H,K) at the 16-M (F-H) and 18-M (I-K) stages, showing the expression of *gfp::sem-2* in the SM mother cell M.vlpa (arrowheads in F,G), the SM cell M.vlpaa and its sister cell M.vlpap (arrowheads in I-J). Expression is also seen in the equivalent cells on the right-hand side (not shown). (L-Q) Ventral views of wild-type animals carrying *gfp::sem-2* (L,O) and *hlh-8::rfp* (M,P) and the corresponding merged images (N,Q) at the 2-SM (L-N) and 4-SM (O-Q) stages. (**R**) Summary of GFP::SEM-2 expression in the M lineage. GFP::SEM-2-expressing cells are in green.

morphology and identity of their sister cells, the body wall muscles (BWM; Fig. 1A,B and data not shown). These cells also fail to migrate anteriorly, never divide and never differentiate into the egg-laying muscles. Therefore, *n1343* hermaphrodites exhibit a SM to BWM fate transformation. The M-lineage phenotype of *n1343* animals may reflect a complete removal of *sem-2* function in the M lineage because *n1343* over a deficiency of the *sem-2* region, *sDf4*, yielded the same M-lineage phenotype as homozygous *n1343* animals (data not shown). The effect of the *sem-2(n1343)* mutation on SM development appears to be sex specific, as the SMs appear unaffected in *n1343* males, and *n1343* males mate efficiently (data not shown).

sem-2 encodes the sole *C. elegans* group C HMG/SRY box-containing protein

We mapped *sem-2(n1343)* to C32E12.5 based on three-factor mapping and cosmid rescue (see Materials and methods). PCR and sequencing analyses showed that *n1343* animals contain a Tc1 transposon insertion located 9185 bp upstream of the predicted ATG of C32E12.5 (Fig. 2A). RNAi of C32E12.5 in wild-type worms by soaking L1s caused 98% (n=500) of worms to be Egl and missing *egl-15::gfp* expression. RNAi-treated worms in the same experiment also exhibited a multivulvae phenotype, with 48% being bi-vulvae and 5% having three vulvae (n=81; Fig. 1G,H and data not shown).

The *n1343* mutation does not appear to be a null allele of *sem-2*, as a deletion allele of *sem-2* generated by the *C. elegans* knockout consortium, *ok2422*, resulted in 100% embryonic lethality (n>100). *ok2422* has most of the C32E12.5-coding sequences deleted and is probably a null allele (Fig. 2A). Time-lapse video microscopy showed that *ok2422* homozygous embryos developed normally until late-comma stage (n=5, Fig. 1I-P). However, these embryos were severely delayed in the elongation process, resulting in noticeable morphological defects, such as enlarged heads relative to control embryos (Fig. 1O). They eventually all reached to, and were arrested at, the threefold stage (Fig. 1L,P; see Movie 1 in the supplementary material). A fosmid (WRM0623cE02) containing C32E12.5 rescued the embryonic lethality of *ok2422*. RNAi soaking of L4s also resulted in 100% embryonic lethality (n>200). Thus, C32E12.5 is an essential gene required in embryogenesis, vulval and M lineage development, and *n1343* is a partial loss-of-function allele of C32E12.5 that specifically affects M-lineage development.

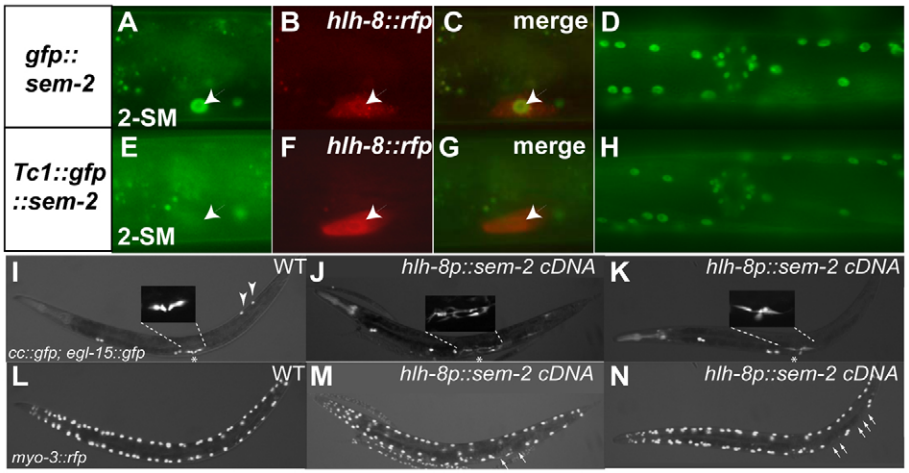

Fig. 3. *sem-2* is both required and sufficient in the M lineage for specifying the SM fate. (**A-H**) Transgenic worms carrying *gfp::sem-2* (A-D) or *Tc1::gfp::sem-2* (E-H) at the 2-SM stage. *gfp::sem-2* (A-C), but not *Tc1::gfp::sem-2* (E-G), is expressed in the SM cells (marked by arrowheads, one focal plane shown) labeled by *hlh-8::rfp* (B,F, with merged images shown in C,G). Both transgenes are expressed in vulval and hypodermal cells (D,H, ventral views). (**I-N**) Wild-type (I,L) or transgenic worms carrying *hlh-8p::sem-2 cDNA* (J,K,M,N) that also express: *cc::gfp*, which labels CC; *egl-15::gfp*, which labels type I vulval muscles (I-K); and *myo-3::gfp*, which labels BWMs (L-N). Compared with wild-type animals, animals expressing *hlh-8p::sem-2* cDNA often have extra type I vulval muscles that are disorganized (enlargement boxes in J,K), and lack M-derived CCs (J,K) and some M-derived BWMs (arrows in M,N indicate regions with missing M-derived BWMs). Arrowheads in I indicate M-derived CCs. Asterisks indicate the location of the vulva.

Sequencing of the cDNA clones for C32E12.5 showed that the C32E12.5 locus is alternatively spliced with two different splice isoforms. Splice form 1 is represented by the cDNA clone *yk657g12*, in which two small exons separated by two large introns; 4.5 kb (containing the Tc1 insertion in *n1343* animals) and 7 kb, respectively, are located upstream of the ATG-containing exon (Fig. 2A). Splice form 2 is represented by two independent cDNA clones *yk1577b07* and *yk1661e08*, which have a SL1 splicing leader sequence trans-spliced to the ATG-containing exon (Fig. 2A). Both splice isoforms are predicted to encode the same open reading frame that contains 404 amino acids. The predicted SEM-2 protein contains a DNA-binding domain, the SRY/HMG box (residue 92-156), and a C-terminal serine-rich region (residue 334-354) that is predicted to be the transcriptional activation domain (Fig. 2B). Based on the homology in the SRY/HMG region, SEM-2 is most similar to group C Sox proteins, including Sox4, Sox11 and Sox12 in vertebrates and Sox14 in *Drosophila* (Bowles et al., 2000).

SEM-2/SoxC is a nuclear protein expressed in the SM precursors and their descendants

To understand how *sem-2* functions, we generated a N-terminal *gfp::sem-2* translational fusion and examined its expression pattern. This *gfp::sem-2* translational fusion is functional because it rescued the Egl phenotype of *n1343* (1/1 line) and the embryonic lethality of *ok2422* (2/2 lines). The GFP::SEM-2 protein was nuclear localized, consistent with its predicted role as a transcription factor (Fig. 2C,E,F,I,L,O).

Expression of *gfp::sem-2* was first detectable in a subset of cells of the E and MS lineages in early gastrulating-stage embryos (Fig. 2C,D). A similar expression pattern has been reported by Broitman-Maduro et al. (Broitman-Maduro et al., 2005) using a transcriptional reporter of *sem-2*. The *gfp::sem-2* expression persisted through embryonic and larval development in many cell types, including vulval, hypodermal and intestinal cells (Fig. 2E, Fig. 3D).

To determine the expression pattern of *gfp::sem-2* in the M lineage, we also labeled the M lineage cells with *hlh-8p::rfp* or anti-FOZI-1 immunostaining (Amin et al., 2007). *gfp::sem-2* expression in the M lineage was first detectable at the 16-M stage in the SM mother cells, M.v(l/r)pa (Fig. 2F-H), and remained in both of their daughter cells, M.v(l/r)paa and M.v(l/r)pap (Fig. 2F-H). The presence of GFP::SEM-2 in M.v(l/r)pap was transient: GFP::SEM-2 was not detectable after M.v(l/r)pap differentiated into BWMs. However, GFP::SEM-2 persisted in the nuclei of the SM cells and all their descendants until the 8-SM stage, and became undetectable at the 16-SM stage (Fig. 2L-Q; data not shown). The expression pattern of GFP::SEM-2 is summarized in Fig. 2R. Thus, *sem-2* expression is turned on in the SM mother cells and is retained in the SMs and their proliferating descendants.

M lineage expression of *sem-2* is specifically disrupted in *n1343* mutants

As *sem-2* is an essential gene, the exclusive M lineage defects observed in *sem-2(n1343)* mutants suggest that the Tc1 transposon insertion may specifically affect *sem-2* expression in the M lineage. To test this hypothesis, we introduced the Tc1 transposon back into the functional *gfp::sem-2* translational fusion, at the same insertion site as found in *n1343* mutants (Fig. 2A), and examined the function and expression of this reporter: *Tc1::gfp::sem-2*. *Tc1::gfp::sem-2* rescued the embryonic lethality of *ok2422* mutants (1/1 line). However, it failed to rescue the Egl phenotype of *n1343* mutants (4/4 lines, n>100). Thus, *Tc1::gfp::sem-2* is functional outside of the M lineage, but not functional in the M lineage. When we examined its expression pattern, we found that *Tc1::gfp::sem-2* was not expressed in the M lineage at all (Fig. 3A-C,E-G; data not shown), whereas its expression outside of the M lineage was largely unaffected (Fig. 3D,H). We also forced the expression of *sem-2* cDNA in the M lineage of *n1343* animals using the M lineage-specific *hlh-8* promoter and found that it rescued the missing egg-laying muscle phenotype in *sem-2(n1343)* animals, as determined by the reappearance of *egl-15::gfp* expression (6/6

lines). Together these observations demonstrate that the Tc1 insertion located in the 4.5 kb intron specifically disrupts the M lineage expression of *sem-2*, and that *sem-2* is required within the M lineage for proper SM fate specification.

SEM-2/SoxC is sufficient to promote the SM fate

We then tested whether *sem-2* is sufficient to promote non-SM cells to adopt the SM fate. To this end, we used the *hlh-8* promoter to force *sem-2* cDNA expression in all the undifferentiated M lineage cells in wild-type animals. We first assayed for the effect of *sem-2* misexpression on the CC and BWM fates using an intrinsic CC-specific reporter *CC::gfp* and a BWM-specific reporter *myo-3p::rfp*. Wild-type worms have four embryonically derived and two M lineage-derived CCs (Fig. 3I). Among animals carrying the *hlh-8p::sem-2* construct, 62.5% had no M-derived CCs (Fig. 3J,K) and 14.1% had only one M-derived CC (n=64). The animals that lack M-derived CCs also lacked, on average, 10 out of the 14 BWMs derived from the M lineage (n=9, Fig. 3M,N). By contrast, 40.6% of the animals carrying *hlh-8p::sem-2* (n=64) had extra *hlh-8::gfp*-expressing SM-like cells and, later, extra *egl-15::gfp*-expressing type I vulval muscle-like cells (Fig. 3J). In most cases the vulval muscles born from animals misexpressing *sem-2* were not attached to the vulva properly or not even located in the vulval region, and were therefore not likely to be functional (Fig. 3J,K). Taken together, the loss of M-derived BWM and CCs and the appearance of excessive SM- and vulval muscle-like cells are strongly suggestive of a transformation of some (if not all) of the M lineage cell types to SMs. Thus, *sem-2* is not only necessary, but also sufficient, to promote the SM fate.

The M lineage-specific expression of *sem-2*/*SoxC* is controlled by elements in the 4.5 kb intron

The M lineage-specific function of *sem-2* in promoting the SM fate coupled with the presence of the Tc1 insertion in the 4.5 kb intron and the loss of M lineage-specific expression of *Tc1::gfp::sem-2* suggest that the 4.5 kb intron may contain element(s) specifically required for *sem-2* expression in the M lineage. To test this hypothesis, we first placed the 4.5 kb intron (Fig. 2A) directly upstream of the *gfp*-coding sequence, but observed no *gfp* expression in transgenic lines carrying the construct (two lines, n>100). However, the 7 kb intron of *sem-2* alone was capable of driving reporter expression in hypodermal, intestinal and vulval cells that express *sem-2* even in the *n1343* mutants (data not shown). Taking into account the two types of *sem-2* transcripts (Fig. 2A), these observations suggest that the 7 kb intron probably has the promoter responsible for the transcription of splice form 2 in hypodermal, intestinal and vulval cells, whereas the 4.5 kb intron has no promoter activity on its own.

To further test whether the 4.5 kb intron contains any M lineage-specific enhancer(s), we placed the entire 4.5 kb intron sequence upstream of the *Δpes-10* basal promoter and *gfp*, followed by the *unc-54* 3′UTR (pCXT12, Fig. 4A). We observed M lineage-specific expression of the reporter (6/8 lines, Fig. 4B-G), suggesting the presence of M lineage-specific enhancer element(s) within the 4.5 kb intron. However, the *gfp* expression pattern from this reporter differs from that of the translational *gfp::sem-2* reporter described earlier; whereas GFP::SEM-2 was only found in the SM mother and the SM lineage cells, transgenic lines carrying the pCXT12 reporter showed *gfp* expression in all the undifferentiated cells in the M lineage, including the early M lineage and the SM lineage (Fig. 4B-G). We reasoned that either additional cis-element(s) are involved in restricting *sem-2* expression to specific cells within the M lineage, or the SEM-2 protein is unstable in the early M lineage. To this end, we tested 2.8 kb sequences upstream of the 4.5 kb intron, 2 kb sequences of the 3′ UTR and the entire 7 kb second intron by placing each of them in cis with the 4.5 kb intron. None of them was able to restrict the expression of the *4.5kb intron::gfp* reporter in the early M lineage (Fig. 4H). Despite this, our data demonstrate that the 4.5 kb intron contains enhancer(s) positively involved in directing *sem-2* expression in the M lineage.

The M lineage expression of *sem-2*/*SoxC* is under the direct control of Hox factors, MAB-5 and LIN-39, and their co-factor CEH-20

To identify the M lineage-specific enhancer(s) within the 4.5 kb intron, we generated a series of deletions in pCXT12 (Fig. 4A), and found a 598 bp region (–11650 to –7097) that is sufficient to drive reporter expression in the M lineage (Fig. 4A). Alignment of this 598 bp region in *C. elegans* and its homologous sequences in three related *Caenorhabditis* species, *C. briggsae*, *C. remanei* and *C. brenneri*, which are more genetically different from *C. elegans* than mouse is from human, identified several blocks of highly conserved sequences (data not shown) that include a site TGATATATCG (Fig. 5A). This site closely matches the consensus PBC/Hox binding sequence TGATNNAT(G/T)(G/A), with TGAT being the PBC-binding site and AT(G/T)(G/A) being the Hox-binding site (Chan and Mann, 1996; Mann and Affolter, 1998). Interestingly, the Tc1 transposon insertion site in the *sem-2(n1343)* mutants disrupts the putative PBC half site (Fig. 5A). Furthermore, inserting Tc1 into the same location as in *n1343* mutants in the reporter construct pCXT33 completely blocked the M lineage expression of the reporter (pCXT173, 6/6 lines, n>100). We further tested the importance of the putative PBC/Hox-binding site by making clustered mutations in each half site and testing the consequences of the mutations on the M lineage enhancer activity. Mutating the PBC half site (mut1) completely abolished the M lineage expression of the reporter (100%, n=45, Fig. 5B). Mutations in the putative Hox-binding site (mut2 and mut3) also led to the loss of the M lineage enhancer activity in all (100%, n=26, for mut2) or most (93.7%, n=63, for mut3) of the transgenic animals (Fig. 5B). Thus, the putative PBC/Hox-binding site, disrupted by the Tc1 insertion in *n1343* mutants, is required for the *sem-2* M lineage enhancer activity.

Previous studies have found that two Hox factors, MAB-5 and LIN-39, and their co-factor the PBC homolog CEH-20, play essential roles in M lineage development (Liu and Fire, 2000; Jiang et al., 2009). Both *mab-5* and *ceh-20* are expressed throughout the M lineage (Liu and Fire, 2000; Jiang et al., 2009), whereas *lin-39* is expressed in the SM lineage (Wagmaister et al., 2006). Furthermore, MAB-5 and LIN-39 together, and CEH-20 are required for the proper specification and differentiation of M-derived CC, BWMs and SMs (Jiang et al., 2009). We therefore tested whether these three factors are required for the *sem-2* M lineage enhancer activity using the minimal pCXT33 reporter. We introduced pCXT33 into the double null mutant *lin-39(n1760) mab-5(e1239)* and a strong loss-of-function mutant *ceh-20(os39)* (Liu and Fire, 2000; Arata et al., 2006; Jiang et al., 2009). *gfp* reporter expression was detected in the M mesoblast in 98.3% (n=56) of *lin-39(n1760) mab-5(e1239)/++* animals (data not shown) and 92.6% (n=54) of *ceh-20(os39)/+* animals (Fig. 5D,E; data not shown). However, *gfp* expression was detected in the M mesoblast in only 4.9% (n=82) of *lin-39(n1760) mab-5(e1239)*

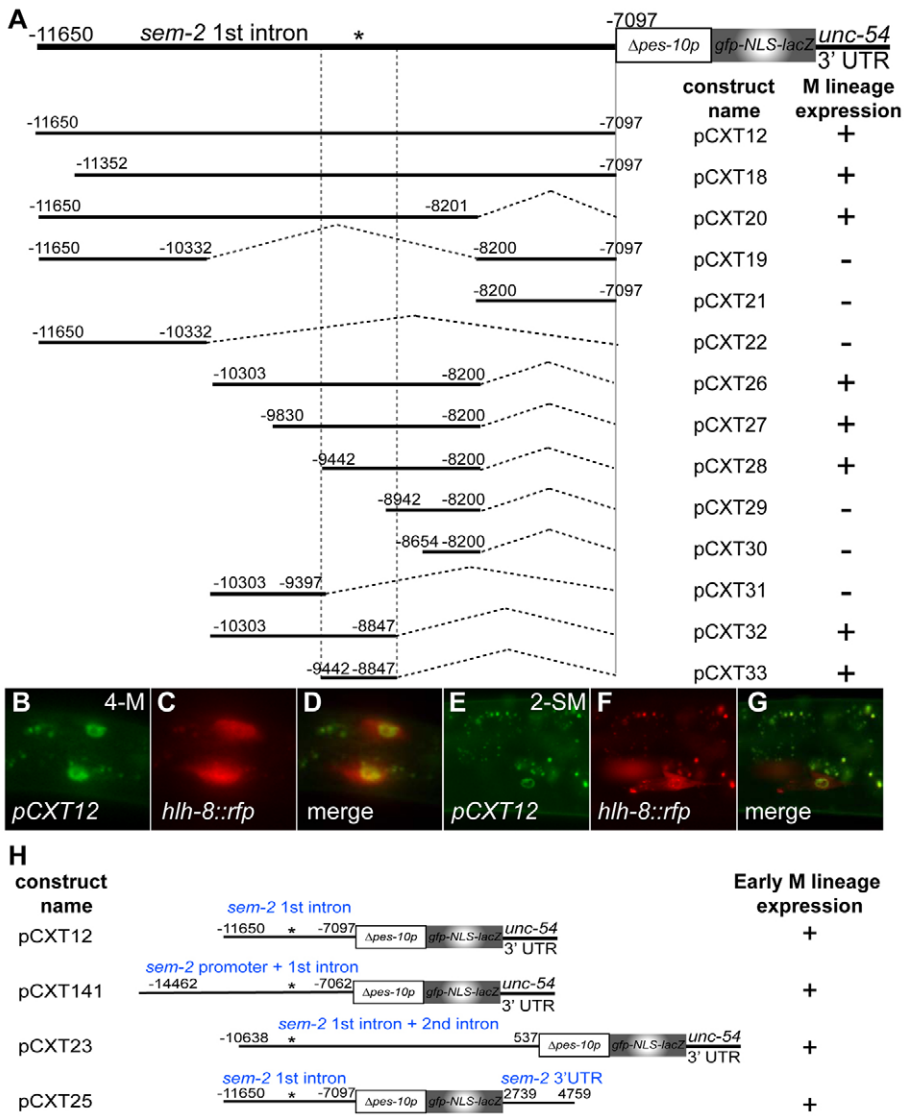

Fig. 4. The 4.5 kb intron of *sem-2* contains an M lineage enhancer element. (**A**) Schematic of deletion series of the *sem-2* 4.5 kb intron in the enhancer analysis. GFP expression was scored in F2 animals in at least three independent lines per construct. Expression is represented as + (consistent expression) or – (no detectable expression). Asterisk indicates the location of the Tc1 insertion in the *n1343* allele (also in H). (**B-G**) Examples showing lateral views (one focal plane) of worms expressing GFP driven by the pCXT12 reporter (B,E) in the M lineage (labeled by *hlh-8::rfp*, C,F) at the 4-M stage (B-D) and the 2-SM stage (E-G). (**H**) Constructs used to identify possible cis-negative elements in the *sem-2* locus. + indicates the presence of reporter expression.

animals and in 8.3% (n=72) of *ceh-20(os39)* animals (Fig. 5F,G; data not shown). Thus, the PBC/Hox factors MAB-5, LIN-39 and CEH-20 are necessary for the *sem-2* M lineage enhancer activity.

To test whether MAB-5, LIN-39 and CEH-20 may directly regulate *sem-2* expression in the M lineage, we used the electrophoretic mobility shift assay (EMSA) to test whether recombinant LIN-39 and CEH-20 proteins can cooperatively bind to the putative PBC/Hox-binding site in vitro. We were not able to generate recombinant full-length MAB-5 proteins in vitro (Liu and Fire, 2000). As shown in Fig. 5, LIN-39 alone, or together with CEH-20, binds oligonucleotides with a canonical ANTP/EXD composite site (Knoepfler et al., 1996) or oligonucleotides containing the putative PBC/Hox site in the *sem-2* M lineage enhancer (Fig. 5B,C). By contrast, mutating the PBC half site (mut1) completely abolished the composite binding of LIN-39 and CEH-20 proteins without affecting LIN-39 binding alone (Fig. 5C). Similarly, mutating the Hox half site (mut2) completely abolished the binding of LIN-39 alone or both LIN-39 and CEH-20. Consistent with the in vivo reporter assay result, mut3 significantly reduced, but did not completely abolish, the binding of LIN-39 alone or LIN-39 and CEH-20 together (Fig. 5C). The composite binding of LIN-39 and CEH-20 to the putative PBC/Hox-binding site can also be competed away using excess of unlabeled oligonucleotides containing the wild-type binding site (Fig. 5C). Taken together, our data suggest that *sem-2* is directly regulated by the Hox factors MAB-5 and LIN-39 and their co-factor CEH-20 in the M lineage. Consistent with our finding, ChIP-seq experiments by modENCODE showed that LIN-39 directly binds to this Hox-PBC site in vivo (Gerstein et al., 2010).

SEM-2/SoxC acts downstream of signaling pathways required for proper SM fate specification

Previous studies have shown that LIN-12/Notch signaling is required for promoting the SM fate on the ventral side, whereas the antagonism of TGF-β signaling by SMA-9 is required for the CC fate on the dorsal side, of the M lineage (Greenwald et al., 1983; Foehr et al., 2006; Foehr and Liu, 2008). We therefore tested whether SM fate and *sem-2* expression in the M lineage is under the control of these dorsoventral patterning mechanisms. As shown in Fig. 6A-C, no *gfp::sem-2* expression was observed in M.v(l/r)pa cells at the 16-M stage (n=44) in *unc-32(e189) lin-12(n676n930ts)* animals at the restrictive temperature, in which the SMs are transformed to CCs (Greenwald et al., 1983; Foehr and Liu, 2008).

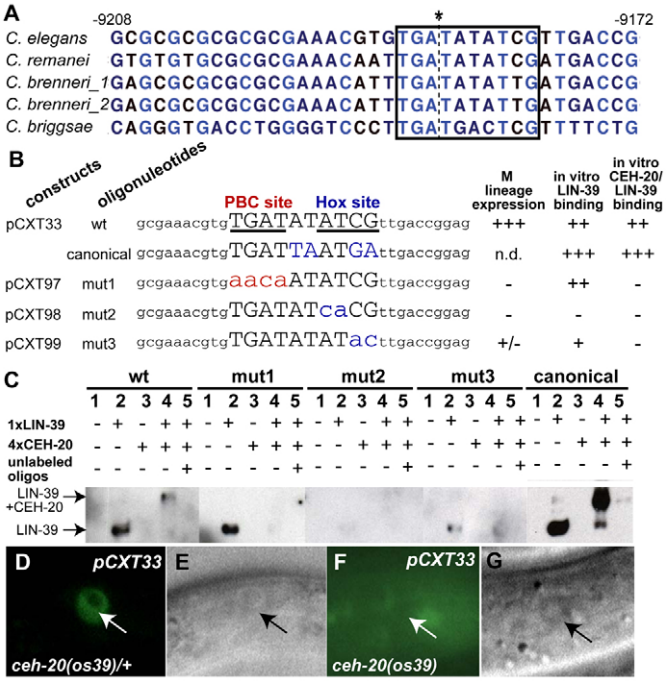

Fig. 5. *sem-2* is a direct target of Hox/PBC factors in the M lineage. (**A**) Alignment of the conserved intronic regions that include the putative Hox/PBC-binding site (box region) from *C. elegans*, *C. remanei*, *C. brenneri* and *C. briggsae* using clustalW. *C. brenneri* has two copies of the *sem-2* gene. Asterisk marks the Tc1 insertion site in *sem-2(n1343)*. Blue color indicates identical nucleotides in all four species. (**B**) Summary of in vivo transgenic reporter assays and in vitro EMSA results. The putative PBC- (red) and Hox- (blue) binding sites are in capital letters. Mutated sites are in lower case. +++, strong binding or robust reporter expression; ++, moderate binding; +/–, faint expression in four out of 63 animals examined; –, no binding or no detectable reporter expression. (**C**) EMSA using indicated oligonucleotides, purified LIN-39 and CEH-20 proteins, and unlabeled competitor wild-type oligonucleotides in 2000-fold excess. The top and bottom arrows indicate the sizes of LIN-39/CEH-20/DNA and LIN-39/DNA complexes, respectively. (**D-G**) pCXT33 expression in *ceh-20(os39)/+* (D,E) and *ceh-20(os39)* (F,G) animals. (D,F) GFP, (E,G) DIC. Arrows indicate the M mesoblast.

By contrast, *gfp::sem-2* expression was detected in both M.d(l/r)pa and M.v(l/r)pa, as well as in their descendants at the 16-M stage (*n*=18, Fig. 6D-F) and the 18-M stage (*n*=4; data not shown), respectively, in *sma-9(cc604)* mutants, which show a M-derived CC to SM fate transformation (Foehr et al., 2006). Furthermore, *sem-2* is required for specifying the dorsal SMs in *sma-9(cc604)* mutants, as both the endogenous and the ectopic SMs in *sma-9(cc604)* mutants were lost upon *sem-2(RNAi)* treatment (97.3%, *n*=75) or in *sem-2(n1343); sma-9(cc604)* double mutants (100%, *n*>200). Thus, the M lineage expression of *sem-2* is downstream of both LIN-12 and SMA-9.

We have recently shown that the FoxF/C forkhead transcription factor LET-381 also functions downstream of SMA-9 and LIN-12 to promote CC fate specification (Amin et al., 2010). *let-381* is expressed in M.d(l/r)pa and their mothers, and *let-381(RNAi)* leads to the fate transformation of M-derived CCs to the SM mother cells (Amin et al., 2010). In *let-381(RNAi)* animals, *gfp::sem-2* was detected in both M.d(l/r)pa and M.v(l/r)pa, as well as in their descendants (Fig. 6G-I, *n*=23/45). Furthermore, no SMs were produced in *sem-2(RNAi) let-381(RNAi)* animals (94.1%, *n*=118), suggesting that *sem-2* functions downstream of *let-381*. Thus, the exclusive ventral expression of *sem-2* in the M lineage is due to the negative regulation of LET-381, the expression of which is repressed by LIN-12 on the ventral side and activated by SMA-9 on the dorsal side (see Fig. 7).

Along the anterior-posterior axis, the Wnt/β-catenin asymmetry pathway is required for proper M lineage fate specification at the 16-M stage (Amin et al., 2009). Specifically, SYS-1/β-catenin is enriched in the nuclei of posterior cells, and mutations of *sys-1* lead to a posterior-to-anterior fate transformation and the production of extra SMs and CCs. POP-1/TCF, however, has a reciprocal localization and loss-of-function phenotypes (Amin et al., 2009). Because *sem-2* expression is first detected only in M.v(l/r)pa cells but not in their posterior sister cells M.v(l/r)pp at the 16-M stage, we looked to see whether *sem-2* expression in the M lineage is under the control of the Wnt/β-catenin asymmetry pathway. As shown in Fig. 6J-L, *gfp::sem-2* was expressed in both M.v(l/r)pa and M.v(l/r)pp (33.3%, *n*=26), and all the endogenous and ectopic SMs (85.7%, *n*=14) in *sys-1(RNAi)* animals. *sem-2* is also required for both the endogenous and the ectopic SMs produced in *sys-1(q544)* animals, as *sem-2(RNAi) sys-1(q544)* animals produced no SMs at all (100%, *n*=20). Thus, *sem-2* acts downstream of, and is negatively regulated by, SYS-1.

sem-2 exhibits mutually repressive interactions with *fozi-1* and *hlh-1*

Previous studies have shown that the MyoD homolog HLH-1 functions redundantly with the zinc-finger protein FOZI-1 to specify M-derived BWMs and CCs while repressing the SM fate (Harfe et al., 1998a; Amin et al., 2007). We found that *gfp::sem-2* was expressed in both the endogenous and the ectopic SMs in *fozi-1(RNAi)* animals (98.6%, *n*=70, Fig. 6M-O) and *hlh-1(RNAi)* animals (100%, *n*=25). Furthermore, no SMs were produced by *sem-2(RNAi); fozi-1(cc609)* animals (100%, *n*=71) or *sem-2(RNAi); hlh-1(cc561ts)* animals at the restrictive temperature (100%, *n*=30). These observations suggest that *sem-2* expression is repressed in the M-derived BWMs and CCs by HLH-1 and FOZI-1.

Because the SMs are transformed to BWMs in *sem-2(n1343)* mutants, we also examined whether *fozi-1* and *hlh-1* are expressed in the ectopic BWMs in *n1343* animals. As shown in Fig. 6P-S, *fozi-1*, which is transiently expressed in all M-derived BWMs prior to their differentiation in wild-type animals (Fig. 6P-Q), is expressed in the ectopic BWMs in *sem-2(n1343)* animals (Fig. 6R,S). Similar results were also obtained for *hlh-1* expression in *sem-2(n1343)* animals (data not shown). As the expression of *sem-2* in the SM mother cells (M.v(l/r)pa) overlaps with those of *hlh-1* and *fozi-1*, the above results suggest that the BWM-specifying factors FOZI-1 and HLH-1 and the SM-specifying factor SEM-2 mutually repress the expression of each other to maintain their proper expression pattern in the respective daughter cells derived from the SM mothers.

DISCUSSION

SEM-2/SoxC acts as a switch in a binary fate decision to promote a proliferative fate over a differentiated muscle fate

We have demonstrated that the single SoxC protein SEM-2 is both necessary and sufficient to specify the SM fate in the *C. elegans* postembryonic mesoderm. SMs are precursors that have the potential to proliferate and then differentiate into 16 non-striated egg-laying uterine and vulval muscles. Disrupting *sem-2* expression

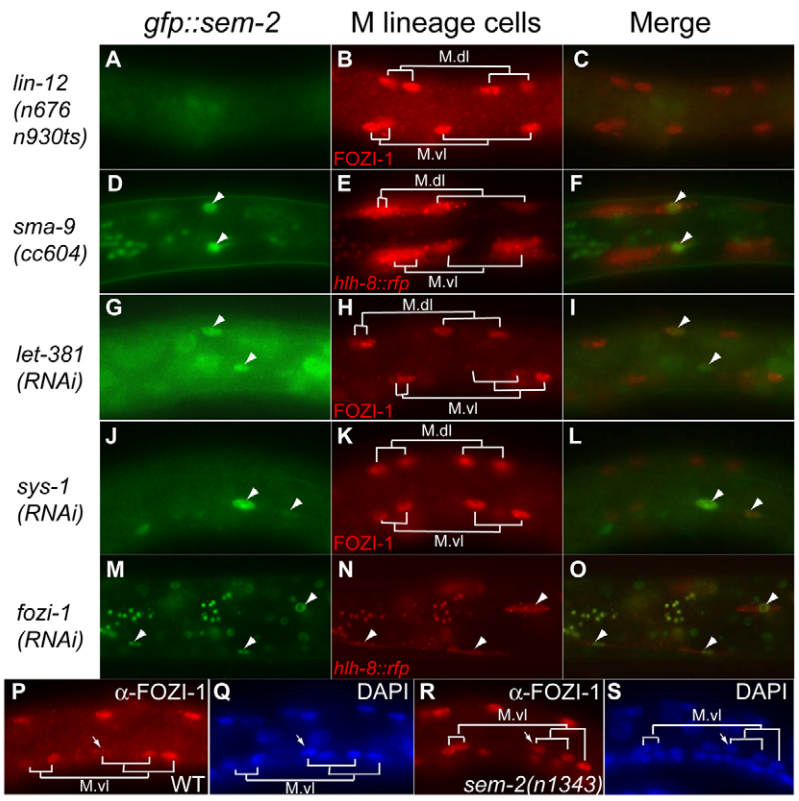

Fig. 6. The M lineage expression of *sem-2* is under the control of multiple signaling pathways and transcription factors. (**A-O**) *sem-2* M lineage expression in *lin-12(n676n930ts)* animals at the restrictive temperature (A-C), and in *sma-9(cc604)* (D-F), *let-381(RNAi)* (G-I), *sys-1(RNAi)* (J-L) and *fozi-1(RNAi)* (M-O) animals. Arrowheads indicate cells with *gfp::sem-2* expression. The animal shown in M-O is at the 2-SM stage. (**P-S**) *fozi-1* is ectopically expressed in M.vlpaa in *sem-2(n1343)* animals (R,S) compared with wild-type animals (P,Q), as shown by α-FOZI-1 immunostaining (P,R) and the corresponding DAPI staining images (Q,S). Arrows indicate M.vlpaa. Similar expression is also seen in the equivalent cells on the right side for A-L and P-S.

specifically in the M lineage leads to a transformation of SMs to cells that differentiate into striated BWMs. *sem-2* is specifically expressed in Mv(l/r)pa, a bi-potent precursor cell that asymmetrically divides to give rise to a SM and a BWM. *sem-2* expression is retained in the SM, perdures in its descendants that remain proliferative and ceases when these cells switch from a proliferation state and differentiate into mature egg-laying muscles (Fig. 2R). Forced overexpression of *sem-2* throughout the M lineage led to the loss of M-derived BWMs and CCs and to the gain of extra SMs (Fig. 3J-N). These experiments demonstrate that SEM-2 is both necessary and sufficient for promoting the egg-laying muscle precursor SM fate while preventing the *sem-2*-expressing cells from differentiating into BWMs. Consistent with this, ectopic *sem-2* expression was detected in mutant animals with extra SMs, and removing *sem-2* in these animals blocked the formation of both ectopic and normal SMs.

sem-2 is also broadly expressed during embryogenesis and essential for embryonic development. *sem-2* null mutant embryos were arrested at the threefold stage. Time-lapse movies showed that this arrest is preceded by delays in hypodermal migration and elongation. Besides the embryonic phenotype, RNAi knockdown of *sem-2* postembryonically caused a multivulval (Muv) phenotype, which may result from a fate specification defect or a proliferation defect in the vulval precursor cells. Further characterization of the embryonic and vulval phenotypes of *sem-2* mutants will help shed light on how *sem-2* functions in these processes.

A model for the specification of the non-striated muscle precursors, the SMs

Our findings provide an example of how lineage information and positional information are integrated to activate the cell type-specific expression of a cell fate determinant (in this case *sem-2*) in the specification of the sex myoblast cells. The M lineage expression of *sem-2* is controlled by a cis-acting M lineage enhancer located in the 4.5 kb intron (Fig. 4A). Disruption of this site by the transposon Tc1 in *n1343* animals specifically disrupted the expression and function of *sem-2* in the M lineage without affecting its expression in other cells. We showed that the Hox factors MAB-5 and LIN-39 and their co-factor, the PBC protein CEH-20, directly bind this site to activate *sem-2* expression in the M lineage (Fig. 5B,C). However, these three genes are also expressed in other cell types in *C. elegans* (Costa et al., 1988; Wagmaister et al., 2006; Jiang et al., 2009). In fact, it has been shown that MAB-5/CEH-20 and/or LIN-39/CEH-20 complexes directly activate the expression of *hlh-8* throughout the M lineage before terminal differentiation (Liu and Fire, 2000), of *egl-18*, *elt-6*, *eff-1* and *lag-2* in the developing vulva (Koh et al., 2002; Shemer and Podbilewicz, 2002; Takacs-Vellai et al., 2007), and of *egl-1* in the P11 lineage and the VC neurons (Liu et al., 2006; Potts et al., 2009). A Hox-independent role of CEH-20 has also been found in the activation of *mls-2* in the early M lineage (Jiang et al., 2008). Thus, additional factors must cooperate with Hox/PBC proteins to refine the cell type-specific expression of Hox/PBC targets.

Our results demonstrate that, in addition to the Hox factors and CEH-20, the integration of dorsal-ventral (D/V) and anterior-posterior (A/P) positional information is crucial in dictating the specific localization of *sem-2* within the M lineage. Previous studies have shown that LIN-12/Notch signaling and SMA-9, which antagonizes the Sma/Mab TGF-β signaling pathway, work independently to control D/V patterning of the M lineage, and that they both regulate the expression of the FoxF/C transcription factor LET-381 for its dorsal M lineage expression (Foehr et al., 2006; Foehr and Liu, 2008; Amin et al., 2010). We showed that the proper pattern of *sem-2* expression is due to the presence of LIN-

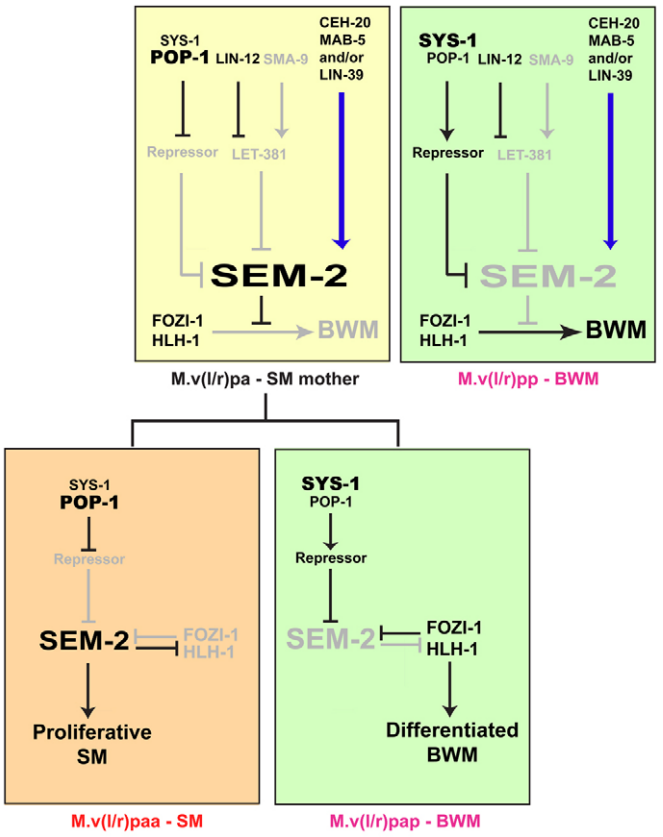

Fig. 7. A model for *sem-2* regulation and SM fate specification. A proposed model on how *sem-2* expression in the M lineage is regulated (see Discussion). Thick blue lines represent direct regulatory relationships, black lines represent relationships based solely on genetic data and do not distinguish between direct and indirect. Gray lines and text indicate a lack of expression in the indicated cell.

12/Notch signaling and the absence of SMA-9 along the D/V axis. Furthermore, *sem-2* functions downstream of *let-381* and is negatively regulated by *let-381* in the dorsal M lineage (Fig. 6G-I). Currently, it is not clear whether this negative regulation of *sem-2* by *let-381* is direct or not.

Along the AP axis at the 16-M stage, POP-1/TCF functions as a repressor in M.v(l/r)pa due to low nuclear levels of SYS-1/β-catenin and is converted to an activator in M.v(l/r)pp due to high nuclear levels of SYS-1/β-catenin (Kidd et al., 2005; Liu et al., 2008; Amin et al., 2009). We showed that *sys-1(RNAi)* caused ectopic expression of *gfp::sem-2* and ectopic production of SMs (Fig. 6J-L). As *sem-2* is expressed in M.v(l/r)pa where POP-1 acts as a repressor, these results suggest that POP-1 must activate the expression of a repressor of *sem-2* in M.v(l/r)pa to restrict *sem-2* expression along the anterior-posterior axis in the M lineage. We hypothesize that an additional factor(s) must be present either in M.v(l/r)aa to repress *sem-2* expression or in M.v(l/r)pa to promote *sem-2* expression in order to ensure specific expression of *sem-2* in M.v(l/r)pa (see model in Fig. 7).

Immediately after the SM mother cell M.v(l/r)pa divides, both its daughter cells express *gfp::sem-2* (Fig. 2I-K). However, *gfp::sem-2* perdures in the anterior SM lineage, but not in the posterior BWMs. *hlh-1* and *fozi-1* exhibit a reciprocal expression pattern: they are turned off in the SMs, but remain expressed in the BWMs (Harfe et al., 1998a; Amin et al., 2007). The initiation of the asymmetric expression of *sem-2* versus *hlh-1* and *fozi-1* may be due to the nuclear POP-1 and SYS-1 asymmetry along the anterior-posterior axis (Amin et al., 2009). Once initiated, *sem-2* and *hlh-1* and *fozi-1* appear to mutually repress each other to maintain their proper expression, as *sem-2* is expressed in all the ectopic SMs in *hlh-1* and *fozi-1* mutant or RNAi animals, while *fozi-1* and *hlh-1* are expressed in the ectopic BWMs in *sem-2(n1343)* animals.

An evolutionarily conserved role of SoxC family members in cell fate specification and cell proliferation

SEM-2 is a member of the SoxC subfamily, which includes Sox14 in *Drosophila* and Sox4, Sox11 and Sox12 in vertebrates (Bowles et al., 2000). *Drosophila* Sox14 exhibits a dynamic expression pattern during development and regulates dendrite severing and other unknown processes essential for fly metamorphosis (Sparkes et al., 2001; Kirilly et al., 2009; Ritter and Beckstead, 2010). Functional studies on vertebrate SoxC members have shown that SoxC proteins play essential roles in multiple lineages such as oligodendrocytes, B lymphocytes, osteoblasts and others to regulate cell fate specification, cell differentiation and cell survival (Penzo-Mendez, 2009). In humans, Sox4 overexpression has often been detected in cases of prostate cancer (Dhanasekaran et al., 2001; Ernst et al., 2002; Lapointe et al., 2004; Liu et al., 2006; Luo et al., 2001; Magee et al., 2001; Rhodes et al., 2002; Welsh et al., 2001). Sox4 is also overexpressed in many other types of cancers, including leukemias (Andersson et al., 2007), melanomas (Talantov et al., 2005), glioblastomas (Sun et al., 2006), medulloblastomas (Lee et al., 2002), bladder cancer (Aaboe et al., 2006) and lung cancer (Friedman et al., 2004). Similarly, Sox11 is highly expressed in medulloblastomas (Lee et al., 2002), gliomas (Weigle et al., 2005), non-B cell lymphomas (Wang et al., 2008) and epithelial ovarian tumors (Brennan et al., 2009). However, the underlying mechanism by which the Sox proteins may contribute to cancer is not fully understood. We have found that *sem-2* is specifically expressed in the proliferating SM mother cells, the SMs and their descendants prior to terminal differentiation (Fig. 2R). Furthermore, increasing *sem-2* level throughout the M lineage is sufficient to transform other M lineage cells into proliferating SM-like cells, even though they express normal levels of CC- and BWM-specifying and differentiating factors. Thus, it appears that increasing *sem-2* level is sufficient to tip the balance between proliferative and differentiative factors and allows for proliferation. This pro-proliferation function of SEM-2 may be a conserved role of the group C Sox proteins.

Hox and PBC proteins have also been implicated in tumorigenesis. There are numerous examples of Hox gene overexpression in various tumor types (Argiropoulos and Humphries, 2007; Moreno, 2010; Shah and Sukumar, 2010). Given our finding that *sem-2* expression in the M lineage is directly activated by Hox-PBC proteins and that there is overexpression of Sox4 and Sox11 in a variety of cancers, it is possible that the oncogenic activity of Hox genes in some cases is due to their direct activation of SoxC gene expression.

Acknowledgements
We thank the *C. elegans* Genetics Center, the knockout consortium and Yuji Kohara for strains and yk clones; Nirav Amin for examining *sem-2* expression in *let-381(RNAi)* animals; Alex Beatty and Rich McCloskey for help with preparing embryos for the time lapse movies; Nirav Amin and Jared Hale for helpful discussions and critical comments on the manuscript. This work was supported by NIH R01 GM066953 (to J.L.). Deposited in PMC for release after 12 months.

Competing interests statement
The authors declare no competing financial interests.

Supplementary material
Supplementary material for this article is available at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.062240/-/DC1

References

Aaboe, M., Birkenkamp-Demtroder, K., Wiuf, C., Sorensen, F. B., Thykjaer, T., Sauter, G., Jensen, K. M., Dyrskjot, L. and Orntoft, T. (2006). SOX4 expression in bladder carcinoma: Clinical aspects and in vitro functional characterization. *Cancer Res.* **66**, 3434-3442.

Amin, N. M., Hu, K., Pruyne, D., Terzic, D., Bretscher, A. and Liu, J. (2007). 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* **134**, 19-29.

Amin, N. M., Lim, S. E., Shi, H., Chan, T. L. and Liu, J. (2009). A conserved six-eya cassette acts downstream of wnt signaling to direct non-myogenic versus myogenic fates in the C. elegans postembryonic mesoderm. *Dev. Biol.* **331**, 350-360.

Amin, N. M., Shi, H. and Liu, J. (2010). The FoxF/FoxC factor LET-381 directly regulates both cell fate specification and cell differentiation in C. elegans mesoderm development. *Development* **137**, 1451-1460.

Andersson, A., Ritz, C., Lindgren, D., Eden, P., Lassen, C., Heldrup, J., Olofsson, T., Rade, J., Fontes, M., Porwit-Macdonald, A. et al. (2007). Microarray-based classification of a consecutive series of 121 childhood acute leukemias: Prediction of leukemic and genetic subtype as well as of minimal residual disease status. *Leukemia* **21**, 1198-1203.

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.

Argiropoulos, B. and Humphries, R. K. (2007). Hox genes in hematopoiesis and leukemogenesis. *Oncogene* **26**, 6766-6776.

Bowles, J., Schepers, G. and Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev. Biol.* **227**, 239-255.

Brennan, D. J., Ek, S., Doyle, E., Drew, T., Foley, M., Flannelly, G., O'Connor, D. P., Gallagher, W. M., Kilpinen, S., Kallioniemi, O. P. et al. (2009). The transcription factor Sox11 is a prognostic factor for improved recurrence-free survival in epithelial ovarian cancer. *Eur. J. Cancer* **45**, 1510-1517.

Broitman-Maduro, G., Maduro, M. F. and Rothman, J. H. (2005). The noncanonical binding site of the MED-1 GATA factor defines differentially regulated target genes in the C. elegans mesendoderm. *Dev. Cell* **8**, 427-433.

Chan, S. K. and Mann, R. S. (1996). A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. *Proc. Natl. Acad. Sci. USA* **93**, 5223-5228.

Costa, M., Weir, M., Coulson, A., Sulston, J. and Kenyon, C. (1988). Posterior pattern formation in C. elegans involves position-specific expression of a gene containing a homeobox. *Cell* **55**, 747-756.

Desai, C., Garriga, G., McIntire, S. L. and Horvitz, H. R. (1988). A genetic pathway for the development of the Caenorhabditis elegans HSN motor neurons. *Nature* **336**, 638-646.

Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A. and Chinnaiyan, A. M. (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822-826.

Ernst, T., Hergenhahn, M., Kenzelmann, M., Cohen, C. D., Bonrouhi, M., Weninger, A., Klaren, R., Grone, E. F., Wiesel, M., Gudemann, C. et al. (2002). Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: A gene expression analysis on total and microdissected prostate tissue. *Am. J. Pathol.* **160**, 2169-2180.

Foehr, M. L. and Liu, J. (2008). Dorsoventral patterning of the C. elegans postembryonic mesoderm requires both LIN-12/Notch and TGFbeta signaling. *Dev. Biol.* **313**, 256-266.

Foehr, M. L., Lindy, A. S., Fairbank, R. C., Amin, N. M., Xu, M., Yanowitz, J., Fire, A. Z. and Liu, J. (2006). An antagonistic role for the C. elegans schnurri homolog SMA-9 in modulating TGFbeta signaling during mesodermal patterning. *Development* **133**, 2887-2896.

Friedman, R. S., Bangur, C. S., Zasloff, E. J., Fan, L., Wang, T., Watanabe, Y. and Kalos, M. (2004). Molecular and immunological evaluation of the transcription factor SOX-4 as a lung tumor vaccine antigen. *J. Immunol.* **172**, 3319-3327.

Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., et al. (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775-1787.

Greenwald, I. S., Sternberg, P. W. and Horvitz, H. R. (1983). The Lin-12 locus specifies cell fates in Caenorhabditis elegans. *Cell* **34**, 435-444.

Harfe, B. D. and Fire, A. (1998). Muscle and nerve-specific regulation of a novel NK-2 class homeodomain factor in Caenorhabditis elegans. *Development* **125**, 421-429.

Harfe, B. D., Branda, C. S., Krause, M., Stern, M. J. and Fire, A. (1998a). MyoD and the specification of muscle and non-muscle fates during postembryonic development of the C. elegans mesoderm. *Development* **125**, 2479-2488.

Harfe, B. D., Vaz Gomes, A., Kenyon, C., Liu, J., Krause, M. and Fire, A. (1998b). Analysis of a Caenorhabditis elegans twist homolog identifies conserved and divergent aspects of mesodermal patterning. *Genes Dev.* **12**, 2623-2635.

Jiang, Y., Shi, H., Amin, N. M., Sultan, I. and Liu, J. (2008). Mesodermal expression of the *C. elegans* HMX homolog *mls-2* requires the PBC homolog CEH-20. *Mech. Dev.* **125**, 451-461.

Jiang, Y., Shi, H. and Liu, J. (2009). Two Hox cofactors, the Meis/Hth homolog UNC-62 and the Pbx/Exd homolog CEH-20, function together during C. elegans postembryonic mesodermal development. *Dev. Biol.* **334**, 535-546.

Kamath, R. S. and Ahringer, J. (2003). Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* **30**, 313-321.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M. et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421**, 231-237.

Kidd, A. R., 3rd, Miskowski, J. A., Siegfried, K. R., Sawa, H. and Kimble, J. (2005). A beta-catenin identified by functional rather than sequence criteria and its role in Wnt/MAPK signaling. *Cell* **121**, 761-772.

Kirilly, D., Gu, Y., Huang, Y., Wu, Z., Bashirullah, A., Low, B. C., Kolodkin, A. L., Wang, H. and Yu, F. (2009). A genetic pathway composed of Sox14 and mical governs severing of dendrites during pruning. *Nat. Neurosci.* **12**, 1497-1505.

Knoepfler, P. S., Lu, Q. and Kamps, M. P. (1996). Pbx-1 Hox heterodimers bind DNA on inseparable half-sites that permit intrinsic DNA binding specificity of the Hox partner at nucleotides 3′ to a TAAT motif. *Nucleic Acids Res.* **24**, 2288-2294.

Koh, K., Peyrot, S. M., Wood, C. G., Wagmaister, J. A., Maduro, M. F., Eisenmann, D. M. and Rothman, J. H. (2002). Cell fates and fusion in the C. elegans vulval primordium are regulated by the EGL-18 and ELT-6 GATA factors – apparent direct targets of the LIN-39 Hox protein. *Development* **129**, 5171-5180.

Lapointe, J., Li, C., Higgins, J. P., van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U. et al. (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl. Acad. Sci. USA* **101**, 811-816.

Lee, C. J., Appleby, V. J., Orme, A. T., Chan, W. I. and Scotting, P. J. (2002). Differential expression of SOX4 and SOX11 in medulloblastoma. *J. Neurooncol.* **57**, 201-214.

Liu, H., Strauss, T. J., Potts, M. B. and Cameron, S. (2006). Direct regulation of Egl-1 and of programmed cell death by the Hox protein MAB-5 and by CEH-20, a C. elegans homolog of Pbx1. *Development* **133**, 641-650.

Liu, J. and Fire, A. (2000). Overlapping roles of two Hox genes and the exd ortholog Ceh-20 in diversification of the C. elegans postembryonic mesoderm. *Development* **127**, 5179-5190.

Liu, J., Phillips, B. T., Amaya, M. F., Kimble, J. and Xu, W. (2008). The C. elegans SYS-1 protein is a bona fide beta-catenin. *Dev. Cell* **14**, 751-761.

Liu, P., Ramachandran, S., Ali Seyed, M., Scharer, C. D., Laycock, N., Dalton, W. B., Williams, H., Karanam, S., Datta, M. W., Jaye, D. L. et al. (2006). Sex-determining region Y Box 4 is a transforming oncogene in human prostate cancer cells. *Cancer Res.* **66**, 4011-4019.

Luo, J., Duggan, D. J., Chen, Y., Sauvageot, J., Ewing, C. M., Bittner, M. L., Trent, J. M. and Isaacs, W. B. (2001). Human prostate cancer and benign prostatic hyperplasia: Molecular dissection by gene expression profiling. *Cancer Res.* **61**, 4683-4688.

Maeda, I., Kohara, Y., Yamamoto, M. and Sugimoto, A. (2001). Large-scale analysis of gene function in Caenorhabditis elegans by high-throughput RNAi. *Curr. Biol.* **11**, 171-176.

Magee, J. A., Araki, T., Patil, S., Ehrig, T., True, L., Humphrey, P. A., Catalona, W. J., Watson, M. A. and Milbrandt, J. (2001). Expression profiling reveals hepsin overexpression in prostate cancer. *Cancer Res.* **61**, 5692-5696.

Mann, R. S. and Affolter, M. (1998). Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* **8**, 423-429.

Mello, C. C., Kramer, J. M., Stinchcomb, D. and Ambros, V. (1991). Efficient gene transfer in C. elegans: Extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959-3970.

Moreno, C. S. (2010). The sex-determining region Y-Box 4 and homeobox C6 transcriptional networks in prostate cancer progression: Crosstalk with the Wnt, Notch, and PI3K pathways. *Am. J. Pathol.* **176**, 518-527.

Penzo-Mendez, A. I. (2009). Critical roles for SoxC transcription factors in development and cancer. *Int. J. Biochem. Cell Biol.* **42**, 425-428.

Phochanukul, N. and Russell, S. (2010). No backbone but lots of Sox: Invertebrate Sox genes. *Int. J. Biochem. Cell Biol.* **42**, 453-464.

Potts, M. B., Wang, D. P. and Cameron, S. (2009). Trithorax, Hox, and TALE-class homeodomain proteins ensure cell survival through repression of the BH3-only gene Egl-1. *Dev. Biol.* **329**, 374-385.
Rhodes, D. R., Barrette, T. R., Rubin, M. A., Ghosh, D. and Chinnaiyan, A. M. (2002). Meta-analysis of microarrays: Interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res.* **62**, 4427-4433.
Ritter, A. R. and Beckstead, R. B. (2010). Sox14 is required for transcriptional and developmental responses to 20-hydroxyecdysone at the onset of drosophila metamorphosis. *Dev. Dyn.* **239**, 2685-2694.
Shah, N. and Sukumar, S. (2010). The Hox genes and their roles in oncogenesis. *Nat. Rev. Cancer* **10**, 361-371.
Shemer, G. and Podbilewicz, B. (2002). LIN-39/Hox triggers cell division and represses EFF-1/fusogen-dependent vulval cell fusion. *Genes Dev.* **16**, 3136-3141.
Sparkes, A. C., Mumford, K. L., Patel, U. A., Newbury, S. F. and Crane-Robinson, C. (2001). Characterization of an SRY-like gene, DSox14, from Drosophila. *Gene* **272**, 121-129.
Sulston, J. E. and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev. Biol.* **56**, 110-156.
Sun, L., Hui, A. M., Su, Q., Vortmeyer, A., Kotliarov, Y., Pastorino, S., Passaniti, A., Menon, J., Walling, J., Bailey, R. et al. (2006). Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell* **9**, 287-300.
Takacs-Vellai, K., Vellai, T., Chen, E. B., Zhang, Y., Guerry, F., Stern, M. J. and Muller, F. (2007). Transcriptional control of Notch signaling by a HOX and a PBX/EXD protein during vulval development in C. elegans. *Dev. Biol.* **302**, 661-669.
Talantov, D., Mazumder, A., Yu, J. X., Briggs, T., Jiang, Y., Backus, J., Atkins, D. and Wang, Y. (2005). Novel genes associated with malignant melanoma but not benign melanocytic lesions. *Clin. Cancer Res.* **11**, 7234-7242.
Timmons, L. and Fire, A. (1998). Specific interference by ingested dsRNA. *Nature* **395**, 854.
Wagmaister, J. A., Miley, G. R., Morris, C. A., Gleason, J. E., Miller, L. M., Kornfeld, K. and Eisenmann, D. M. (2006). Identification of cis-regulatory elements from the C. Elegans Hox gene Lin-39 required for embryonic expression and for regulation by the transcription factors LIN-1, LIN-31 and LIN-39. *Dev. Biol.* **297**, 550-565.
Wang, X., Asplund, A. C., Porwit, A., Flygare, J., Smith, C. I., Christensson, B. and Sander, B. (2008). The subcellular Sox11 distribution pattern identifies subsets of mantle cell lymphoma: Correlation to overall survival. *Br. J. Haematol.* **143**, 248-252.
Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. and Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, e36.
Wegner, M. (2010). All purpose Sox: The many roles of Sox proteins in gene expression. *Int. J. Biochem. Cell Biol.* **42**, 381-390.
Weigle, B., Ebner, R., Temme, A., Schwind, S., Schmitz, M., Kiessling, A., Rieger, M. A., Schackert, G., Schackert, H. K. and Rieber, E. P. (2005). Highly specific overexpression of the transcription factor SOX11 in human malignant gliomas. *Oncol. Rep.* **13**, 139-144.
Welsh, J. B., Sapinoso, L. M., Su, A. I., Kern, S. G., Wang-Rodriguez, J., Moskaluk, C. A., Frierson, H. F., Jr. and Hampton, G. M. (2001). Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res.* **61**, 5974-5978.