Published in final edited form as:

Gene Expr Patterns. 2018 December ; 30: 71–81. doi:10.1016/j.gep.2018.10.003.

A tissue-specific enhancer of the *C. elegans nhr-67/tailless* gene drives coordinated expression in uterine stem cells and the differentiated anchor cell

Shari Bodofsky[#], Katarina Liberatore[#], Lauren Pioppo, Dominic Lapadula, Lily Thompson, Susanna Birnbaum, George McClung, Akshara Kartik, Sheila Clever, and Bruce Wightman^{*}
Biology Department, Muhlenberg College, Allentown, PA, 18104, USA.

[#] These authors contributed equally to this work.

Abstract

The *nhr-67* nuclear receptor gene of *Caenorhabditis elegans* encodes the ortholog of the *Drosophila tailless* and vertebrate *Tlx* genes. In *C. elegans*, *nhr-67* plays multiple roles in the development of the uterus during L2 and L3 larval stages. Four pre-VU cells are born in the L2 stage and form the precursor complement for the ventral surface of the mature uterus. One of the four pre-VU cells becomes the anchor cell (AC), which exits the cell cycle and differentiates, while the remaining three VU cells serve as stem cells that populate the ventral uterus. The *nhr-67* gene functions in the development of both VU cell lineages and AC differentiation. Hypomorphic mutations in *nhr-67* identify a 276bp region of the distal promoter that is sufficient to activate *nhr-67* expression in pre-VU cells and the AC. The 276bp region includes 8 conserved potential cis-acting sites, including two E boxes and a nuclear receptor binding site. Mutational analysis demonstrates that the two E boxes are required for expression of *nhr-67* in uterine precursor cells. The *E/daughterless* ortholog HLH-2 binds these sites as a homodimer, thus playing a central role in activating *nhr-67* expression in the uterine precursors. At least two other binding activities, one of which may be the *nhr-25/Ftz-F1* nuclear receptor transcription factor, also contribute to uterine precursor cell expression. The organization of the *nhr-67* uterine precursor enhancer is compared to similar conserved enhancers in the *egl-43*, *lag-2*, and *lin-3* genes, which contain the same HLH-2-binding E boxes and are similarly expressed in both pre-VU cells and the AC. This basic regulatory module allows the coordinated expression of at least four genes. Expression of genes in different cells that must coordinate to form a mature organ is driven by a shared set of promoter elements, which integrate multiple transcription factor inputs.

Graphical Abstract

^{*}Corresponding author: wightman@muhlenberg.edu.

1. Introduction

Evolution of animal forms occurs through changes in developmental programs that lead to morphological variation. These changes are, in part, the outcome of genetic modifications in the regulation of key transcription factors and the cis-acting sites they bind (McLean et al., 2011; Maher, 2012; Levine et al., 2014). The study of developmental regulatory pathways in invertebrate model systems has been a powerful approach to understanding how changes in gene regulation lead to cell differentiation and pattern formation (Carroll, 2008). Tissue-specific patterns of gene expression during development are achieved through specialized enhancer sequences that are bound by trans-acting developmental control genes.

The development of the *C. elegans* uterus has served as a model for understanding the mechanisms of organogenesis. A series of cell-cell signaling events utilize a *Notch* pathway to create the adult hermaphrodite uterus, which consists of sixty cells that contain a lumen and connect to the dorsal side of the vulva (Newman et al, 1995, 1996). The entire ventral uterus is generated from four cells (the pre-VU cells: Z1ppa, Z1ppp, Z4aaa, Z4aap) that are born in the early L2 stage (Fig. 1A). These four precursor cells are also referred to as α and β cells (Sallee and Greenwald, 2015). The anchor cell (AC) of the ventral uterus plays a pivotal role in coordinating the development of the uterus. The AC is generated from one of two medial equipotent pre-VU cells (Z1ppp or Z4aaa; the “ α ” cells) in the late L2 somatic

gonad. The cell that does not become the AC becomes a VU (ventral uterine) cell, similar to the other two VU cells. The capability of differentiating into an AC depends on the bHLH transcription factor *hlh-2*, which encodes the ortholog of human E2A and *Drosophila daughterless* (Karp and Greenwald, 2003, 2004). The selection of which medial pre-VU cell executes each fate depends on birth order and reciprocal signaling between a *delta*-like signal, LAG-2, and a *Notch*-like receptor, LIN-12 (Greenwald et al., 1983; Kimble and Simpson, 1997; Karp and Greenwald, 2003). The two other pre-VU cells become VU cells (without a requirement for AC signal), for a total complement of three VU and one AC in late L2 animals. The AC does not divide and is terminally differentiated in the early L3, but the three VU cells divide multiple times during the L3 and L4 stages to produce 36 cells that form adult ventral uterine tissue (Kimble and Hirsh, 1979; Newman et al., 1996). The AC, in turn, signals to adjacent uterine cells and to underlying vulval cells to coordinate the differentiation of multiple cell types. Therefore, the AC-VU regulatory system provides a relatively well-understood iterative signaling system where a single coordinating cell, the AC, orchestrates the development of surrounding cells.

The *C. elegans* *tailless* ortholog *nhr-67* encodes a nuclear receptor transcription factor that plays key roles in the differentiation of the AC and VU cells (Verghese et al., 2011; Matus et al., 2015). While deletions of the *nhr-67* coding sequence cause embryonic lethality and L1 arrest, three hypomorphic mutations that map to the promoter region of the gene reduce or eliminate *nhr-67* function in the AC and VU, yet are viable and retain *nhr-67* function in other tissues (Verghese et al., 2011). This observation suggested the existence of a tissue-specific promoter element located within a 276bp region about 1.4 kb upstream of the *nhr-67* transcription start site. Analysis of the promoter element predicted by the location of these mutations provides an opportunity to understand the mechanisms by which the dynamic *nhr-67* expression pattern in pre-VU cells and the AC is achieved, serving as a case study for how genes come to be regulated differentially over time in a subset of related cells.

2. Results and Discussion

2.1 *nhr-67* is expressed and functions in pre-VU cells and the AC

nhr-67 is expressed in the four pre-VU cells shortly after their birth in early L2 (Verghese et al., 2011) (Fig. 1A). In late L2, after the *lin-12*-mediated AC-VU decision, *nhr-67* expression increases to high levels in the AC, but decreases sharply in VU descendants. *nhr-67* functions at multiple steps in ventral uterine development. First, *nhr-67* is required for expression of *lin-12* in the four pre-VU cells (Verghese et al., 2011). Second, *nhr-67* functions in the differentiation of the AC. Mutations in *nhr-67* cause a loss of expression of the *zmp-1* zinc metalloproteinase gene in the AC, and a failure of the AC to invade the underlying pre-vulval tissue (Verghese et al., 2011; Matus et al., 2015). The AC's of *nhr-67(RNAi)*-treated animals also undergo extra divisions, indicating that *nhr-67* is required for exit from the cell cycle (Matus et al., 2015). Third, *nhr-67* is required in VU lineages for the proper development of VU descendants (Verghese et al., 2011). Taken together, these data demonstrate that *nhr-67* functions in the development of both the AC and VU cells.

Because *nhr-67* is required for the proper differentiation of the precursor to the AC (pre-VU), as well as the AC itself, it is possible that the AC defects are secondary to earlier

defects in the pre-VU precursor cell. We previously described a two-AC phenotype caused by mutations in *nhr-67* and *nhr-67(RNAi)* (Verghese et al., 2011). Because expression of *lin-12/Notch* in pre-VU and VU cells depends on *nhr-67*, we reasoned that the additional AC's reflect conversion of a VU cell into an AC, due to the absence of *lin-12/Notch* signaling. Loss-of-function mutations in *lin-12* cause both medial pre-VU cells to become AC's, while gain-of-function mutations in *lin-12* cause both medial pre-VU cells to become VU cells and no AC's to form (Greenwald et al., 1983; Kimble and Simpson, 1997). However, Matus et al. (2015) observed more than two AC-like cells in *nhr-67(RNAi)* animals, and demonstrated that the *nhr-67* differentiation defect includes a failure of the defective AC to exit the cell-cycle. Furthermore, this function appeared to require *nhr-67* after AC specification. Therefore, some or all of the extra AC-like cells in *nhr-67* mutants could be due to the AC dividing instead of VU to AC conversion.

In order to test whether additional AC's caused by the loss of *nhr-67* depend on *lin-12* transcriptional regulation by *nhr-67*, we performed an epistasis test with a *lin-12(gf)* mutant. In *lin-12(137gf)* mutants, both potential AC's become VU cells due to excessive *Notch* signaling response (Greenwald et al., 1983). If the additional AC-like cells caused by loss of *nhr-67* are a result of a loss of *lin-12* expression in pre-VU and VU cells, then we would expect that *nhr-67(RNAi)* would be able to suppress the *lin-12(gf)* phenotype. Loss of *nhr-67* should result in a loss of expression of the activated LIN-12(gf) product in pre-VU and VU cells, and therefore partially or completely suppress the *lin-12(gf)* phenotype. If *lin-12* is not transcribed in pre-VU and VU cells, then no hyperactive LIN-12 protein can be made. In contrast, we found that *nhr-67(RNAi)* treatment was not able to suppress the *lin-12(gf)* AC to VU conversion phenotype (Fig. 2). This result is consistent with the evidence of Matus et al. (2015) showing that loss of *nhr-67* results in supernumerary division of the AC, rather than VU to AC conversion. Regulation of *lin-12* in pre-VU cells and their descendants may still be an important feature of *nhr-67* function, but it is not sufficient to explain the two-AC phenotype. This result, combined with the previous analysis, argues that *nhr-67* functions in both pre-VU and AC's. Therefore, understanding how *nhr-67* comes to be expressed in both pre-VU and the AC is important for illuminating the regulatory pathways that control uterine development.

2.2 A regulatory region in the *nhr-67* promoter is required for gonadal expression in larva

Three independent *nhr-67* mutations (*pf2*, *pf88*, and *pf159*) delete overlapping portions of the *nhr-67* promoter (127, 398, and 494 bp, respectively) (Fig. 1B). All three alleles are homozygous viable, unlike *nhr-67* coding region deletions, and display protruding-vulva and egg-laying defects that are the consequence of earlier AC and VU developmental defects (Verghese et al., 2011) and/or defects in vulval morphogenesis (Fernandes and Sternberg, 2007). The smaller *pf2* deletion causes a weaker phenotypic effect, while *pf88* and *pf159* show similar strong effects and behave as genetic null alleles for gonadal function (Verghese et al., 2011). These observations suggested that the viable *nhr-67* alleles result from the deletion of a tissue-specific promoter element that affects *nhr-67* expression in the developing uterus.

We used an evolutionary sequence comparison approach to identify conserved potential cis-acting promoter elements in the region defined by the three deletions. We aligned the sequence of the *nhr-67* promoter from four different related nematode species (*C. elegans*, *C. briggsae*, *C. brenneri*, and *C. remanei*), identifying eight conserved candidate elements in a 276bp region of the promoter located about 1.4kb upstream of the presumptive transcription start site (Figs. 1B, 3, 4A). These elements are hereafter referred to as elements 1 through 8, from upstream to downstream. The total size of this region includes some 5' and 3' flanking DNA sequences that are not well-conserved. In order to test whether this region is necessary for *nhr-67* expression in pre-VU cells and AC, we created three constructs that replicate the three *pf* deletions in the context of a *nhr-67::GFP* translational fusion reporter. The two larger deletions resulted in a complete loss of GFP expression in pre-VU cells and AC, although GFP expression was observed in some head neurons (Fig. 4C). Consistent with its weaker phenotype, the smaller *pf2* deletion (which removes only the first 4 elements) showed weak and variable GFP expression in both cell types. All GFP reporters, wild-type and deleted, were expressed in several head neurons, which we have not identified (data not shown). Therefore, the loss of the 276bp region eliminates expression in the pre-VU cells and AC without eliminating expression in some *nhr-67*-positive neurons. Additional conserved elements present more distal and more proximal to the core promoter (Fig. 1B) are presumably responsible for *nhr-67* expression in neurons and other tissues. These observations link the loss of *nhr-67* function in the pre-VU cells and AC to loss of *nhr-67* expression in these cells.

In order to test the sufficiency of the 276bp region for expression in pre-VU and AC, we created an “enhancer-trap” construct that fused the 276bp region directly adjacent to the GFP gene under the control of a *pes-10* minimal promoter (Fig. 4A). This construct was able to drive variable expression of GFP in the pre-VU cells and AC that recapitulates the cell-specificity and temporal regulation of the intact *nhr-67* promoter, demonstrating that this region sufficient for gonadal expression (Fig. 4B,D). Levels of expression from this construct were less robust than GFP constructs with the larger native *nhr-67* promoter; therefore other promoter regions are required for full expression levels. A smaller construct including only 160bp and elements 1 through 7 gave slightly higher levels of expression with a similar temporal dynamic (Fig. 4D), indicating that element 8 is dispensable and may have a slight repressive effect. Both constructs also activated GFP expression in one to two pairs of head neurons, which we have not investigated further. These observations indicate that the 160bp region acts as a cell-specific enhancer to activate expression of *nhr-67* in the uterine precursor cells (as well as a small number of neurons). While other promoter elements probably contribute to *nhr-67* transcription in pre-VU cells and the AC, this region is sufficient to confer the spatial and temporal pattern of expression observed with the intact promoter.

2.3 Sites 1 and 7 are E boxes that are necessary for pre-VU and AC expression

Site 1 contains a conserved CACCTG sequence and site 7 contains the same sequence on the opposite strand (Fig. 3). This sequence is a canonical binding site (E box) for bHLH transcription factors (Massari and Murre, 2000). A genome-wide approach for identifying candidate bHLH binding sites in *C. elegans* identified these sequences as candidate targets

for HLH-2 heterodimers with 9 candidate HLH partners (HLH-3, -4, -10, -14, -15, -19, HND-1 and LIN-32) (Grove et al., 2009). Directed studies on the binding of HLH-2 have demonstrated that this sequence will also bind HLH-2 homodimers both *in vivo* and *in vitro* (Krause et al., 1997; Harfe et al., 1998; Sallee and Greenwald, 2015; Sallee et al., 2017).

In order to test the hypothesis that sites 1 and 7 are important for regulation of *nhr-67* in the pre-VU and AC, we created a construct that included the 160bp region without site 1 and site 7 (Fig. 4D). This construct was unable to activate detectable GFP expression in the pre-VU cells or AC (although neuronal expression and background intestinal expression were observed). Therefore, sites 2 through 6 are not sufficient for pre-VU or AC expression, highlighting the dominant role of the two E boxes. Reintroduction of site 8 did not restore pre-VU or AC expression, indicating that loss of the two E boxes is the cause of loss of expression. Constructs that contain only sites 1 and 7 (immediately adjacent rather than in their normal separation from each other) and an artificial construct with three E boxes (two copies of site 1 and one copy of site 7) in their normal half-turn helical register with respect to each other, did not activate expression (Fig. 4D). Furthermore, a construct that includes site 1 and site 7 separated by 140 bp of unrelated and non-conserved DNA (from a downstream segment of the *nhr-67* promoter) also did not activate expression, indicating that two E boxes are not sufficient, even when separated by their normal distance. Likewise, the full promoter construct that includes the *pf159* deletion retains a single site 1, but was not able to activate detectable expression in the gonad (Fig. 4C). In contrast, the full promoter construct that includes the *pf2* deletion retains a single copy of site 7 along with sites 5, 6, and 8 was able to activate modest expression in both pre-VU cells and AC. Taken together, these data demonstrate that sites 1 and 7 are necessary, but not sufficient for gonadal transcriptional activation by the 160bp region.

2.4 HLH-2 appears to bind E box sites 1 and 7 as a homodimer

The *hlh-2* gene has been shown to play a critical role in both AC differentiation and AC-VU signaling (Karp and Greenwald, 2003, 2004). It is expressed in all four pre-VU cells at a time similar to the appearance of *nhr-67* in the pre-VU cells (Karp and Greenwald, 2003; Sallee and Greenwald, 2015). Late in the L2 stage HLH-2 protein becomes destabilized in the three VU cells, but is stabilized at high levels in the AC (Sallee and Greenwald, 2015). Degradation of HLH-2 protein parallels the reduction of expression of *nhr-67* in the VU cells at this time (Fig. 4B) (Verghese et al., 2011). Therefore, if HLH-2 activity is responsible for activation of *nhr-67* expression, loss of HLH-2 protein in the VU cells may serve as the primary explanation for the decline in *nhr-67* expression in late L2 and L3 VU cells.

Sallee and Greenwald (2015) demonstrated that HLH-2 homodimers were responsible for activation of *lin-3* expression in the AC. Since the *lin-3* promoter also has a pair of CACCTG E boxes, just like the *nhr-67* 276bp enhancer, we reasoned that HLH-2 homodimers might also be responsible for activation of *nhr-67* expression. The CACCTG E box can bind both heterodimers and homodimers (Krause et al., 1997; Harfe et al., 1998; Grove et al., 2009). Consistent with the previous work, we found that *C. elegans* HLH-2 can transactivate expression in a yeast system without any additional HLH partners via

CACCTG E boxes (Fig. S1). In contrast, CATATG E boxes bind only HLH-2 heterodimers (Krause et al., 1997; Sallee and Greenwald, 2015). We made a version of the 160bp enhancer region that includes sites 1 through 7, but has both E box 1 and E box 7 mutated to the CATATG heterodimer-only sequence (“Het” in Fig. 4D). A parallel construct changed both E boxes to CATTTG, which is not a known E box, and served as a negative control (red X’s in Fig. 4D). We observed very little expression with either the heterodimer-only construct or the negative control (Fig. 4D), indicating that HLH-2 regulation of *nhr-67* requires homodimers, as previously shown for the *lin-3* enhancer (Sallee and Greenwald, 2015). Because similar very low levels of GFP expression were observed with both the heterodimer only and the negative control constructs, the residual expression obtained with the heterodimer only construct does not appear to reflect a low level of HLH heterodimer activation.

We previously showed that expression of an *nhr-67::gfp* construct in the AC depends on *hlh-2* (Verghese et al., 2011). However, interpretation of this experiment was complicated by the inability to separate differentiation of the AC from regulation of expression—knockdown of *hlh-2* causes an AC differentiation defect which is upstream of *nhr-67* expression. We performed *hlh-2(RNAi)* on our *276bp::GFP* enhancer construct and found that while AC expression at L3 was significantly reduced, as expected given the AC differentiation defect caused by loss of *hlh-2*, we observed no significant effect on the earlier L2 pre-VU expression (Table 1). Because eliminating maternal *hlh-2* is lethal, we performed RNAi on newly-hatched L1 animals. Therefore, it is possible that the failure of *hlh-2(RNAi)* to eliminate L2 pre-VU expression reflects residual HLH-2 activity. The *hlh-2 (bx108)* mutation, which sharply compromises *hlh-2* function in the somatic gonad (Sallee and Greenwald, 2015), significantly reduced *276bp::GFP* in the AC and had a modest effect on expression in L2 pre-VU cells (Table 1). The *hlh-2(tm1768)* mutation, which has very mild effects on *hlh-2* function in the gonad (Schindler and Sherwood, 2011), had no effect on *276bp::GFP* expression, but did enhance the strength of *hlh-2(RNAi)*. Taken together, these results argue that reduction of *hlh-2* does not have as large an effect on 276bp enhancer expression as removal of the two E boxes. This disparity could simply reflect residual low levels of HLH-2 after knockdown or mutation, but the experiment also raises the possibility that other transcription factors may contribute to regulation of gonadal expression via the E boxes. When we examined other candidate bHLH regulatory factors that are predicted to bind the CACCTG sites and may be expressed in somatic gonad tissue (*hlh-3*, *-14*, *-19*, *hnd-1*, *lin-32*) we found no effect on expression from the *276bp::GFP* enhancer construct (Table 1). This result is consistent with the results of Sallee and Greenwald (2015), who found no role for 41 other bHLH genes in AC development, and no pre-VU or AC expression for 17 other bHLH genes.

It is possible that some of other bHLH transcription factors could substitute for HLH-2 homodimers, but only when HLH-2 is absent (thus accounting for the low level of expression we observed in *hlh-2* mutants and RNAi-treated animals). To test this possibility, we knocked down *hlh-2* function by RNAi in the presence of the weak *hlh-2(tm1768)* mutation, allowing for viability after L1, but with reduced pre-VU and AC function and examined the consequences of individually knocking-down *hlh-3*, *hlh-14*, *hlh-19*, *hnd-1*, and *lin-32* (Table 1). Reducing the function of these individual *hlh* genes did not lower

276bp::gfp expression beyond the level seen in *hlh-2* knock-down alone. Therefore, we did not observe evidence for substitution of function for any of these *hlh* genes. While it is possible that combinations of different HLH proteins can substitute for HLH-2 homodimers, *hlh-2* appears to be the primary, if not the sole, trans-acting factor that acts via sites 1 and 7 to regulate *nhr-67* expression in pre-VU and AC.

2.5 Site 2 is a candidate nuclear receptor binding site

The conserved site 2 has a sequence that matches a core nuclear hormone binding site: AGGTCU (on the complementary strand in Fig. 3). Site-directed mutagenesis of this site to a non-nuclear receptor binding sequence (indicated by red X in Fig. 4D) reduced *276bp::GFP* expression in the AC, but did not have a significant effect on earlier L2 pre-VU expression. Therefore, we considered the possibility that a nuclear receptor might contribute to AC expression. Consistent with this expectation, one-hybrid experiments in yeast showed that the nuclear receptors NHR-2 and SEX-1 are capable of binding site 2 (data not shown; J. Reece-Hoyes and M. Walhout, unpublished results). The *nhr-2* gene is not known to be expressed in the somatic gonad, but we found that *sex-1* is expressed broadly in the hermaphrodite gonad, including pre-VU cells and AC, indicating that it has temporal and spatial opportunity to regulate *nhr-67* expression (Fig. S2). However, knockdown of *sex-1* by RNAi (Table 1), did not result in a significant change in *276bp::GFP* expression in either AC or pre-VU cells. Knockdown of *nhr-67* itself also did not change *276bp::GFP* expression, indicating that site 2 is not a positive autoregulatory site. Therefore, neither SEX-1 nor NHR-67 appear to be involved in regulation via the site 2 element.

The nuclear receptor *nhr-25*, the ortholog of Ftz-F1 and SF1, has been shown to be expressed in the AC and is required for normal uterine morphogenesis (Gissendanner and Sluder, 2000; Hwang and Sternberg, 2004; Asahina et al., 2006). We found that knockdown of *nhr-25* resulted in a modest reduction in *276bp::GFP* expression in the AC (Table 1) that was similar to the modest reduction in AC expression resulting from site-directed mutagenesis of site 2. Knockdown of *nhr-25* also resulted in a decrease in GFP expression in the pre-VU cells of L2 larvae. Because mutation of site 2 did not result in a significant decrease in pre-VU expression, this result raises the possibility that effects of *nhr-25* knockdown on pre-VU cells may be indirect and not mediated via site 2. Nonetheless, these results suggest that NHR-25 may bind site 2 in the AC and contribute to activation of *nhr-67* expression.

2.6 At least one other transcription factor is required for uterine expression

In addition to the two E box sites (1 and 7) and the NR-binding site (2), the 276bp fragment contains four other conserved sites (3 through 6) that could contribute to expression of *nhr-67* in pre-VU cells and the AC (Fig. 3). These sites are clearly not sufficient, since GFP reporters that include them but lack the E boxes are not expressed at all in either cell type (Fig. 4D). Site 3 appears to be a compound binding site that includes a site that is the reverse complement of site 5 (ATCGAGAA) and a site that is identical to site 6 (ATTGCGY). Thus sites 3, 5, and 6 together may constitute binding sites for two pairs of distinct transcription factors. In order to test the requirement for sites 3 through 6, we created a GFP reporter construct that deleted all four sites. The deletion of these four sites resulted in a significant

loss of GFP expression, although the effect was not as large as removal of the E boxes (Fig. 4D). Therefore, one or more of the binding activities represented by the four internal binding sites is required for normal expression of *nhr-67* in the pre-VU cells and AC. Restoring a single copy of site 6 was sufficient to restore normal levels of pre-VU and AC expression, indicating that at least site 6, but presumably the summed site 3/site 6 binding activity, is the critical additional regulatory input required. In contrast, deletion of sites 3 through 5 had no detectable effect on pre-VU and AC expression, indicating that these sites are not individually or collectively required. A role for the proposed site 3/6 binding activity is also supported by the weak somatic gonad expression of the *pf2* deletion (Fig 4C), as this deletion retains both a single E box (site 7) and site 6. Therefore, the weaker *nhr-67(pf2)* homozygous phenotype (Verghese et al., 2011) may result from the low level of *nhr-67* expression in pre-VU cells and the AC made possible by the combination of site 6 and a single E box. Inclusion of the compound site 3 along with a single site 1 E box and site 2 NR site resulted in only very low levels of pre-VU and AC expression (Fig. 4D), supporting the notion that the site 3/6 activity normally operates in combination with a pair of E boxes.

2.7 A model for a recurring uterine precursor expression enhancer

The *nhr-67* gonadal expression pattern is strikingly similar to the expression patterns that have been described for three other genes: *egl-43*, *lin-3*, and *lag-2* [Wilkinson et al., 1994; Hwang and Sternberg, 2004; Hwang et al., 2007]. All are expressed in the four pre-VU cells during the late L2, shortly after these cells are born, and all are expressed at high levels in the AC during the L3, with much lower or absent levels in the VU cells. In each case two or more E boxes have been identified in upstream promoter or intronic regions for these genes, suggesting that regulation of all four genes in the pre-VU and AC depends on HLH-2. Furthermore, this particular expression pattern closely mirrors expression of *hlh-2* in pre-VU and AC, and the destabilization of HLH-2 protein in VU cells during L3 (Karp and Greenwald, 2003; Sallee and Greenwald, 2015). We aligned the relevant promoter regions using the UCSC Genome Browser utility and evaluated them for common elements with the 160bp *nhr-67* uterine precursor enhancer (Fig. 5; Fig. S3). As previously reported, all four genes contain at least one pair of the same evolutionarily-conserved CACCTG E box (Hwang and Sternberg, 2004; Hwang et al., 2007; Sallee and Greenwald, 2015). The E boxes are found in the varying orientations with respect to each other and from 46 bp (*lin-3*) to 415 bp (*egl-43*) apart in *C. elegans*. Distance between E box sites was not evolutionarily conserved for any enhancer region. The *lin-3* and *lag-2* promoters also include a third conserved CACCTG E box that is several kb distant from the pair of E boxes. The putative uterine enhancer regions can be found in the 5' flanking region of the distal promoter (*nhr-67* and *lag-2*) or the first intron of the transcription unit (*lin-3* and *egl-43*). While the *lag-2* and *nhr-67* were located in the distal 5' flanking regions of all species examined, the *lin-3* and *egl-43* sequences could be located in either the near 5' flanking region or first intron, depending on species (Sallee and Greenwald, 2015). Therefore, the E boxes appear to be able to regulate pre-VU and AC gene expression without respect to orientation or separation from each other or position relative to transcription start site, as expected for a classic enhancer element (Levine et al., 2014; Zabidi et al., 2015).

We also examined *lin-3*, *lag-2*, and *egl-43* for the presence of a NR (site 2)-like sequence and a site 3/6-like sequence (Fig. 5; Fig.S3). As previously reported (Hwang and Sternberg, 2004), the *lin-3* enhancer region has a conserved GGGTCA NR-binding site that can bind NHR-25 and is located between the two E boxes. The *egl-43* enhancer region has a conserved ACGTCU between the two E boxes, which could also possibly serve a NR-binding function. We did not identify any candidate NR-binding sites among the conserved elements in the *lag-2* enhancer region. The *egl-43* enhancer region has a single conserved site 3/6 sequence, ATTGCGY, 65bp upstream of the first E box, which is included in a promoter fragment that drives *egl-43* expression in the AC (Rimann and Hajnal, 2007). No conserved sequences matching this motif appear among those in the *lin-3* or *lag-2* enhancer elements. Our data are consistent with a model in which the pair of E boxes serves as the key feature for driving gene expression in the L2 pre-VU cells and L3 AC. Additional elements, which appear to vary from one enhancer to another, provide additional required activities and presumably contribute to robustness or regulation in other cells. The variable nature of the location and separation of these elements suggests that the enhancer regions could be examples of convergent evolution rather than whole-scale duplications of a single conserved regulatory cassette.

Tissue-specific enhancers were first described in the early 1980's (Gillies et al., 1983, Banerji et al., 1983; Mercola et al., 1983). Like these archetypal enhancers, the uterine precursor enhancer has large effects on expression rates in specific cells and can act from different locations relative to the transcription start site. The uterine precursor enhancer of the *nhr-67* promoter controls a tight developmental progression of expression in multiple precursor cell types, but appears to maintain expression in only one type of mature differentiated cell, the AC. The E boxes, primarily or exclusively binding HLH-2 homodimers, appear to be the key elements in driving expression in a group of initially equivalent stem cells (pre-VU), followed by high levels of expression in a single, differentiated descendant (the AC). The uterine precursor enhancer does require additional activities for high level expression (NHR-binding and other activities), but comparisons from one gene to another suggest that different elements may be recruited for specific enhancers to sum to the same overall regulatory activity by somewhat different combinations.

The conserved uterine precursor enhancer is a key element in a complex regulatory network governing development of the VU lineages and AC (Fig. 6). Evidence from this and previous studies places the transcription factors *hlh-2* and *fos-1* in primary roles in coordinating gene expression in multiple cell types. Accordingly, mutations in *hlh-2* have the most severe effect on VU lineages and the AC (Karp and Greenwald, 2003, 2004), while mutations in the next level of downstream regulators, such as *nhr-67* and *egl-43* cause later and more modest defects (Hwang et al., 2007, Rimann and Hajnal, 2007; Verghese et al., 2011; Matus et al., 2015). Mutations in effector genes, such as *egl-13* and *zmp-1*, cause more subtle defects (Sherwood et al., 2005; Oommen and Newman, 2007; Ghosh and Sternberg, 2014). Not long after the birth of the pre-VU cells in L2, the HLH-2 transcription factor activates expression of the transcription factors *nhr-67* and *egl-43*, and the signaling factors *lag-2/delta* and *lin-3/EGF receptor*. In late L2 and early L3 stages, as the AC differentiates and the VU cells begin dividing, destabilization and proteolysis of HLH-2 is a major event leading to a

progressive loss of transcription of *nhr-67*, *egl-43*, *lag-2* and *lin-3*. Additional unidentified transcription factors in the VU lineage must help maintain *lin-12/Notch* expression. Expression of VU lineage transcription factors, including *egl-13* and *lin-11* depend on the *nhr-67* and/or *egl-43* transcription factors, although this may reflect expression and function of the upstream regulators in L2 pre-VU cells, rather than the mature VU descendants. In the AC, a network of interdependent transcription factors, in particular HLH-2, FOS-1, EGL-43, and NHR-67, coordinate regulatory control of differentiation. This case illustrates how batteries of transcription factors can be recruited to distinct developmental events through differential employment of shared promoter elements at different stages within a group of related cell lineages.

3. Experimental Procedures

3.1 Genetics and Molecular Analysis

C. elegans were cultured and crosses performed as described by Brenner (1974) and Stiernagle (2006). Translational reporter constructs were created using the pPD95.75 plasmid backbone to fuse PCR-derived promoter fragments to GFP. The *pf* deletion series was created by amplifying equivalent fragments from genomic DNA obtained from homozygous deletion strains. The enhancer-trap constructs were created by using PCR amplification and cloning into the pPD122.53 plasmid backbone to fuse the promoter region to the *pes-10* minimal promoter element and GFP. Mutated versions were obtained either by PCR or by use of a Quick-Change II kit (Agilent). New transgenic strains were created by microinjection using a pRF4 co-transforming marker plasmid (Mello and Fire, 1995). Both experimental constructs and co-transforming marker plasmid DNA were injected consistently at 100 µg/ml. For all constructs a minimum of 2 independent extrachromosomal array lines were created and evaluated; 3–6 lines were typical. Strains were grown at 20°C and evaluated at room temperature. DNA sequences were recovered from Wormbase (www.wormbase.org) and aligned using ClustalW (Thompson et al., 1994) and the UCSC Genome Browser utility (genome.ucsc.edu).

3.2 RNAi Treatment

We obtained plasmids containing *C. elegans* cDNA inserts in vector L4440 flanked by inducible promoters from Dharmacon/GE Life Sciences or SourceBioScience. Experiments were performed either by picking L4 animals or placing synchronized embryos onto RNAi feeding lawns of HT115 bacteria on NGM-IA plates (Ahringer, 2006; Stiernagle, 2006). We performed controls in which each strain was also grown on L4440 control plasmid-transformed HT115 feeding RNAi bacteria on NGM-IA plates. Treated worm strains were grown at 20°C.

3.3 Microscopy

To examine and record GFP fluorescence patterns we used a Nikon UD Optiphot 2 microscope and captured images using a Nikon DS camera. We cropped images and adjusted for optimum contrast and brightness using Adobe Photoshop software. We evaluated statistical differences in expression of GFP by pairwise comparison with matched wild-type controls using a t-test with two-tailed distribution.

3.4 One-hybrid Analysis

We used the Clontech one-hybrid Matchmaker system to evaluate DNA-binding activity in yeast. We fused a full-length *hlh-2* cDNA clone to the GAL4 activation domain coding region. We transformed and integrated the pGAD424-derived activation plasmids and pLacZi-derived reporter plasmids into yeast strain YM 4271 using the Li-acetate procedure described by the manufacturer. We assayed for β -galactosidase activity using a quantitative ONPG spectrophotometric assay, and calculated β -Gal units using the method of Miller (1972): $1000 \times \text{OD}_{420}/\text{time of incubation} \times \text{concentration factor} \times \text{OD}_{600}$.

Strains and molecular clones used in this study are listed in the Key Resources Table.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The following former Muhlenberg College students made important experimental contributions to this study: Alex Breiding, Emily Brennan, Jessica Chu, Stephanie Eng, Catherine Ezzio, Evan Fletcher, Jenny Hall, Michael Kho, Mackenzie Kramer, Kimberly Lightfoot, Emily Lisco, Brittany Sanford, Peter Shartel, John Shocken, Zachary Shuler, and Katie Skwirut. Emily A. Bayer and Maria Sallee provided editorial comments on the manuscript. We are very appreciative of John Reece-Hoyes and Marian Walhout for sharing unpublished results, and Iva Greenwald, David Matus, and Maria Sallee for helpful advice. The *Caenorhabditis* Genetics Center, funded by the NIH Office of Research Infrastructure Programs (P40 OD010440), provided several nematode strains. The *tm1768* allele was generated by the National Bioresource Project of Japan. Michael Krause provided an *hlh-2* cDNA. Funding: This work was supported by NIGMS 1R15GM107799 grant to BW. Student researchers were supported by the Vaughan, Lake Road, and Crist Family research scholarships of Muhlenberg College.

References

- Ahringer J, (ed.), 2006 Reverse genetics, in: WormBook (ed.), The *C. elegans* Research Community, WormBook, doi:10.1895/wormbook.1.47.1.
- Asahina M, Valenta T, Silhankova M, Korinek V, Jindra M. Crosstalk between a nuclear receptor and beta-catenin signaling decides cell fates in the *C. elegans* somatic gonad. *Dev Cell*. 2006 8;11(2): 203–11. PubMed PMID: . [PubMed: 16890160]
- Banerji J, Olson L, Schaffner W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*. 1983 7;33(3):729–40. PubMed PMID: . [PubMed: 6409418]
- Benassayag C, Montero L, Colombie N, Gallant P, Cribbs D, Morello D: Human c-Myc isoforms differentially regulate cell growth and apoptosis in *Drosophila melanogaster*. *Mol Cell Biol*. 2005, 25:9897–909. PMID: PMCID: PMC1280252. [PubMed: 16260605]
- Brenner S, The genetics of *Caenorhabditis elegans*. *Genetics*. 1974 77: 71–94. [PubMed: 4366476]
- Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*. 2008, 134(1):25–36. PMID: . [PubMed: 18614008]
- Fernandes JS, Sternberg PW: The tailless ortholog *nhr-67* regulates patterning of gene expression and morphogenesis in the *C. elegans* vulva. *PLoS Genet*. 2007, 3(4):e69 PubMed PMID: ; PubMed Central PMCID: PMC1857733. [PubMed: 17465684]
- Ghosh S, Sternberg PW. Spatial and molecular cues for cell outgrowth during *C. elegans* uterine development. *Dev Biol*. 2014 12 1;396(1):121–35. doi:10.1016/j.ydbio.2014.09.028. Epub 2014 Oct 2. PubMed PMID: . [PubMed: 25281934]
- Gillies SD, Morrison SL, Oi VT, Tonegawa S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*. 1983 7;33(3): 717–28. PubMed PMID: . [PubMed: 6409417]

- Gissendanner CR, Sluder AE. *nhr-25*, the *Caenorhabditis elegans* ortholog of *ftz-f1*, is required for epidermal and somatic gonad development. *Dev Biol.* 2000 5 1;221(1):259–72. PubMed PMID: . [PubMed: 10772806]
- Greenwald IS, Sternberg PW, Horvitz HR: The *lin-12* locus specifies cell fates in *Caenorhabditis elegans*. *Cell* 1983, 34: 435–444. PMID: . [PubMed: 6616618]
- Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJ. A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. *Cell.* 2009 7 23;138(2):314–27. PubMed PMID: ; PubMed Central PMCID: PMC2774807. [PubMed: 19632181]
- Harfe BD, Vaz Gomes A, Kenyon C, Liu J, Krause M, Fire A. Analysis of a *Caenorhabditis elegans* Twist homolog identifies conserved and divergent aspects of mesodermal patterning. *Genes Dev.* 1998 8 15;12(16):2623–35. PubMed PMID: ; PubMed Central PMCID: PMC317087. [PubMed: 9716413]
- Hunter CP, Kenyon C: Specification of anteroposterior cell fates in *Caenorhabditis elegans* by *Drosophila* Hox proteins. *Nature* 1995, 377:229–32. PMID: . [PubMed: 7675107]
- Hwang BJ, Meruelo AD, Sternberg PW. *C. elegans* EVI1 proto-oncogene, EGL-43, is necessary for Notch-mediated cell fate specification and regulates cell invasion. *Development.* 2007 2;134(4): 669–79. Epub 2007 Jan 10. PubMed PMID: . [PubMed: 17215301]
- Hwang BJ, Sternberg PW. A cell-specific enhancer that specifies *lin-3* expression in the *C. elegans* anchor cell for vulval development. *Development.* 2004 1;131(1):143–51. Epub 2003 Dec 3. PubMed PMID: . [PubMed: 14660442]
- Karp X, Greenwald I: Multiple roles for the E/Daughterless ortholog HLH-2 during *C. elegans* gonadogenesis. *Dev. Biol* 2004, 272: 460–469. PMID: . [PubMed: 15282161]
- Karp X, Greenwald I: Post-transcriptional regulation of the E/Daughterless ortholog HLH-2, negative feedback, and birth order bias during the AC/VU decision in *C. elegans*. *Genes Dev.* 2003, 17: 3100–3111. PMID: PMCID: PMC305261. [PubMed: 14701877]
- Kimble J, Hirsh D: The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Dev. Biol* 1979, 70: 396–417. PMID: . [PubMed: 478167]
- Kimble J, Simpson P: The LIN-12/Notch signaling pathway and its regulation. *Ann Rev Cell Dev Biol* 1997, 13: 333–61. PMID: . [PubMed: 9442877]
- Kopan R, Ilagan MX: The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell.* 2009, 137(2):216–33. PMID: PMCID: PMC2827930. [PubMed: 19379690]
- Krause M, Park M, Zhang JM, Yuan J, Harfe B, Xu SQ, Greenwald I, Cole M, Paterson B, Fire A. A *C. elegans* E/Daughterless bHLH protein marks neuronal but not striated muscle development. *Development.* 1997 6;124(11):2179–89. PubMed PMID: . [PubMed: 9187144]
- Krause M, Techniques for analyzing transcription and translation. *Methods Cell Biol.* 1995 48:513–29. Review. PubMed PMID: . [PubMed: 8531740]
- Levine M, Cattoglio C, Tjian R. Looping back to leap forward: transcription enters a new era. *Cell.* 2014 3 27;157(1):13–25. doi: 10.1016/j.cell.2014.02.009. Review. PubMed PMID: ; PubMed Central PMCID: PMC4059561. [PubMed: 24679523]
- Maher B ENCODE: The human encyclopaedia. *Nature.* 2012, 489(7414):46–8. PMID: . [PubMed: 22962707]
- Marri S, Gupta BP. Dissection of *lin-11* enhancer regions in *Caenorhabditis elegans* and other nematodes. *Dev Biol.* 2009 1 15;325(2):402–11. doi:10.1016/j.ydbio.2008.09.026. Epub 2008 Oct 7. PubMed PMID: . [PubMed: 18950616]
- Massari ME, Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol.* 2000, 20(2):429–40. PubMed PMID: ; PubMed Central PMCID: PMC85097. [PubMed: 10611221]
- Matus DQ, Li XY, Durbin S, Agarwal D, Chi Q, Weiss SJ, Sherwood DR. In vivo identification of regulators of cell invasion across basement membranes. *Sci Signal.* 2010 5 4;3(120):ra35. doi: 10.1126/scisignal.2000654. PubMed PMID: ; PubMed Central PMCID: PMC3917318. [PubMed: 20442418]
- Matus DQ, Lohmer LL, Kelley LC, Schindler AJ, Kohrman AQ, Barkoulas M, Zhang W, Chi Q, Sherwood DR. Invasive Cell Fate Requires G1 Cell-Cycle Arrest and Histone Deacetylase-

Mediated Changes in Gene Expression. *Dev Cell*. 2015 10 26;35(2):162–74. PMID: . [PubMed: 26506306]

McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, Wenger AM, Bejerano G, Kingsley DM. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature*. 2011, 471(7337):216–9. PMID: PubMed Central PMCID: PMC3071156. [PubMed: 21390129]

Mello C, Fire A, DNA transformation. *Methods Cell Biol*. 1995 48: 451–482. [PubMed: 8531738]

Mercola M, Wang XF, Olsen J, Calame K. Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. *Science*. 1983 8 12;221(4611):663–5. PubMed PMID: . [PubMed: 6306772]

Miller JH: Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press; 1972.

Newman AP, Sternberg PW: Coordinated morphogenesis of epithelia during development of the *Caenorhabditis elegans* uterine-vulval connection. *Proc Natl Acad Sci USA* 1996, 93: 9329–33. PMID: . [PubMed: 8790329]

Newman AP, White JG, Sternberg PW: Morphogenesis of the *C. elegans* hermaphrodite uterus. *Development* 1996, 122: 3617–26. PMID: . [PubMed: 8951077]

Newman AP, White JG, Sternberg PW: The *Caenorhabditis elegans* lin-12 gene mediates induction of uterine specialization by the anchor cell. *Development* 1995, 121: 263–71. PMID: . [PubMed: 7768171]

Oommen KS, Newman AP. Co-regulation by Notch and Fos is required for cell fate specification of intermediate precursors during *C. elegans* uterine development. *Development*. 2007 11;134(22): 3999–4009. Epub 2007 Oct 17. PubMed PMID: . [PubMed: 17942488]

Rimann I, Hajnal A. Regulation of anchor cell invasion and uterine cell fates by the egl-43 Evi-1 proto-oncogene in *Caenorhabditis elegans*. *Dev Biol*. 2007 8 1;308(1):187–95. Epub 2007 May 25. PubMed PMID: . [PubMed: 17573066]

Sallee MD, Greenwald I. Dimerization-driven degradation of *C. elegans* and human E proteins. *Genes Dev*. 2015 7 1;29(13):1356–61. doi: 10.1101/gad.261917.115. PubMed PMID: ; PubMed Central PMCID: PMC4511211. [PubMed: 26159995]

Sallee MD, Littleford HE, Greenwald I. A bHLH Code for Sexually Dimorphic Form and Function of the *C. elegans* Somatic Gonad. *Curr Biol*. 2017 6 19;27(12):1853–1860.e5. doi: 10.1016/j.cub.2017.05.059. Epub 2017 Jun 8. PubMed PMID: ; PubMed Central PMCID: PMC5531766. [PubMed: 28602651]

Shindler AJ, Sherwood DR. The transcription factor HLH-2/E/Daughterless regulates anchor cell invasion across basement membrane in *C. elegans*. *Dev Biol*. 2011 9 15;357(2):380–91. doi: 10.1016/j.ydbio.2011.07.012. Epub 2011 Jul 18. PubMed PMID: ; PubMed Central PMCID: PMC3164387. [PubMed: 21784067]

Sherwood DR, Butler JA, Kramer JM, Sternberg PW. FOS-1 promotes basement-membrane removal during anchor-cell invasion in *C. elegans*. *Cell*. 2005 6 17;121(6):951–62. PubMed PMID: . [PubMed: 15960981]

Stiernagle T, 2006 Maintenance of *C. elegans* In: *WormBook* (Ed.), The *C. elegans* Research Community. *WormBook*. doi:10.1895/wormbook.1.101.1.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 1994;22(22):4673–4680. [PubMed: 7984417]

Timmons L, Court DL, Fire A, 2001 Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* 263, 103–112. [PubMed: 11223248]

Verghese E, Schocken J, Jacob S, Wimer AM, Royce R, Nesmith JE, Baer GM, Clever S, McCain E, Lakowski B, Wightman B: The tailless ortholog nhr-67 functions in the development of the *C. elegans* ventral uterus. *Dev Biol*. 2011, 356(2):516–28. PMID: . [PubMed: 21718694]

Wilkinson HA, Fitzgerald K, Greenwald I. Reciprocal changes in expression of the receptor lin-12 and its ligand lag-2 prior to commitment in a *C. elegans* cell fate decision. *Cell*. 1994 12 30;79(7): 1187–98. PubMed PMID: . [PubMed: 8001154]

Zabidi MA, Arnold CD, Scherhuber K, Pagani M, Rath M, Frank O, Stark A. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*. 2015 26;518(7540):556–9. doi: 10.1038/nature13994. Epub 2014 Dec 15. PubMed PMID: . [PubMed: 25517091]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Outline of early uterine development and organization of the *nhr-67* gene

A) Schematic of early ventral uterine development in the L2 and L3 stages. The developing gonad is outlined and the positions of ventral uterine precursor cells are shown in a ventral view. Low level expression of *nhr-67* is represented by stippled cells identifying the pre-VU (at mid L2) and VU cells (at late L2). High level expression of *nhr-67* in the AC of late L2 and L3 is represented by dark shading. The 12 proliferative descendants of the VU stem cells in the L3 stage are shown as unfilled circles—these cells have very low or no detectable *nhr-67* expression. The L3 VU descendants continue to divide in the L4 stage and ultimately produce 36 cells that constitute the ventral body of the uterus. B) Schematic of the *nhr-67* gene with conserved promoter regions and mutations annotated. Hatched conserved blocks identify potential functional promoter regions consisting of clustered multiple short, conserved apparent cis-acting elements. The location of a predicted, but functionally undefined non-coding RNA region in intron 4 is also annotated.

Figure 2. Knockdown of *nhr-67* fails to suppress a *lin-12(gf)* mutation

The number of AC in L3 hermaphrodites was evaluated using an *syIs50* transgene that expresses *cdh-3::gfp*, which is expressed in differentiated AC's and additional cells. The function of *nhr-67* was knocked down by feeding RNAi. The homozygous *lin-12(n137gf)* mutation causes conversion of the presumptive AC into a VU due to activated *Notch* signaling. A) Percentages of L3 animals that display 1 or 2 AC. Although *nhr-67(RNAi)* treatment resulted in a high penetrance 2 AC phenotype and high penetrance visible protruding vulva and egg-laying defective phenotypes (data not shown), *nhr-67(RNAi)* treatment of *lin-12(n137gf)* animals failed to suppress the AC specification defect. The *syIs50* transgene does not reveal the three or more defective AC's in *nhr-67(RNAi)* animals that was reported by Matus et al. (2015) at this stage. B) Fluorescence microscopy images of GFP expression in representative L3 animals.

Figure 3. Sequence alignment of 276bp enhancer region

Aligned sequences from *Caenorhabditis elegans*, *C. brenneri*, *C. remanei*, and *C. briggsae*. Genomic DNA sequences aligned using ClustalW (Thompson et al., 1994) and the UCSC Genome Browser utility (genome.ucsc.edu). Boxed and bolded sequence identify conserved regions of 6bp or longer that are conserved among all four species. Yellow shading identifies sequences that are identical among all four species. The breakpoints of the *pf2*, *pf159*, and *pf88* promoter deletions are shown above the alignment. The E boxes and NHR binding sites are annotated below the alignment. The 3/5 and 3/6 proposed binding activities are annotated by blue and red lines over the alignment.

Figure 4. Reporter expression in uterine precursor cells
GFP expression as visualized by fluorescence microscopy of transgenic animals carrying reporter constructs. (A) Schematic of *nhf-67* reporter construct strategy. The WTprom::GFP construct serves as the basis for translational reporters to test deletions of the endogenous *nhf-67* promoter shown in part C. The complete *nhf-67* expression pattern requires additional introns (Fernandes and Sternberg, 2007), but the region shown is sufficient for pre-VU and AC expression. The 276bp::gfp and 160bp::gfp constructs serves as the basis for transcriptional reporters to study enhancer requirements shown in part D. (B) Representative examples of pre-VU expression in an L2 hermaphrodite (top pair) and AC expression in an

L3 hermaphrodite (bottom pair) carrying the *276bp::gfp* reporter fusion. Top panels of each pair show GFP fluorescence, bottom panels of each pair show same image overlaid atop bright field. (C) Translational reporters using the *nhr-67* proximal promoter. The pProx control construct includes only 500bp of 5' flanking *nhr-67* proximal promoter sequences. The remaining constructs include up to 2.5 kb of upstream promoter sequences. The various deletion constructs recreate exactly the three *in vivo* promoter mutations that have been functionally characterized. (D) Transcriptional reporters that fuse the 276 bp or 160 bp enhancer regions to GFP via the *pes-10* minimal promoter, which does not support pre-VU and AC expression alone (Krause, 1995). Schematic highlights the 8 numbered sites in the 276 bp region. The site 1 and site 7 E boxes are shaded in black. Red X's indicate the sequence has been mutated to a non-consensus sequence that is not expected to bind the predicted trans-acting protein. "HET" designated point mutations in the E boxes that change the sites recognized by HLH heterodimers but not HLH-2 homodimers. Columns indicated the average number of pre-VU cells per animal that express detectable GFP during the L2 stage and the percentage of animals that show detectable GFP in the AC during the L3 stage. Parenthetical numbers indicate total number of larval animals examined for each construct. Data shown are a compilation of multiple experiments using up to three independent transgenes per construct. Asterisks indicate data points that are significantly different than matched wild-type controls at $p < 0.01$ as judged by two-tailed t test.

Figure 5. Organization of uterine precursor enhancers

Schematic shows the overall organization of enhancer elements in the *C. elegans* genome. *nhr-67* and *lag-2* are located on chromosome V. *egl-43* is located on chromosome II. *lin-3* is located on chromosome IV. The *nhr-67* and *lag-2* enhancers are located in the distal 5' flanking region of the gene; the *lin-3* and *egl-43* enhancers are located in the first intron. Shaded boxes indicate evolutionarily-conserved regions in each promoter region that may constitute binding sites for transcription factors: black- E boxes for HLH-2, green- NHR binding sequence, yellow- unidentified 3/6 site from *nhr-67* enhancer, gray- other conserved sites. The slightly-diverged potential NHR conserved site in the *egl-43* enhancer is indicated by stippled green. Coordinates of these regions in different species are annotated in Fig. S2.

Figure 6. Regulatory model for specification of ventral uterine cells

A model of gene regulatory pathways for pre-VU, VU and AC development incorporating data from this study and previous work (Hwang and Sternberg, 2004; Hwang et al., 2007; Rimann and Hajnal, 2007; Oommen and Newman, 2007; Marri and Gupta, 2009; Verghese et al., 2011; Matus et al., 2015; Sallee and Greenwald, 2015). Solid arrows indicate direct positive regulation. Upper-case identifies demonstrated or proposed protein activities (i.e., HLH-2 protein is believed to directly regulate expression of target genes shown). Dashed arrows indicate presumed indirect positive regulation. Upregulation of *nhr-67* in the AC is symbolized by larger bold font. In VU lineages destabilization of degradation of HLH-2 protein (symbolized by slash and smaller outline font) leads to downregulation of *nhr-67*, *lag-2*, *lin-3* and *egl-43* (outline font). Because *lin-12* expression is maintained in VU lineages, transcription factors in addition to NHR-67 must positively regulate *lin-12*. The *egl-13* and *lin-11* transcription factors are expressed in descendants of the VU cells.

TABLE 1

Knockdown effects on expression of 276bp::GFP

Stage	L2		L3		
	# pre-VU	St Dev	N	%AC	N
wild-type	2.0	1.5	29	74%	39
<i>hlh-2 (RNAi)</i>	1.9	1.5	33	41% [*]	27
<i>hlh-2 (bx108)</i>	0.9 [*]	1.3	31	23% [*]	26
<i>hlh-3 (RNAi)</i>	2.5	1.5	43	80%	35
<i>hlh-14(RNAi)</i>	2.5	1.5	39	74%	27
<i>hlh-19(RNAi)</i>	2.5	1.6	46	81%	27
<i>hnd-1 (RNAi)</i>	2.8	1.5	38	76%	41
<i>lin-32 (RNAi)</i>	2.5	1.5	41	59%	27
<i>sex-1 (RNAi)</i>	2.0	1.5	27	71%	21
<i>nhr-25 (RNAi)</i>	1.0 [*]	1.2	23	51% [*]	32
<i>nhr-67 (RNAi)</i>	2.1	1.5	25	80% ^a	46
<i>hlh-2 (tm1768)</i>	1.9	1.3	35	ND	
<i>hlh-2 (tm1768); hlh-2(RNAi)^b</i>	0.8 [*]	1.1	22	ND	
<i>hlh-2 (tm1768); hlh-3(RNAi)^b</i>	0.8 [^]	1.0	22	ND	
<i>hlh-2 (tm1768); hlh-14(RNAi)^b</i>	1.1 [^]	1.2	27	ND	
<i>hlh-2 (tm1768); hlh-19(RNAi)^b</i>	1.3 [^]	1.6	21	ND	
<i>hlh-2 (tm1768); hnd-1(RNAi)^b</i>	0.9 [^]	1.3	22	ND	
<i>hlh-2 (tm1768); lin-32(RNAi)^b</i>	1.3 [^]	1.5	21	ND	

Average number of pre-VU cells that express GFP at L2 per animal and the percentage of L3 animals that display GFP in the AC. N indicates the number of larval animals examined at each stage.

^{*} Significantly different from wild-type controls at p<0.05 by two-tailed t test.

[^] Not significantly different from *hlh-2 (tm1768); hlh-2(RNAi)* by two-tailed t test.

^a *nhr-67* knockdown results in two AC's that express GFP in some animals.

^b *hlh-2(tm1768) hlh-2(RNAi)* background knocks down *hlh-2* in L1 and L2.

ND = Not determined.

KEY RESOURCES TABLE

Reagent or resource	Source	Identifier
Bacterial and Virus Strains		
HT115(DE3) E. coli for feeding RNAi	Caenorhabditis Genetics Center	HT115(DE3)
Experimental Models: Organisms/Strains		
wild-type <i>Caenorhabditis elegans</i>	Caenorhabditis Genetics Center	N2
<i>hlh-2(bx108); him-5(e1490); lin-32(e1926)</i>	"	EM 496
<i>hlh-2(tm1768)</i>	"	JK4099
<i>lin-12(n137gf)</i>	"	MT688
<i>nhr-67(pf2), nhr-67(pf88), nhr-67(pf159)</i>	Verghese et al., 2011	LA275, LA691, MU1217
<i>syIs50 [cdh-3::gfp], lin-12(n137gf); syIs50</i>	Caenorhabditis Genetics Center, this paper	PS3352, MU1299
Transgenic reporter strains:	This paper	
wild-type <i>nhr-67::gfp</i>	"	
<i>bwEx92</i>		MU1303
<i>bwIs6</i>		MU1268
<i>nhr-67(pf2)::gfp</i>	"	
<i>bwEx137, bwEx138, bwEx139</i>		MU1370, MU1371, MU1372
<i>nhr-67(pf159)::gfp</i>	"	
<i>bwEx140, bwEx141</i>		MU1374, MU1375
<i>nhr-67(pf88)::gfp</i>	"	
<i>bwEx93, bwEx94, bwEx95, bwEx96</i>		MU1305, MU1306, MU1307, MU1308
wild-type 276bp::gfp	"	
<i>bwEx91</i>		MU1302
<i>bwEx132</i>		MU1355
sites 2-3-4-5-6::gfp	"	
<i>bwEx112, bwEx113, bwEx114, bwEx115</i>		MU1324, MU1325, MU1326, MU1327
sites 2-3-4-5-6::gfp	"	

Reagent or resource	Source	Identifier
<i>bwEx154</i> , <i>bwEx155</i>		MU1400, MU1401
sites 1-2-6-7-8::gfp	"	
<i>bwEx178</i> , <i>bwEx179</i>		MU1502, MU1503
sites 1-2-7-8::gfp	"	
<i>bwEx180</i> , <i>bwEx181</i> , <i>bwEx182</i>		MU1504, MU1505, MU1506
sites 1-2-3::gfp	"	
<i>bwEx189</i> , <i>bwEx190</i> , <i>bwEx191</i> , <i>bwEx192</i>		MU1545, MU1546, MU1547, MU1548
sites 1-7::gfp	"	
<i>bwEx156</i> , <i>bwEx157</i> , <i>bwEx158</i> , <i>bwEx159</i>		MU1402, MU1403, MU1404, MU1405
sites 1-7-1::gfp	"	
<i>bwEx183</i> , <i>bwEx184</i> , <i>bwEx184</i>		
sites 1-3-4-5-6-7-8::gfp	"	MU1518, MU1519
<i>bwEx129</i> , <i>bwEx130</i> , <i>bwEx131</i>		MU1351, MU1352, MU1353
sites 1-2-3-4-5-6-7::gfp	"	
<i>bwEx166</i> , <i>bwEx167</i>		MU1473, MU1474
sites 1het-2-3-4-5-6-7het::gfp	"	
<i>bwEx168</i> , <i>bwEx169</i> , <i>bwEx170</i> , <i>bwEx171</i>		MU1476, MU1477, MU1478, MU1479
sites 1X-2-3-4-5-6-7X::gfp	"	
<i>bwEx172</i> , <i>bwEx173</i> , <i>bwEx174</i> , <i>bwEx175</i>		MU1493, MU1494, MU1495, MU1496
sites 1-140bp separation-7::gfp	"	
<i>bwEx200</i> , <i>bwEx201</i> , <i>bwEx202</i> , <i>bwEx203</i>		MU1565, MU1566, MU1567, MU1568
Oligonucleotides		
for amplification of large <i>nhr-67</i> promoter fragment	ThermoFisher	
CAGCGGCCGCCGAAATCTAACGCCGCTCTGAATGAA		OC0858
CAGCGGCCGCCGCCCTGATTGATTTGAGA		OC0833
for amplification of 276bp region	"	

Reagent or resource	Source	Identifier
ACCATTTTCGGCTAAAGTTTTC		ON67PROM51
AATTCCTCAAAATGCTCATCA		ON67PROM31
for amplification of site 1/7 deletions	"	
ACCCCTCGTCTGTCTATCTCTC		ON67MINE52
TTCATTTTGATGCTTACGCCA		ON67MINE32
for site-directed mutagenesis of site 2	"	
GAGAACGATGCAGGACTGCAGTGAAGAGAGATAGACAGACGAG		ODSITE2F51
GAGAACGATGCAGGACTGCAGTGAAGAGAGATAGACAGACGAG		ODSITE2R31
for site 7 alone construct	"	
AGCTGAAATGCAGGTGTTC		ON67SITE6SH
CTAGTGAAACACCTGCAITTC		ON67SITE6AX
for site 1-7 alone construct	"	
AGCTACACCTGTATGGAAATGCAGGTGTTC		ON67SITE16SH
CTAGTGAAACACCTGCAITTCATACAGGTGT		ON67SITE16AX
for 160bp constructs and E box mutagenesis	"	
ACAAGCTTACACCTGTATGACCCCTCGTCTGTCTA		O160WTF1
CATCTAGAGAACACCTGCAITTCATTTTGATGCTT		O160WTR1
ACAAGCTTACATATGATGACCCCTCGTCTGTCTA		O160HETF1
CATCTAGAGAACATATGATTTCAITTTTGATGCTT		O160HETR1
ACAAGCTTACATTTTATGACCCCTCGTCTGTCTA		O160NEGF1
CATCTAGAGAACAATTTCAITTTCAITTTTGATGCTT		O160NEGR1
for triple E box construct	"	
AGCTACACCTGTATGACCGAAATGCAGGTGTCAGAAATTCACACCTGTA		ON67SITE161S
CTAGTACAGGTGTGAATTTCTGAACACCTGCATTTCCGGTCAATACAGGTGT		ON67SITE161A
for 1-2-3 construct	"	
AGCTCAAAAGTTTCACACCTGTATGACCCCTCGGAATTCATCTCTCTCAACGACCTTTTCCCTGCATCGTTCTCGATAATTCGCGTA		ON6712351
CTAGTAGGCAATATCGAGAACGATGCAGGAAAGGTGTTGAAGAGAGATAGAAATCCGAGGGTTCATACAGGTGTGAACCTTTTG		ON6712331
for 1-150bp-7 construct	"	
ACAAGCTTCACCTGTGAAGAGGGGTAGAACATTAATAACG		O16STUF5

Reagent or resource	Source	Identifier
ACTCTAGAGAAACACCTGCA ^{CCAGAA} CAAAATCGAACCAA		O16STUF3
Recombinant DNA		
pPD95.75	Addgene	Plasmid Cat#1494
pPD122.53	Addgene	Plasmid Cat#1631
Matchmaker one-hybrid plasmids and reagents	Clontech	Cat# K1603-1
<i>hlh-2</i> RNAi feeder plasmid in L4440 backbone	GE Health Sciences	M05B5.5
<i>hlh-3</i> RNAi feeder plasmid in L4440 backbone	SourceBioScience	T24B8.6
<i>hlh-14</i> RNAi feeder plasmid in L4440 backbone	SourceBioScience	C18A3.8
<i>hlh-19</i> RNAi feeder plasmid in L4440 backbone	GE Health Sciences	F57C12.3
<i>hnd-1</i> RNAi feeder plasmid in L4440 backbone	GE Health Sciences	C44C10.8
<i>lin-32</i> RNAi feeder plasmid in L4440 backbone	GE Health Sciences	T14F9.3
<i>sex-1</i> RNAi feeder plasmid in L4440 backbone	GE Health Sciences	F44A6.2
<i>nhr-25</i> RNAi feeder plasmid in L4440 backbone	SourceBioScience	F11C1.6
<i>nhr-67</i> RNAi feeder plasmid in L4440 backbone	GE Health Sciences	C08F8.8
<i>hlh-3</i> RNAi feeder plasmid in L4440 backbone	SourceBioScience	T24B8.6
<i>hlh-14</i> RNAi feeder plasmid in L4440 backbone	SourceBioScience	C18A3.8