THE CATHOLIC UNIVERSITY OF AMERICA

Investigating the control of *hlh-8* gene expression

A DISSERTATION

Submitted to the Faculty of the

Department of Biology

School of Arts and Sciences

Of The Catholic University of America

In Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

By

Stephany Gale Meyers

Washington, D.C.

2010

Investigating the control of *hlh-8* gene expression

Stephany Gale Meyers, Ph.D.

Director: Ann K. Corsi, Ph.D.

The *hlh-8* gene encodes a basic helix-loop-helix transcription factor called Twist that is involved in mesoderm development and the morphogenesis of enteric and sex muscles in *C. elegans*. Twist binds to the canonical binding site CANNTG referred to as an E box. The *hlh-8* gene is composed of 5 exons with a 2 kb intron after the first exon. Because the function of *hlh-8* is dependent on the gene being expressed in the appropriate cell type, it is important to understand how the expression of the *hlh-8* gene is regulated. Typically, elements in the promoter of a gene determine the regulation of gene expression in different tissues. However, only elements that control expression in a subset of Twist containing tissues have been discovered in the promoter. Therefore, additional elements must exist elsewhere and a hypothesis was developed that regulatory elements are present in the large first intron. This study explores the *hlh-8* intron using a construct that contains a basal promoter and can be activated to express GFP in a variety of tissues by juxtaposition to a tissue-specific enhancer. Constructs containing portions of intron 1 reveal two regions, with a single E box each, which are sufficient to drive expression of *gfp* in a subset of tissues that express Twist. Furthermore, expression of *gfp* is lost when both E boxes are disrupted. My hypothesis is that these E boxes are important for *hlh-8* autoregulation. Results from expressing these reporters in *hlh-8* null mutants and examining Twist binding to the E boxes by *in vitro* gel shift analysis support this

hypothesis. Additionally, this study characterizes a mutant with a large deletion (646 bp) in the first intron of *hlh-8*. Altogether, results from this study lead to an understanding of tissue-specific regulation of *hlh-8*. Intron elements appear to control expression in differentiated tissues, whereas it has been shown previously that other factors regulate expression in undifferentiated cells. Moreover, since there is homology between *C. elegans* and human Twist proteins, understanding the regulation of *hlh-8* will elucidate the control of expression for the gene that encodes for the human Twist protein.

This dissertation by Stephany Gale Meyers fulfills the dissertation requirement for the doctoral degree in Cell and Microbial Biology approved by Ann K. Corsi, Ph.D., as Director, and by John E. Golin, Ph.D., and J. Michael Mullins, Ph.D. as Readers.

Ann K. Corsi, Ph.D., Director

John E. Golin, Ph.D., Reader

J. Michael Mullins, Ph.D., Reader

This work is dedicated to

My Family: Marty Meyers, Sheri Sorensen, Clifford Meyers and Marlena Muffoletto, for giving me the true understanding and appreciation of perseverance and whose unwavering belief in me made this work possible.

To

Patricia Heileman, Mike Jansen and Tricia Jansen,
whose endless encouragement and support inspired me to achieve more than I anticipated.

And

To my fiancé, Douglas Small, and my friends,
whose beautiful friendship, love and prayers have allowed me to get through difficult times and brought me joy throughout this process.

TABLE OF CONTENTS

Preliminaries:

Text:

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A/P: anterior/posterior

bp: base pair

bHLH: basic helix-loop-helix

bwm: body wall muscle

°C: degrees Celsius

Cat: catalog

cc: coelomocyte

CeE/DA: *C. elegans* E/Daughterless

Con: constipated

Egl: egg-laying deficient

EMSA: Electrophoretic Mobility Shift Assay

EMT: epithilial-mesenchymal transition

GFP: green fluorescent protein

HIF: hypoxia-inducible factors

hlh-8 (-): *hlh-8 (nr2061)*

hlh-8 (iΔ): *hlh-8 (tm726)*

hmc: head mesodermal cell

HREs: hypoxia-response elements

IDC: invasive ductal carcinoma

ILC: invasive lobular carcinomas

L1: first larval stage

LIST OF ABBREVIATIONS (CONT.)

L4: fourth larval stage

kb: kilobase

mg: milligram

ml: milliliter

mM: millimolar

mu anal: anal depressor

mu ints: intestinal muscles

mu sph: anal sphincter

mut: mutant/mutation

PCR: polymerase chain reaction

Pvl: protruding vulva

RNAi: RNA interference

rpm: rotations per minute

RT-PCR: Reverse transcriptase polymerase chain reaction

SDM: Site-Directed mutagenesis

Semi-Egl: semi-egg-laying deficient

SM: Sex Myoblast

Spu: extreme protruding vulva

µl: microliter

um: uterine muscle

vm: vulval muscle

ACKNOWLEDGEMENTS

The author would like to acknowledge the following:

Dr. Ann K. Corsi: for hiring me as a laboratory technician and for fostering my scientific development with her constant support and generous opportunities of education and research independence.

Drs. Michael Krause and Andy Golden: for valuable suggestions and advice and for providing a surplus of reagents and equipment.

Dr. Tetsu Fukushige and Tommy Brodigan: for the vast amount of sequencing performed for me.

Drs. David Eisenmann and Javier Wagmaister: for the pKKMCS *gfp* vector and associated material.

Drs. John E. Golin and J. Michael Mullins: for participation as my committee members and for the guidance in my research and dissertation.

Dr. John E. Golin: for the opportunities and freedom he provided me as the Genetics TA.

Peng Wang: for his efficient microinjections.

Peng Wang, Jie Zhao, Mary Philogene: for being great, supportive and intelligent lab mates when I entered the Corsi Lab as a naïve technician.

Baltimore-Washington Worm Club and NIH Data Club: for the incredible support through reagents, helpful advice and suggestions.

INTRODUCTION

Transcription factors are intrinsic players in a multitude of developmental processes. The regulation of key transcription factors is central to proper development. Twist is a basic helix-loop-helix (bHLH) transcription factor that is essential in mesoderm development. The importance of understanding the regulation of this dynamic transcription factor is seen in human disease. Misregulation of Twist has been implicated in the role of cancer metastasis and mutations in the gene that encodes for Twist cause a congenital craniosynostotic disorder (Wilkie, 1997; Yang et al., 2004). Craniosynostosis is a human developmental disorder that is characterized by premature closure of the cranial sutures (Wilkie, 1997). A better understanding of the regulation and role that Twist plays during development will provide important insights into the defects associated with these diseases.

The Twist pathway is conserved from a microscopic non-parasitic nematode, *Caenorhabditis elegans,* to humans (Wang et al., 2006). Furthermore, there is one homolog of Twist in *C. elegans* named CeTwist. Mutations in this protein cause the animals to become constipated (Con) and egg-laying deficient (Egl) due to CeTwist's important role in the development of enteric and sex muscles, respectively (Harfe et al., 1998b; Corsi et al., 2000). The study of the regulation of CeTwist, serves as a guide for understanding mesoderm development and the role of human Twist in craniosynostotic disorders and tumor metastasis.

bHLH factors in *C. elegans* and other organisms

Twist is a basic helix-loop-helix (bHLH) transcription factor. These factors are proteins that are required to regulate gene transcription. Transcription factors can either promote or inhibit expression of a gene. bHLH proteins regulate their target genes by the basic domain of two monomers binding to DNA at the canonical site of CANNTG. This binding site is referred to as an E box. The helix-loop-helix domain is important for the formation of a dimer of Twist to itself as a homodimer or to a partner as a heterodimer. The known binding partner for CeTwist is CeE/Daughterless (CeE/DA) (Fig. 1). Furthermore, the bHLH domain in CeTwist is 59%-63% identical to Twist in other species (Harfe et al., 1998b). Due to this conservation of the bHLH domains, the target sequences and dimer partners are conserved between humans and *C. elegans*. In humans, Twist acts as both a repressor and an activator to regulate target gene expression. However, to date in *C. elegans*, CeTwist is shown only to have activator function. Therefore, transcriptional activity and corresponding sequences outside of the bHLH domain are not conserved between *C. elegans* and humans. On the other hand, there is currently limited information regarding the human gene regulation. Therefore, it remains to be seen how this is conserved between *C. elegans* and humans, until the regulation in both species is better understood. This study addresses *C. elegans* Twist regulation in hopes that there will be similarities with the human gene regulation.

Twist was first identified in *Drosophila melanogaster* where knockout mutations lead to embryonic lethality due to the complete lack of mesoderm (Simpson, 1983; Thisse et al., 1987). Many orthologs have also been identified in *Mus musculus* with a wide

variety of complex functions (Wilson-Rawls et al., 2004). CeTwist is the only homolog in *C. elegans* and a presumptive null mutation results in viable animals. The combined simplicity of one homolog that is viable when mutated and the advantages of the model organism discussed below, makes *C. elegans* an excellent model to use to study Twist regulation (Harfe et al., 1998b; Corsi et al., 2000). Therefore, understanding the regulation of CeTwist may elucidate the control of expression for the gene that encodes for the human Twist protein.

The implication of Twist in human developmental disease

Understanding the proper control of Twist gene expression is important because certain human diseases are caused by alterations in Twist level. Decreased expression of Twist results in craniosynostotic disorders (Wilkie, 1997). Craniosynostosis is a human developmental disorder that is characterized by premature closure of the cranial sutures leading to an abnormal skull shape (Fig. 2). This phenotype occurs in about one out of every 2,500 births (Wilkie, 1997). Mutations in Twist result in a specific congenital craniosynostotic disorder called Saethre-Chotzen syndrome. Twist-loss-of-function mutations lead to this syndrome that is an autosomal dominant disorder. Hence, the proper amount of Twist protein is critical to proper development in an individual. Along with craniosynostosis, most commonly of the coronal suture, this syndrome is identified by other cranial feature defects such as low hairline, facial asymmetry, small ears and ptosis, or drooping eyes. Saethre-Chotzen syndrome is also linked to limb abnormalities

that include brachydactyly, or shortness of digits, and cutaneous syndactyly, which is the joining of two adjacent digits (Ratisoontorn and Cunningham, 2007).

The developmental phenotypes seen in those patients with Twist mutations are due to the protein's role as a critical embryonic morphogenesis regulator. Twist is essential for the induction of cell migration and tissue reorganization in embryogenesis (Yang et al., 2004). In *Drosophila,* Twist acts during gastrulation by inducing the ventral cells to migrate and subsequently create the mesoderm layer (Leptin and Grunewald, 1990). In mammals, Twist is necessary for neural tube morphogenesis via accurate cell migration (Chen and Behringer, 1995). The migration of cells and formation of mesoderm across species is due to a Twist-dependant process called epithelial-mesenchymal transition (EMT) (Yang et al., 2004). In this process, epithelial cells lose cell polarity and cell adhesion, and a reorganization of the cytoskeleton occurs. These dramatic morphological changes lead to fibroblast-like mesenchymal cells that are able to migrate (Yang and Weinberg, 2008). One feature of EMT is the loss of epithelial markers, such as E-cadherin, α-catenin and other adherens junction proteins, and the gain of mesenchymal markers, including N-cadherin and fibronectin (Yang et al., 2004). Hence, the regulation of Twist gene expression is important to elucidate because proper Twist levels are essential for controlling mesoderm development through EMT.

Twist's role in the metastasis of tumors

Not only is Twist down-regulation associated with human disease, but up-regulation is as well. Up-regulation of Twist has been shown to have a role in multiple

cancer types including melanoma, T-cell lymphoma, rhabdomyosarcoma, gastric carcinoma, head and neck squamous cell carcinoma, breast cancer, prostate cancer, and pancreatic cancer (Rosivatz et al., 2002; Yang et al., 2004; Kwok et al., 2005; Alexander et al., 2006; Satoh et al., 2008; Yang et al., 2008). In many of these cancers, Twist has been shown to play an important role in metastasis. Metastasis of a tumor is one of the hallmarks that correspond to the worst prognosis and the most difficulty in treatment of the cancer. There are four distinct steps in the metastatic process of a primary tumor. First, invasion is characterized by the loss of cell-cell contact and gain of cell mobility. This is followed by the cancer cells entering the circulatory system in the process of embolism. Third, the cells leave the circulatory system to enter a new tissue environment; this step is called extravasation. Lastly, some cells that survive the previous steps must proliferate to induce secondary tumors at the new site (Fidler, 2003; Fig. 3). Most studies to date have shown that Twist's primary role in the process is to cause tumors to start the first step of metastasis. Twist's essential function in the metastatic process is shown by the presence of increased Twist expression in breast cancer cells that are undergoing all steps of metastasis. However, Twist is not expressed in non-metastatic tumors. Furthermore, in a mouse model, down-regulation of Twist via siRNA treatment of potential-metastatic cells results in the lack of metastasis of the tumor (Yang et al., 2004).

Interestingly, the developmental function of Twist in the EMT process is conserved in the metastatic progression of cancer (Yang et al., 2004; Yang and Weinberg, 2008). Specifically, ectopic expression of Twist in kidney epithelial cells and in human

mammary epithelial cells, causes a loss of cell-cell adhesion and cell polarity. The morphological change in the cells is moreover seen by a loss of epithelial cell markers, such as E-cadherin, coupled with a gain of N-cadherin and other mesenchymal markers (Yang et al., 2004). A similar correlation of the loss of epithelial characteristics and gain of mesenchymal features in the presence of an increase of Twist has been seen in human gastric cancers and prostate cancer (Rosivatz et al., 2002; Kwok et al., 2005).

Additionally, the parallel of increased Twist expression with a decrease in E-cadherin has been shown in metastatic cancers compared to non-metastatic cancers. Twist and E-cadherin expression was analyzed in different samples of invasive human breast tumors (Yang et al., 2004). Invasive lobular carcinomas (ILC) are clinically more likely to metastasize than invasive ductal carcinomas (IDC) (Fackler et al., 2003). Tumor sample analysis showed that Twist is highly expressed in the majority of ILC. Inversely, the expression of E-cadherin is dramatically diminished to almost a complete loss in ILC compared to IDC samples (Yang et al., 2004). Furthermore, it has been shown that the Twist promoter is much less methylated in ILC than in IDC (Fackler et al., 2003). Thus, suggesting that Twist activity through its promoter is higher in metastatic breast cancer and emphasizing the importance of thoroughly understanding the transcription of this gene.

The strong correlation between Twist, E-cadherin and N-cadherin in the process of EMT and the metastasis of tumors suggests the regulation of E-cadherin and N-cadherin is dependent on Twist. This is indeed the case. Twist inhibits E-cadherin through three E boxes found in the promoter region of the gene that encodes for E-

cadherin; however, it is unclear if this regulation is direct or indirect (Yang et al., 2004; Yang and Weinberger, 2008). Furthermore, N-cadherin is shown to be directly activated by Twist through an E box found in the first intron of the *N-cadherin* gene (Alexander et al., 2006).

Recent data suggest that Twist's role in the progression of cancer is not strictly limited to EMT mechanisms. Twist is linked to protecting tumor cells against apoptosis. Evidence for this is not only due to the increased Twist levels found in the most aggressive type of prostate cancer, but also because up-regulation of Twist protects cells from undergoing apoptosis when treated with a common cancer therapeutic drug, taxol (Kwok et al., 2005). The molecular mechanisms have yet to be worked out, but experiments establish a correlation between Twist, p53 and HOXA5. p53 is an important tumor suppressor that controls cell cycle regulations and HOXA5 is a transcription factor that directly activates p53 expression. In a breast cancer tissue culture model, over-expression of Twist leads to inhibition of p53 mediated gene expression and can override the cell cycle check-point arrest. This suppression is implicated to be partly due to Twist directly binding to a p53 transactivator, HOXA5 (Stasinopoulos et al., 2005). Similarly, in neuroblastoma, Twist is shown to antagonize the function of a pro-apoptotic factor, called MYC, by down-regulation of the p53 pathway (Peinado et al., 2007).

The individual roles of Twist in development and cancer progression seem to be interrelated more than just on a mechanistic level. Interestingly, Sahlin and colleagues found a statistical link between an increased risk in developing breast cancer and women with Saethre-Chotzen Syndrome (Sahlin et al., 2007). The level of Twist has dramatic

consequences in both development and the progression of cancer. Thus, understanding how concentrations of Twist are controlled is critical to future implications of disease progression of both cancer and craniosynostotic disorders.

The regulation of human Twist

The regulation of the human Twist gene is not completely understood. However, the recent identification of Twist as an oncogene has permitted tissue culture and animal models to be exploited in a more direct way to study Twist's regulation.

Regulation of Twist as a metastatic inducer has been shown to be under the control of hypoxia-inducible factors (HIFs). Hypoxia is the condition of oxygen depletion. Tumors use the surrounding tissue's blood supply as an oxygen source. Hypoxia is a feature of tumors that have outgrown their host's blood supply. When tumors reach this stage, it correlates to reduced therapeutic achievement and increased metastatic potential (Gort et al., 2007). The HIFs complex is composed of α and β subunits, which are both constitutively expressed. However, under oxygen-rich conditions prolyl hydroxylases are expressed that hydroxylate HIFα. The hydroxylated HIFs are degraded through the targeting of a tumor suppressor factor. However, in hypoxia conditions, the non-hydroxylated HIFα subunit binds to the HIFβ subunit and then the transcription factor complex regulates target genes (Gort et al., 2007). Twist is the direct target of HIFs in both mouse and humans (Gort et al., 2007; Yang et al., 2008). HIFs bind to canonical DNA sequences called hypoxia-response elements (HREs) (Yang et al., 2008). HIF-2α binds to two HREs located in the only intron in the human gene that

encodes for Twist (Gort et al., 2007). Similarly, HIF-1α binds to the HRE located in the promoter region of the *Mus musculus* Twist (Yang et al., 2008).

A pancreatic cancer model suggests MSX2 is a direct regulator of Twist. MSX2 is a Hox gene and this family of genes is described below. Over-expression of MSX2 induces Twist expression. Furthermore, MSX2 is linked to inducing EMT through the expression of Twist (Satoh et al., 2008). MSX2 and Twist interact to control the differentiation and proliferation of the frontal bone skeletogenic mesenchyme in mice (Ishii et al., 2003). However, in humans, Twist is expressed in non-hypoxic situations and MSX2 is unlikely to be entirely responsible for Twist expression. Thus, work in a model organism may provide further information regarding the regulation of Twist.

C. elegans are advantageous model organisms

C. elegans were first introduced in 1974 as a model organism for scientific inquiry for numerous reasons (Brenner, 1974). The organisms are microscopic non-parasitic, free-dwelling roundworms that at full growth are about one millimeter long. They can easily be manipulated in the laboratory on Nematode Growth Media agar petri plates and eat a non-virulent strain of *E. coli* called OP50.

These animals are ideal for developmental studies due to their lifecycle and brood size. A single animal can produce an average of 300 to 1,000 progeny in its lifetime, depending on the mating strategy, and the lifecycle is a short three days at room temperature. *C. elegans* have an embryonic and a postembryonic developmental phase. Oocytes are fertilized and begin development within the mother. When the embryos

reach the 28-cell stage, they are laid out into the environment where they continue to develop for about nine hours before hatching. The embryonic phase of development is followed by four distinct larval stages where development continues for 38 hours until the nematode matures into a young adult (Fig. 4).

A genetic advantage of the nematodes is the two sexes of *C. elegans*: self-fertilizing hermaphrodites and males. This means a homozygous genetic population can easily be maintained from one single hermaphrodite animal. Also, another genotype can easily be incorporated into the population by mating a hermaphrodite to a male. Furthermore, the complete 97 Mb genome of *C. elegans* was sequenced in 1998 (*C. elegans* Sequencing Consortium, 1998). Interestingly, about 40% of genes associated with human disease are conserved in *C. elegans* (Culetto and Sattelle, 2000). Altogether, the organism's simple genetics provide for an excellent model to understand human disease and development.

Another advantage of this model system is that *C. elegans* are transparent; thus every cell in the live animal can be visualized under a microscope (Fig. 5). This is important to this study for two key reasons. First, mesoderm is found throughout the animal. Furthermore, the transparency of cells allows one to use Green Fluorescent Protein (GFP) in a dynamic way to look at mesodermal cells throughout development. GFP is used by two main approaches to study CeTwist's regulation. First, it can be used with tissue-specific promoters to mark a specific cell or cell-type. Second, GFP can be utilized as either a fusion protein or as a transcriptional construct in order to read-out gene expression and regulation (Chalfie et al., 1994). In particular, this study will

employ promoters plus enhancer elements to drive the expression of *gfp* in the location where the promoter is activated.

The transparency of *C. elegans* has allowed the entire cell lineage to be mapped (Sulston and Horvitz, 1977). The division of each cell is tightly controlled and this leads to a completely fixed cell lineage and somatic cell number. Hence, factors that control cell division and differentiation can easily be explored. The organisms are simplistic with 959 somatic cells in hermaphrodites and 1031 in males. This study will focus on the development and characterization of the hermaphrodite.

In conclusion, *C. elegans* are an exceptional model system to use for this study due to their ease of simple manipulations and upkeep, unique features, well-annotated fully-sequenced genome, and well-developed molecular tools such as GFP.

C. elegans mesoderm and CeTwist expression in the mesoderm

A further advantage to using *C. elegans* to study Twist is the organism's simple mesodermal development. Multicellular eukaryotic organisms have three germ layers in early embryogenesis. Ectoderm, the outermost germ layer, develops into the skin and the nervous system in mammals. Endoderm is the middlemost germ layer and differentiates into the lining of organs. Mesoderm is located between the other two layers and is the focus of this research. Mesoderm becomes a variety of tissues including bone, muscle, cartilage, and blood. However, *C. elegans* do not have many of the same tissues as mammals, instead in these animals the mesoderm differentiates into somatic gonad and non-gonadal tissues.

CeTwist is expressed in a subset of non-gonadal mesodermal derived tissues, including muscles. There are two types of muscles in *C. elegans*: striated and non-striated. Body wall muscles (bwm) make up the striated muscles and are responsible for the body movements of the animal. There are a total of 95 of these cells in a single organism. There are three different types of non-striated muscles in *C. elegans*: pharyngeal muscles, sex muscles and enteric muscles. The pharyngeal muscles are important for food uptake. The 16 sex muscles are composed of eight uterine muscles and eight vulval muscles (vm), which are responsible for proper deposit of embryos outside the animal. The enteric muscles are composed of four different muscles that are required for the defecation of waste from the animal: right and left intestinal muscles (mu ints), anal depressor (mu anal), and anal sphincter (mu sph). Of these muscle types, CeTwist is expressed in the sex muscles and all four of the enteric muscles (Harfe et al., 1998b).

Along with muscles, there are two other non-gonadal cell types that are derived from mesoderm in *C. elegans*. One of these, the head mesodermal cell (hmc), is found in the anterior portion of the animal and is of unknown function. The other cells are six coelomocytes that reside in the pseudocoelomic region of the animal and have a macrophage, scavenging function. CeTwist is expressed in two of the coelomocytes and in the hmc.

Differentiation of the tissues described above happens both embryonically and post-embryonically. Furthermore, CeTwist is expressed in cells that are born in both developmental periods and in non-lineally related cells. The enteric muscles and the hmc

are born in the embryo and all cells arise out of three distinct lineages (Zhao et al., 2007; Fig. 6). The other tissues that express CeTwist differentiate post embryonically but arise from one cell called the M mesoblast. This cell is born embryonically, divides and differentiates postembryonically into two coelomocytes, 14 body wall muscles, eight vms and eight uterine muscles (Sulston and Horvitz, 1977; Fig. 7). CeTwist is expressed in the M mesoblast cell and all undifferentiated cells that arise from this blast cell called the M lineage. Overall, the expression pattern of CeTwist is throughout the undifferentiated M lineage, the hmc, sex muscles, and enteric muscles.

CeTwist regulation in *C. elegans*

The gene that encodes for CeTwist, *hlh-8*, is made up of 5 exons and the first intron is 2 kb. The entire gene is 3.4 kb in length. The promoter of this gene has been well characterized and elements that control the expression of *hlh-8* in the coelomocytes and undifferentiated M lineage have been identified (Harfe *et al.*, 1998b). Furthermore, the expression of CeTwist in the M lineage is under the control of two Hox factors, LIN-39 and MAB-5, and a PBC homology cofactor, CEH-20, which bind to a site in the promoter (Liu and Fire, 2000). Hox genes are a class of genes that are important in the patterning and formation of the anterior/posterior (A/P) axis in many organisms. Typically, Hox genes reside in clusters and the A/P location of expression of individual genes corresponds to their chromosomal location. For example, the most 5' gene is expressed in the anterior portion of the animal and the most 3' gene corresponds to posterior expression (Gilbert, 2006). Due to the anterior, middle, and posterior location

of the hmc, vm, and enteric muscles, respectively, it is unlikely that the same Hox genes control CeTwist expression in all of these tissues.

As in humans, there is a link between hypoxia and CeTwist in *C. elegans*. In the nematode there is a single homolog for each of the major hypoxia components. HIFα is encoded by *hif-1*, HIFβ is encoded by *aha-1*, and there is a single gene, *egl-9*, that controls the prolyl hydroxylase activity. Therefore, there is constitutive HIF activity in *egl-9* mutants and *egl-9* mutant animals are Egl. Moreover, the *egl-9* mutant phenotype is suppressed in animals when *hif-1* is not expressed. Interestingly, knock-down of CeTwist via RNAi also suppresses the Egl phenotype of *egl-9* mutant animals (Gort et al., 2007). This suggests CeTwist, as in humans, is in the hypoxia pathway. Furthermore, there are numerous potential HRE sites in the promoter and introns of *hlh-8*. However, it is unlikely that HIFs are the sole regulators of CeTwist expression in the hmc, vms, and enteric muscles because expression of *hlh-8* is seen in these tissues under non-hypoxia situations. Therefore, investigation needs to be conducted to elucidate the regulation of CeTwist in the hmc, vms, and enteric muscles.

The aims of this study

This study focuses on additional regulation of *hlh-8* beyond control of expression in the undifferentiated M lineage by Hox genes. I identified two E boxes, E1(Tw) and E2(Tw), in the first intron of *hlh-8* that were necessary for expression of CeTwist in the vulval muscles, enteric muscles and the hmc. I show through the use of an *hlh-8*

presumptive null mutant and *in vitro* binding studies that CeTwist directly regulates its own expression through the E2(Tw) site.

I also characterized a mutant that is the result of a 646 nucleotide deletion in the first intron of *hlh-8,* called *hlh-8 (tm726)* [referred to here as *hlh-8 (iΔ)*]. The 5' deletion break point is nine nucleotides downstream from the E2(Tw) site and the 3' deletion break point is two nucleotides from exon 2, preserving the nucleotides that are required for the splicing (Aroian et al., 1993). I found through RT-PCR that the proximity to the splicing acceptor site of the deletion break point, leads to splicing defects in *hlh-8 (iΔ)* animals, overall decreasing the level of wild type CeTwist mRNA. The *hlh-8 (iΔ)* animals have an attenuated phenotype from the *hlh-8* presumptive null animals. *hlh-8 (iΔ)* animals are Con, can lay their embryos at a non-wild-type rate (Semi-Egl), and are able to activate some but not all downstream targets of CeTwist.

I propose a model that the promoter and Hox factors provide a basal level of CeTwist in specific tissues, for example in the undifferentiated M lineage. Autoregulation takes place once a threshold of CeTwist molecules is obtained in other tissues that require a higher level of CeTwist for cell function, for example in the vms that are derived from M lineage cells. Such autoregulation has been shown for other bHLH factors in mammals; however this is the first time it has been demonstrated for any gene in the Twist family.

MATERIALS AND METHODS

Construction of *gfp* transgenic lines

Reporter constructs were made from regions of the first intron of *hlh-8* that were amplified via Polymerase Chain Reaction (PCR) and inserted into the multiple cloning site of the *gfp* vector pKKMCS (gift from J. Wagmaister and D. Eisenmann). pKKMCS contains the basal promoter of the well-characterized gene *egl-18* and is not sufficient to cause expression of *gfp* in most tissues, with the exception of some intestinal cells. However, the addition of DNA that contains an enhancer element(s) to the vector allows for expression of *gfp* in a temporospatial pattern corresponding to the activity of the enhancer element(s) (Wagmaister et al., 2006).

The DNA inserts were constructed with wild-type *C. elegans* genomic DNA as the template. DNA was isolated from lysing the wild-type strain called N2. Single-worm lysis buffer was made following standard procedures and stored at -20°C (Plasterk, 1995). The lysis mixture was made by adding 2 µl of proteinase K (20 mg/ml) to 150 µl of single-worm lysis buffer. Five to ten N2 animals were added to 2.5 µl of the lysis mixture and put at -20°C for at least one hour followed by 95°C for 15 minutes and 60°C for one hour in a thermocycler. To avoid amplifying the OP50 bacteria in the background, nested PCR was employed. PCR reaction mixture and procedure was conducted according to the manufacturer's protocol (New England Biolabs Cat# M0267L). Nested PCR is a procedure of using primers corresponding to the flanking region of the desired DNA product in a preliminary PCR. The product from this reaction

is then used as a template in a second round of PCR with primers designed to complement internal sequences within the original product.

Primers against exon 1 and exon 3 were used as the outside primers in the first step of nested PCR. To assure the resulting template for the second round of PCR was not contaminated with other sequences of DNA, gel extraction was performed. Gel extraction was executed by electrophoresis of an agarose gel loaded with the preliminary PCR product. Once the DNA had migrated to the midpoint of the gel, a slit was made directly beneath the DNA of interest and filter paper backed with dialysis tubing (size: 10,000 molecular weight cutoff) was inserted into the slit. The PCR product was deposited onto the filter paper by running the gel for 20 minutes longer at 125 volts. The isolated DNA was extracted from the filter paper by using a punctured 0.5 ml microcentrifuge tube placed inside a 1.5 ml microcentrifuge tube that was centrifuged at 14,000 rpm for a few minutes in a microcentrifuge. Water was then added to bring the volume of the extract to 110 µl. To this, 60 µl of chloroform was added, vortexed and centrifuged at 14,000 rpm for 4 minutes. The top layer was then placed into a new tube and DNA precipitation was carried out to remove any additional contaminants. This purified section of wild-type genomic *hlh-8* DNA from exon 1 through exon 3 was diluted 1:50 and used as a template to create inserts for the pKKMCS vector.

For the second round of PCR, primers were designed to add *Sal*-I and *Bgl*-II restriction sites to the 5’ and 3’ flanking region of the specific *hlh-8* insert, respectively. The first construct used primers to amplify the entire first intron of *hlh-8*. Successively smaller overlapping portions of intron 1 were used as inserts to narrow down the region

of DNA that contained enhancer element(s) that were able to drive expression of *gfp*. A list of constructs with corresponding intron location and PCR primers used to create the inserts is found in Table 1 and 2, respectively.

Following amplification of the various DNA inserts, 500 ng of PCR product along with the pKKMCS vector was used in restriction digestion. The digestion reaction mixture and length of incubation followed the manufacturer's protocol (New England Biolabs Cat# R0144S and R0138S). Due to the close proximity of the *Sal*-I and *Bgl*-II restriction sites in the multiple cloning site of the pKKMCS vector, two single digestions were performed. Gel extraction was done following digestion to isolate properly cut pieces (see above for details). Ligation of the cut pKKMCS vector and inserts was conducted according to the manufacturer's protocol (New England Biolabs Cat# M0202S). Engineered constructs were transformed into ultracompetent *E. coli* (Stratagene Cat# 200315). A PCR-based screen was conducted of lysed individual colonies to locate the desired recombinant DNA. PCR primers were designed against the DNA flanking the multiple cloning site of the pKKMCS vector and used to identify inserts of the expected sizes (Table 3). Minipreps (Qiagen) of isolated colonies were used to sequence the junction of the cloning regions utilizing the same primers to verify the proper insert was ligated into the vector.

Once the reporter plasmids were made and verified, N2 animals were transformed with the *gfp* reporter constructs (100 µg/ml) and the transformation marker pRF4 plasmid (50 µg/ml) that encodes the dominant *rol-6 (su1006)* allele by standard microinjection techniques (Mello et al., 1991). The pRF4 marker allowed animals that contain the *gfp*

reporter plasmids to be identified by a rolling phenotype of the transgenic animals. At least two independent lines were isolated and minimally 30 animals per line were scored for each construct.

Site-directed mutagenesis of E1 and E2 E boxes

The construct pSM7(E1E2) was used for site-directed mutagenesis with mutant primers and the Quick Change Site-Directed Mutagenesis Kit (Stratagene Cat# 200516 and Table 4). The E box sequence of E1 and/or E2 was changed from CATCTG to AATCAG, but the remaining DNA of the plasmid remained unchanged. Constructs were sequenced to confirm that only the designated mutation was made. The mutated plasmids were injected into N2 animals following the above methods for making transgenic lines.

Homologous alignments of distantly related nematodes

Sequences were obtained and BLASTs were performed on WormBase (www.WormBase.org). BLASTs were conducted using the first intron of *hlh-8* from *C. briggsae, C. remanei*, *C. brenneri*, *C. japonica* against the *C. elegans* genomic database. ClustalW alignment of homologous regions was generated from http://www.ebi.ac.uk/Tools/clustalw2/index.html. Shading of the alignment was produced from BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html).

Analyzing the binding affinity of CeTwist dimers to E boxes through EMSA

Gel shift assays were performed to investigate the binding affinity of CeTwist homodimers and CeTwist/CeE/DA heterodimers to the intron E1 and E2 E boxes. CeTwist and CeE/DA from recombinant *E. coli* strains were purified previously in the laboratory as described in Zhao et al., 2007. Two sets of four pairs of 20mers were designed that contained the six nucleotide E box and seven flanking nucleotides. Probes corresponded to E1, E2, E2 (*arg-1*) as a positive control, and Control (Tw) as a negative control. One set of probes had the wild-type sequence and the other set had the same E boxes mutations as the reporters that were modified by site-directed mutagenesis (Table 5). One complementary strand of the probes was radiolabeled with γ-AT^{32}P and then the entire probe was incubated with bacterially expressed purified CeTwist and/or CeE/DA protein according to Harfe et al., 1998b. The input concentrations of proteins were determined previously from SDS-PAGE examination. The protein-probe mixture was run on a 6% native polyacrylamide gel (Invitrogen Cat# EC63655BOX) and visualized through autoradiography.

hlh-8 (-) strain crosses to investigate CeTwist autoregulation

A line containing the construct pSM10(E2a) was crossed into *hlh-8 (-)* animals. First, N2 males were crossed with transgenic pSM10(E2a) hermaphrodites. Then, 20 of the resulting transgenic males were mated with 30 *hlh-8 (-)* hermaphrodites. In the F1 generation, rolling heterozygous transgenic hermaphrodites were isolated and allowed to self-propagate. In the following F2 generation, Con, Egl, rolling transgenic

hermaphrodites were picked singly to allow for a new population of homozygous *hlh-8 (-)* containing pSM10(E2a) to propagate.

Knock down of CeE/DA by RNA interference

RNA interference (RNAi) feeding treatment of transgenic animals containing the pSM10(E1a) plasmid was carried out following modified protocols provided by A. Fire lab (Kamath et al., 2000). The *E. coli* strain called HT115 was used because it lacks the ability to degrade RNA. HT115 was transformed with an empty vector, L4440, or the *hlh-2* RNAi vector (gifts from A. Golden). The strains were cultured overnight with LB media containing 12.5 µg/ml tetracycline and 100 µg/ml ampicillin at 37°C shaking at 250 rpm. Nematode Growth Media agar plates containing 100 µg/ml ampicillin and 0.35 mM IPTG were used to transfer a small amount of cultured HT115. *hlh-2* dsRNA expression in the HT115 strain was induced by room temperature incubation for one day. CeE/DA is required early in embryogenesis and to circumvent this requirement, L1s were treated with *hlh-2* dsRNA (Kamath et al., 2000). A synchronous population of animals was obtained by taking advantage of the durable outer shell that surrounds the developing embryo. A population of mixed-stage wild-type animals was treated with a solution of 1 mM NaOH and 10% pure bleach (Lewis and Fleming, 1995). Following this hypochlorite treatment, only the eggs protected by their shells were able to survive. Embryos were allowed to hatch on a plate lacking food to cause developmental halt at the L1 stage. This L1 synchronized population was fed either the *E. coli* strain HT115 expressing *hlh-2* dsRNA or containing an empty L4440 vector. Animals were moved

every 24-hour period to a new RNAi feeding plate. Adult animals containing the pSM10(E2a) construct were scored for *gfp* expression and associated *hlh-2* RNAi phenotypes.

C. elegans strains and maintenance

Animals were maintained according to standard conditions and techniques (Brenner, 1974). Investigations were done at 20°C unless otherwise noted. Three *C. elegans* strains were used in this study: N2 wild type, *hlh-8 (nr2061)* referred to in this document as *hlh-8 (-)* (Corsi et al., 2000), and *hlh-8 (tm726)* denoted here as *hlh-8 (iΔ)*.

The *hlh-8 (iΔ)* allele was isolated by and obtained from the National Bioresource Project of Japan. Upon arrival, the strain was backcrossed minimally eight times and the deletion was confirmed through PCR amplification followed by sequencing. A list of primers used for sequencing is found on Table 6. Primers used in confirming the homozygous mutation were AC3, AC22, and AC61. AC61 is located in the region deleted in *hlh-8 (iΔ)* animals. Thus, a wild-type copy of the gene will amplify a product size of 456 bp with primers AC61 and AC3 and the *hlh-8 (iΔ)* allele will amplify a product size of 839 bp with primers AC22 and AC3. To circumvent the wild-type locus from making a larger product from primers AC22 and AC3, the thermocycler was set to have a 72°C amplification time of one minute which was not long enough for the 1485 bp product to be amplified (Table 7).

Previously integrated *gfp* reporter constructs were introduced into *hlh-8 (iΔ)* animals by standard genetic mating. PCR was used to confirm that the *hlh-8 (iΔ)* allele

was homozygous (see above for details) and outcrossed to N2 animals to confirm the *gfp* was homozygous. A minimum of 20 individual hermaphrodites were mated to N2 males in each strain that did not express *gfp* in the *hlh-8 (iΔ)* background. The following *gfp* reporter lines were used in this study: *arg-1::gfp ccIs4443(II)* (Corsi et al., 2002) is expressed in the hmc, vm1 and vm2 vulval muscles, uterine muscles and the four enteric muscles; *egl-15::gfp ayIs2(IV)* (Harfe et al., 1998a) is expressed in the vm1 vulval muscles; from the regulatory DNA of *ceh-24, NdEbox::gfp ccIs4656(IV)* (Harfe and Fire, 1998) is expressed in vm1 and vm2 vulval muscles*;*, *hlh-8::gfp ayIs7(IV)* (Harfe et al., 1998b) shows *gfp* expression in the M mesoblast cell, the 16 myoblast descendants, the sex myoblasts and descendants; *ccIs4438 [hlh-8::gfp] (IV)* (Yanowitz et al., 2004) is expressed in all 6 coelomocytes.

Observation of *C. elegans* animals

General manipulation of animals such as mating, picking and observation of brood size, life span and egg-laying rate were conducted with the use of a stereomicroscope. General outward phenotypes, such as constipation, egg-laying defects, sterility, and protruding vulvas were scored also with the use of a stereomicroscope. Observations of integrated *gfp* lines were accomplished with a fluorescent stereomicroscope to score the GFP pattern in free-moving animals. Visualization of GFP patterns of non-integrated lines were performed with a fluorescent compound microscope.

Brood size experiments were carried out on individual animals. Specifically, L4 animals were picked individually to a plate. Each consecutive day the animal was moved to a new plate and the embryos that were laid on the plate were counted. This was continued until the mother ceased to lay progeny. At the same time, the phenotype of the individual animals was analyzed.

To gain an accurate expression period of the *egl-15::gfp* reporter, animals were scored every hour to two hours for GFP expression. Concurrently, the exact developmental age of the animal was determined by observing the morphology of the developing vulva. This allowed for accurate assessment of when the expression of *egl-15::gfp* began. Once the animals became adults, they were scored a few times throughout the day to determine when the *egl-15::gfp* was no longer expressed. In tandem with GFP observation, the animals were also scored for constipation, protruding vulva, sterility, and egg-laying defects. Animals were scored for the M lineage division defects using the *hlh-8::gfp ayIs7(IV)* reporter and the same methodology as described above for *egl-15::gfp*.

Reverse Transcription PCR and splice product cloning

The RNA of mixed-stage populations of *hlh-8 (iΔ)* and N2 animals was isolated as described from Wang and colleagues with modifications (Wang et al., 2006). Glass beads (Sigma) and Trizol Reagent (Invitrogen Cat# 15596-018) were used to extract total RNA from the animals. M-MuLV Reverse Transcriptase (New England Biolabs Cat# M0253S) was used with a poly-A primer to make cDNA-mRNA hybrids from equal

quantities of total RNA from the two populations following the manufacturer's protocol. The Reverse Transcriptase products were used in a standard nested PCR reaction with primers against exon 1 and exon 5 of *hlh-8*. Actin primers were also used for total mRNA quantity control (Table 8).

Spliced products were individually extracted from an agarose gel and subjected to TA-cloning using vector pCR®2.1 (Invitrogen Cat# K2020). Isolated cDNA clones containing spliced products were sequenced to identify the location of splicing in aberrant and wild-type products (Table 8).

RESULTS

Evidence from the literature regarding regulatory elements in intron 1 of *hlh-8*

Elements that control the expression of *hlh-8* in undifferentiated cells of the M lineage and in coelomocytes have previously been identified in the upstream promoter region of *hlh-8* (Harfe et al., 1998b). Expression data from a variety of *gfp* reporter constructs suggest that DNA in addition to the promoter region is necessary in *hlh-8::gfp* reporter constructs to show expression in vms, the hmc, and the four enteric muscles (Fig. 8). A transcriptional reporter (pBH47.01) that contains 8.3 kb of DNA from the upstream region of *hlh-8* was reported to express in the undifferentiated M lineage descendents and the coelomocytes, but not in the enteric muscles. Expression in the vms and hmc was not scored in this construct (Harfe et al., 1998b). Another transcriptional reporter (pBH47.70) that contained 1.8 kb of the promoter is expressed in only the M lineage, not in the enteric muscles, the vms, the hmc, nor in the coelomocytes. A translational construct (pBH47.08) that contains the 9.4 kb upstream region, the entire genomic sequence of *hlh-8*, and the 3.6 kb downstream region was reported to be expressed in the undifferentiated M lineage descendents, the coelomocytes, and also in enteric muscles; vms and hmc expression was not scored in this construct (Harfe et al., 1998b). Furthermore, a construct (pAK95) that contains the 2.7 kb upstream region, exon 1, intron 1, and a portion of exon 2 of *hlh-8* was expressed in the enteric muscles, the hmc, and the vms (Peng Wang, unpublished). This construct contains the elements that are responsible for expression of CeTwist in the M lineage, but there was no expression of

gfp in these cells. This result is unlikely to be due to positional effects of the reporter gene *in vivo* because multiple lines were observed. However, I report the pAK95 result because it is consistent with the other reports in that elements located in regions other than the promoter are controlling expression of CeTwist in the enteric muscles, the hmc, and the vms. Therefore, regulatory elements controlling *hlh-8* expression in vms, the hmc, and the four enteric muscles are likely to be found in exon 1, intron 1, or a portion of exon 2.

Other evidence for regulatory elements in the first intron of *hlh-8* comes from rescue experiments. A presumptive null allele of CeTwist, *hlh-8 (nr2061)*, referred to as *hlh-8 (-)* in this dissertation, contains a large 1267 nucleotide deletion that removes 95% of the DNA responsible for coding the helix-loop-helix domain (Corsi et al., 2000). The *hlh-8 (-)* animals are Egl and Con due to the improper development of the vms and enteric muscles, respectively (Corsi et al., 2000). Rescue assays were performed by injecting the mutant with a plasmid that contained either *hlh-8* cDNA or genomic DNA and then scoring for rescue of Con and Egl phenotypes in stable lines. The cDNA construct (pAC1) contains the *hlh-8* cDNA, plus the 500 base pair *hlh-8* upstream region and the *unc-54* 3' UTR. The *hlh-8* genomic DNA plasmid (pBH64) contains the 9 kb upstream, the entire open reading frame, and the 3 kb downstream of *hlh-8*. The *hlh-8* cDNA in the *hlh-8 (-)* background rescued the Con and Egl phenotypes in 20% and 61% of the animals, respectively (n=100). On the other hand, *hlh-8 (-)* animals receiving the *hlh-8* genomic DNA were rescued for the Con and Egl phenotypes in 78% and 89% of the animals, respectively (n=75) (Corsi et al., 2000). Therefore, a reasonable assumption

can be made that constructs containing genomic DNA rescue *hlh-8 (-)* more completely, in particular the Con phenotype, than those containing the cDNA. The rescue data are consistent with the *gfp* reporter data in that they suggest intronic regions, specifically intron 1, of *hlh-8* may contain regulatory element(s) to control expression of CeTwist in a subset of differentiated tissues. However, these data do not exclude the possibility that the extra DNA in the larger *gfp* reporter constructs contains repressor element(s) or that important element(s) in exon 1, in conjunction with DNA greater than 500 nucleotides upstream, are important. Therefore, further study of the *hlh-8* intron 1 was required to determine whether this DNA was important for gene expression.

Intron 1 sequences of *hlh-8* control expression in a subset of differentiated mesodermal tissue

To test whether intron 1 controlled any of the *hlh-8* expression, I used a plasmid, pKKMCS, containing an *egl-15::gfp* minimal promoter that could drive basal *gfp* expression. pKKMCS can be activated in tissues by juxtaposition to a tissue-specific enhancer element. Thus, this reporter will express *gfp* in a temporal and spatial orientation according to the activity of the element (Wagmaister et al., 2006). The first construct contained the entire 2 kb *hlh-8* intron 1. The transgenic animals had expression of *hlh-8::gfp* in the hmc, the vms, and the four enteric muscles (Fig. 9; Fig. 10A). As expected, this construct did not express in the M lineage or in the coelomocytes. This lack of expression is due to the lack of previously isolated elements in the promoter

region of *hlh-8* that control for M lineage and coelomocyte expression (Harfe et al., 1998b).

Next, to investigate individual intron elements that control *hlh-8* expression, it was important to isolate a smaller region of intron 1 that was sufficient to drive expression in specified mesodermal tissues. To accomplish this, a series of increasingly smaller, overlapping construct pairs was used to isolate a minimal enhancer region (Fig. 10B). Each time expression activity was attributed to a fragment, it was divided into two smaller pieces. A 503 nucleotide fragment that expressed *gfp* in all analyzed tissues was identified (pSM7(E1E2); Fig. 10B). There are a total of 8 E-boxes in the first intron of *hlh-8* (Fig. 10A). pSM7(E1E2) contained two of these E-boxes that have been named E1 and E2. pSM7(E1E2) was divided into two constructs that contained approximately 300 nucleotides each, pSM9(E1a) and pSM10(E2a). pSM9(E1a) contains E1 and had expression in a subset of the enteric muscles. pSM10(E2a) contains E2 and was strongly expressed in all scored tissues (Fig. 10B). Constructs made from smaller portions of pSM7(E1E2) revealed two regions, both 163 nucleotides long, which were sufficient to drive *gfp* in a subset of tissues where *hlh-8* is expressed (pSM14(E1b) and pSM15(E2b); Fig. 10C). Furthermore, pSM14(E1b) and pSM15(E2b) contain E1 and E2, respectively (Fig. 10C). In conclusion, reporter constructs that contain either E1 or E2 were able to express *hlh-8::gfp* and constructs that contained E2 alone expressed *gfp* in more tissues and at a higher frequency of animals than those that contained E1 alone (Fig. 10B, C). Importantly, constructs containing other E boxes from intron 1 did not express in any of

these tissues, emphasizing the importance of these 2 E boxes (pSM3 and pSM4(*iΔ*1); Fig. 10B).

An interesting allele of *hlh*-8, whose phenotype is described below, contains a deletion of 646 nucleotides of the 3' region of intron 1 (Fig. 10A). To predict the expression for the *hlh-8 (tm726)* locus, referred to here as *hlh-8 (iΔ)*, additional *gfp* reporters were examined. The DNA that is deleted in *hlh-8 (iΔ)* animals was not sufficient to drive *gfp* expression (pSM4(*iΔ*1); Fig. 10B). The 5' deletion breakpoint of the *hlh-8 (iΔ)* is nine base pairs away from E2. A modified construct of pSM10(E2a) was made that removed the nucleotides that would be missing in the *hlh-8 (iΔ)* locus (pSM20(*iΔ*2); Fig. 10D). Interestingly, the DNA that is adjacent to E2 but is absent in *hlh-8 (iΔ)* animals was necessary for strong expression in all tissues (compare pSM10(E2a) to pSM20(*iΔ*2) and pSM15(E2b); Fig. 10B-D). Therefore, it is unlikely that the DNA deleted in the *hlh-8 (iΔ)* locus contained any enhancer elements, but seems to be important in conferring strong expression with constructs that contained E2.

E1 and E2 E boxes regulate *hlh-8* expression

Next, I wished to examine the contributions of E1 and E2 to *hlh-8* expression. Site Directed Mutagenesis (SDM) was performed to mutate E1 and E2 in pSM7(E1E2). The E box sequences were changed from CATCTG to AATCAG. This sequence change is expected to eliminate the E box function (Karp and Greenwald, 2003). Mutating E1 did not affect the reporter from being expressed in all scored tissues (pSM24(E1mut); Fig. 10E). In fact, the expression level of *gfp* in animals that contain pSM24(E1mut) closely

resembles that of animals with a construct that contains only E2 (pSM10(E2a); Fig. 10B). Interestingly, the expression level increases in tissues when E1 is mutated, and so could represent repressor activity of E1 (compare pSM7(E1E2) to pSM24(E1mut); Fig. 10B, E). Furthermore, when E2 is mutated the only tissue that continues to express *gfp* is the anal sphincter in 11% of the animals (pSM25(E2mut); Fig. 10E). This expression pattern is similar to constructs that contain only E1, in that *gfp* is only expressed in a subset of enteric muscles (pSM9(E1a), pSM14(E1b); Fig. 10B, C). Furthermore, when both E1 and E2 are mutated, *gfp* expression is obliterated in all cell types (pSM26(E1E2mut); Fig. 10E). Therefore, SDM confirms that E1 and E2 are necessary for expression of *hlh-8* in the specific scored mesoderm tissues and that E2 is more important than E1 in this function.

Extensive homology of a portion of intron 1 between distantly related nematodes

To provide further evidence for the importance of the E1 and E2 elements, the sequence of intron 1 was examined. A nucleotide comparison analysis was performed to determine the degree of conservation between the first intron of *hlh-8* in *C. elegans* with those of four distantly related species: *C. brenneri, C. japonica, C. briggsae, and C. remanei*. Interestingly, there was a long stretch of intron homology of 470 nucleotides (1128-1597) in *C. elegans* with the other nematodes. When the *C. elegans* sequence was compared with each individual species, *C. elegans* and *C. brenneri* had 74% identity; *C. elegans* and *C. briggsae* had 72% identity; *C. elegans* and *C. japonica* had 57% identity; and *C. elegans* and *C. remanei* had 74% identity. However, when all five nematodes

were aligned there was a 32% nucleotide identity between the sequences. In support of the importance of the two intron E boxes, E1 and E2 were perfectly conserved in all five distantly related species (Fig. 12). Furthermore, the last 124 nucleotides of the homologous region are deleted in the *hlh-8* mutant, *hlh-8 (iΔ)* discussed below.

CeTwist and CeE/DA proteins bind to E1 and E2 E boxes *in vitro*

Since it was established that E1 and E2 were necessary for *hlh-8::gfp* expression in a subset of tissues (Fig. 10E), the next step was to determine which bHLH protein(s) were binding to the elements. An *in vitro* Electrophoretic Mobility Shift Assay (EMSA) was used to ask if CeTwist and/or its known binding partner, CeE/DA, are able to bind to E1 and E2. Radiolabeled 20mers containing single E boxes were incubated with purified, bacterially expressed, recombinant CeTwist and/or CeE/DA and run on a native gel. In addition to E1 and E2, two additional E boxes were tested in this assay (Fig. 13A, B). The first is an E box that does not confer expression when placed into the pKKMCS expression vector (pSM22; Fig. 10F). This E box, referred to as Control (Tw), is found at nucleotide location 475 in intron 1 and was used as a negative control. E2 (*arg-1*) is an E box that is found in the promoter region of a known downstream target of CeTwist, *arg-1* and was used as a positive control in this experiment. E2 (*arg-1*) has been shown to be required for *arg-1* expression and is bound by both CeTwist and CeTwist plus CeE/DA *in vitro* (Zhao et al., 2007). Furthermore, E1 (Tw), E2 (Tw) and E2 (*arg-1*) all have the same E box sequence, CATCTG (Fig. 13B). Interestingly, CeTwist homodimers and CeTwist/CeE/DA heterodimers bound with greater affinity to E2 (Tw) than E1 (Tw) (Fig.

13C). Specifically, using a phosphorimager it was found that CeTwist homodimers bind to 4.5 times more E2 (Tw) than E1 (Tw) and 3.3 times more E2 (*arg-1*) than E1 (Tw). CeTwist/CeE/DA heterodimers bound 3.6 fold more E2 (Tw) and E2 (*arg-1*) than E1 (Tw). Also, CeTwist/CeE/DA heterodimers preferentially bound more to the probes than did CeTwist homodimers when increasing amounts of CeE/DA protein were added. Furthermore, the proteins were not able to bind to the 20mer that corresponds to Control (Tw) E box, nor to mutant probes that had the E box sequence changed to AANNAG (Fig. 13C). Thus, the *in vitro* gel shift data correspond with the expression data from the minimal promoter constructs in that E2 had greater affinity for CeTwist and CeE/DA in EMSA and led to a broader tissue expression in reporter constructs when compared to results with E1 (Fig. 10; Fig. 13C). Furthermore, when the E box sites were mutated, there was no *gfp* expression in the scored tissues, and these mutated sites did not bind CeTwist and/or CeE/DA *in vitro* (Fig. 10E; Fig. 13C).

hlh-8 undergoes autoregulation through E2

Since both CeTwist homodimers and CeTwist/CeE/DA heterodimers bind E1 (Tw) and E2 (Tw) *in vitro*, it was important to address whether these proteins were required for *hlh-8* expression *in vivo*. The presumptive null mutant, *hlh-8 (-)*, was used to address whether *hlh-8* is able to undergo autoregulation through the intron. The Egl phenotype of *hlh-8 (-*) animals is due to improper development of the vms. However, when reporter constructs of non-target genes that express in the vms are introduced into the *hlh-8 (-)* background, vulval muscle-like cells can be seen. The gene, T12D8.9 is one

such non-target gene that is expressed in the vms of wild-type animals (Fox et al., 2007; Wang et al., 2006). A *gfp* reporter construct of either T12D8.9 or pSM10(E2a) was crossed into *hlh-8 (-)* animals. The extrachromosomal T12D8.9::*gfp* reporter was expressed in the vms of 97% of wild-type young adults (n=30) and in vulval muscle-like cells of 100% of *hlh-8 (-)* young adults (n=24). In a wild-type background, pSM10(E2a) was expressed in the vm cells of 92% young adults scored (n=66). However, there was no expression of pSM10(E2a) in the vulval muscle-like cells of *hlh-8 (-)* young adults scored (n=59) (Fig. 14). Therefore, the expression of pSM10(E2a) in the vms depends upon the presence of wild-type CeTwist molecules.

Regulation of *hlh-8* by CeE/DA

The EMSA analysis revealed that E1 (Tw) and E2 (Tw) preferentially were bound by the CeTwist/CeE/DA heterodimers compared to CeTwist homodimers (Fig. 13C). Since a null mutation of the gene that encodes for CeE/DA, *hlh-2*, has not been isolated, *hlh-2* RNAi was performed to investigate whether CeE/DA is also responsible for *hlh-8* expression. *hlh-2* RNAi treatment of L4 animals results in embryonic lethality of progeny (Krause et al., 1997). Therefore, synchronized L1 animals carrying the pSM10(E2a) transgene were fed bacteria expressing either double stranded *hlh-2* RNA or an empty control vector. Animals carrying the transgene were scored for *hlh-2* associated phenotypes and expression of *hlh-8::gfp*. Previously, it was shown that *hlh-2* RNAi treated animals are sterile and have a protruding vulva (Pvl) (Kamath et al., 2000; Karp and Greenwald, 2004). Thus, the *gfp* pattern of RNAi treated animals was scored in

conjunction with those phenotypes to be confident the animals had a sufficient decrease in the amount of CeE/DA present. There was a reduction of *hlh-8::gfp* expression in *hlh-2* RNAi animals. In 58% of the *hlh-2* RNAi treated animals (n=86), the *gfp* was expressed in the vms, compared to 92% of the control animals (n=92). There was less of a reduction in the hmc or enteric muscles with treatment. The hmc expression decreased from 91% in control animals to 84% in treated animals. Furthermore, expression in any of the enteric muscles decreased from 100% in control to 93% in treated animals. However, the hmc and enteric muscles are already born and expressing the *gfp* at the time of treatment so this experiment would only detect whether CeE/DA was required for maintenance rather than initial expression in these cells. This decrease of *hlh-8::gfp* expression in *hlh-2* RNAi treated animals was not as dramatic as seen in *hlh-8 (-)* animals. This result may reflect the CeE/DA knock-down technique, versus CeTwist knock-out approach, or perhaps illustrates a partial requirement of CeE/DA in CeTwist transcriptional regulation. Regardless, knockdown experiments of CeE/DA confirmed an important role of CeE/DA in CeTwist regulation.

An *hlh-8* intron mutant has a subset of *hlh-8 (-)* defects

hlh-8 (iΔ) phenotype:

To explore the contribution of the first intron of *hlh-8* to CeTwist function, animals containing a deletion mutation in intron 1 were characterized. Upon receiving the mutant animals, they were backcrossed to wild-type animals a minimum of eight times to alleviate any secondary mutation that may have been present. In *hlh-8 (iΔ)*

animals there is a 646 nucleotide deletion of the first intron of *hlh-8*. The 3' break point preserves the AG splicing acceptor site adjacent to exon 2, known to be required for splicing in a variety of genes (Aroian et al., 1993; Fig. 15A). The *hlh-8 (iΔ)* animals were Con and Semi-Egl (Fig. 15D, E). The animals were able to lay embryos, but not at a wild-type rate, thus leading to embryos becoming stacked within the uterus (Fig. 15C vs 15E) and an overall lower brood size. *hlh-8 (iΔ)* animals laid an average of 26, 50, and 12 embryos on day 1, 2, and 3 of adulthood, respectively and had an average brood size of 90 progeny (n=20). In comparison, wild-type animals laid 77, 177, and 24 embryos on day 1, 2, and 3 of adulthood, respectively and had an average brood size of 278 progeny (n=10). In addition, 72% of *hlh-8 (iΔ)* animals developed either a Pvl phenotype or an extreme protruded vulva (Spu) within 5 days of adulthood (n=46). Wild-type animals did not show a Pvl or Spu phenotype within the first 5 days of adulthood (n>100). Furthermore, to determine if the Pvl/Spu phenotype was a result of the constipation, animals that were constipated due to an unrelated mutation were also assayed. A population of animals harboring an *aex-1 (sa9)* mutation did not show any Pvl or Spu phenotypes. Also, the Pvl or Spu phenotypes were not seen in the *hlh-8 (-)* animals. Altogether, in contrast to *hlh-8 (-)* animals, which were Con and Egl, *hlh-8 (iΔ)* animals were Con, Semi-Egl, and Pvl/Spu.

gfp *reporter expression:*

To test the expression of downstream targets of CeTwist in *hlh-8 (iΔ)* animals, *gfp* reporter constructs were employed. Standard genetic techniques were used to cross

integrated *gfp* reporter constructs into the *hlh-8 (iΔ)* background. The *arg-1*, *ceh-24*, and *egl-15* genes are three downstream targets of CeTwist (Harfe et al., 1998b; Corsi et al., 2000; Zhao et al., 2007). The promoter regions and expression patterns of these three genes have been well characterized (Harfe et al., 1998b; Zhao et al., 2007). Reporter constructs in wild-type animals of the targets were all expressed in the vms. Furthermore, *arg-1::gfp* was also expressed in the hmc and the four enteric muscles of wild-type animals. However, in the *hlh-8 (-)* animals, no *gfp* was expressed in any of the downstream target *gfp* reporters (Corsi et al., 2000). Similarly, in *hlh-8 (iΔ)* animals, *arg-1::gfp* and *NdEbox::gfp* (*ceh-24*) were not expressed in the animals (n>100). Conversely, *egl-15::gfp* was expressed in the vms in 15% of the animals (n=131). In wild-type animals, *egl-15::gfp* continued to express for at least 2 days of adulthood (n=32). However, of the *hlh-8 (iΔ)* animals that did express the construct, the *gfp* prematurely turned off in 74% of those animals (n=19) (Table 9). In general, *hlh-8 (iΔ)* animals were not able to express CeTwist downstream targets in a wild-type pattern. However, as with the other characterized phenotypes, *hlh-8 (iΔ)* animals are not as severe as *hlh-8 (-)* animals, in that they were able to partially activate one of the CeTwist downstream targets.

Along with exploring the activation of downstream targets in *hlh-8 (iΔ)* animals, I was also interested in characterizing the pattern of the M lineage in these animals. To accomplish this, reporters that are expressed in the M lineage and also a reporter that marks the coelomocytes as an output of proper M lineage division and differentiation were employed. As described above, the patterning and differentiation of the M lineage

is tightly controlled in *C. elegans* to become body wall muscles, coelomocytes and sex muscles, including vms (Fig. 7). It has been shown previously that CeTwist has an intimate role with the M lineage (Corsi et al., 2000). When an *hlh-8* promoter fused to a *gfp* reporter construct is used, the M cell and descendants express *gfp* prior to cell differentiation. This non-rescuing construct was used to view the patterning and division of the M cell in different genetic backgrounds: wild-type, *hlh-8 (-),* and *hlh-8 (iΔ)* (Table 10). As shown previously, in *hlh-8 (-)* animals there is a characteristic mispatterning of the early M cell divisions (Corsi et al., 2000). Specifically, in wild-type animals the first M cell division was 100% dorsal/ventral (n=72). However, normal dorsal/ventral division of the M cell occurred only 30% of the time in *hlh-8 (-)* animals (n=98). Furthermore, in *hlh-8 (iΔ)* animals dorsal/ventral division of the M cell occurs 79% of the time (n=97) (Table 10). Wild-type animals have two SM cells that arise out of the M lineage (n=22). It has also been shown previously that *hlh-8 (-)* animals have extra SMs (Corsi et al., 2000). Similar results were observed in this study, in which only 32% of *hlh-8 (-)* animals had two SMs (n=38). The *hlh-8 (iΔ)* animals also displayed this phenotype to a lesser degree where 48% of *hlh-8 (iΔ)* animals had two SMs (n=44) (Table 10). Lastly, the division of the SMs was scored. In *hlh-8 (-)* and wild-type animals, the SMs undergo three divisions (wild type n=26, *hlh-8 (-)* n=27). However, in *hlh-8 (iΔ)* animals 8% of the time the SMs went undivided (n=53) (Table 10). Therefore, *hlh-8 (iΔ)* animals do display M mesodermal patterning and differentiation defects, but not as frequently as the defects as in *hlh-8 (-)* animals.

Of the differentiated cell types to arise from the M lineage one is the coelomocytes. Two of the six coelomocytes arise out of the M lineage. An intrinsic coelomocyte marker was used to express *gfp* in the 6 coelomocytes in wild-type animals (Yanowitz et al., 2004). This construct was used in an *hlh-8 (iΔ)* background and revealed no significant difference in the number of coelomocytes (n=46) compared to wild-type animals (n=33). However, in *hlh-8 (-)* animals only 77% of the time there were the correct number of 6 coelomocytes. There were too few coelomocytes 21.5% of the time (n=130) (Table 9). This data was similar to what was reported earlier (Corsi et al., 2002). Therefore, it is unlikely the deletion of intron 1 in *hlh-8 (iΔ)* animals is affecting the differentiation of these cells.

Splicing defects in *hlh-8 (iΔ)* animals

Due to the incomplete penetrance of the phenotype of *hlh-8 (iΔ)* and the position of the intron deletion, I investigated whether there could be splicing defects at the *hlh-8* locus in the mutant animals. The 3' deletion break point of *hlh-8 (iΔ)* is adjacent to two nucleotides that are required for the splicing acceptor site. Reverse transcription was performed to determine if *hlh-8 (iΔ)* animals had aberrant splice products due to the disruption of the splicing acceptor site. Primers that correspond to exon 1 and exon 5 were used to amplify the cDNA and revealed five spliced products from the *hlh-8* locus in *hlh-8 (iΔ)* animals. The spliced products were cloned and sequenced (Fig. 16). Two of the products had a larger molecular weight than the wild-type spliced product, and the sequence revealed that the splicing occurred into intron 1. Protein formation is not

predicted to occur in these fragments due to numerous stop codons in all three frames throughout the first intron. The third spliced product corresponded to wild-type splicing, in which only exons remained in the mRNA. The *hlh-8 (iΔ)* animals had a decreased level of the wild-type product compared to wild-type animals, when normalized against actin transcript levels. The remaining two splice products corresponded to splicing occurring into exon 2 or splicing out exon 2 along with all introns. Splicing into exon 2 led to a frameshift followed by a stop codon and thus, is not predicted to form a functional protein. However, the smallest molecular weight transcript does not cause a frameshift and potentially could result in a protein product that contained the intact basic domain, but is lacking the majority of the helix-loop-helix domain. Altogether, the *hlh-8 (iΔ)* animals had *hlh-8* splicing defects that caused an overall decrease in CeTwist and four alternative splice products (Fig. 16). This decrease in CeTwist is likely to contribute to the phenotype of the *hlh-8 (iΔ)* animals.

DISCUSSION

In this study, it was shown that two conserved E boxes in the first intron of *hlh-8*, E1 and E2, were necessary for *hlh-8::gfp* expression in the enteric muscles, hmc, and vms (Fig. 10). Furthermore, *in vitro* and *in vivo* results showed that E2 was more critical for the expression of *hlh-8* in these tissues. CeTwist homodimers and CeTwist/CeE/DA heterodimers bound to both E1 and E2 *in vitro*. However, both types of dimers had stronger affinity for E2 than to E1 (Fig. 13). The *in vivo* expression data revealed more *gfp* expression when constructs were used that contained E2 alone than those that contained E1 alone (Fig. 10). Using *hlh-8 (-)* animals and pSM10(E2a), it was determined that *hlh-8* undergoes autoregulation through E2 (Fig. 14). Additionally, *hlh-2* RNAi revealed an important role for CeE/DA in the expression of *hlh-8* through E2.

Furthermore, characterization of the *hlh-8 (iΔ)* allele revealed that the animals had attenuated phenotypes when compared to presumptive null *hlh-8 (-)* animals. It was shown that *hlh-8 (iΔ)* animals were Con and Semi-Egl, could partially activate downstream genes, and had M lineage defects (Fig. 15; Table 9; Table 10). *hlh-8 (iΔ)* animals had splicing defects due to the close proximity of the deletion break point to the splicing acceptor site (Fig. 16).

Two E boxes regulate *hlh-8* expression

Minimal promoter *gfp* constructs containing portions of the *hlh-8* first intron revealed two important E boxes. Expression of *hlh-8::gfp* was contingent on DNA that

contained either E1 or E2 (Fig. 10). Site-directed mutagenesis revealed the necessity of E1 and E2 because *hlh-8::gfp* expression was obliterated when both sites were mutated (Fig. 10E).

This study also revealed that additional DNA is important to enhance expression of *hlh-8::gfp* in specific tissues. Intriguingly, the 3' DNA adjacent to E2 was important for expression in the vms and hmc. Specifically, a construct that contained E2 and 133 nucleotides of the adjacent E2 3' DNA had high expression in all tissues. However, removing 60 bps from the 3' region resulted in no expression in the hmc and the vms expression was decreased from 87% to 22% in animals examined. A construct that removed an additional 65 nucleotides of the 3' DNA was not expressed in the hmc nor the vms (pSM10(E2a); pSM15(E2b); pSM20(*iΔ*2); Fig. 10B-D). These constructs clearly emphasize the importance of the 3' flanking DNA of E2.

There are two possibilities to explain the importance of the E2 3' flanking DNA. First, the sequence may be critical for CeTwist dimers to properly bind. Changing the three nucleotides flanking the E boxes in the promoter of a CeTwist target, *ceh-24*, disrupts the activity of the E boxes (Harfe et al., 1998b). However, the additional DNA of the constructs used in this study that affected the activity of E2 was minimally nine nucleotides away from E2 (pSM20(*iΔ*2); Fig. 10D). Furthermore, the *in vitro* gel shift assay results clearly demonstrated that the seven flanking nucleotides are important for CeTwist homo and heterodimers to bind, since both E1 and E2 contain the same E box of CATCAG yet they were bound by the proteins with differing affinities (Fig. 13). However, this explanation would have to be tissue-dependent because there was not a

dramatic change in the enteric muscle expression in constructs that removed the 3' flanking DNA of E2, compared to expression in the hmc and vms. CeTwist dimer selection may explain the tissue-specific expression dependence on the additional 3' DNA. Perhaps, CeTwist homodimers control expression in the enteric muscles and binding of this dimer to the E boxes is not sensitive to the 3' DNA sequence. On the other hand, CeTwist/CeE/DA heterodimers may control hmc and vms expression and this dimer binding is dependent on the 3' sequence. The second explanation is that the adjacent DNA is required for a tissue-specific co-factor to bind, to either allow for CeTwist dimers to bind or to function properly. This type of tissue-specific regulation has previously been proposed in *arg-1*, a downstream target of CeTwist (Zhao et al., 2007). The *arg-1* gene is expressed in the hmc, vms, and enteric muscles, similar to the expression of *hlh-8*. Three E boxes, E1 (*arg-1*), E2 (*arg-1*), and E3 (*arg-1*), and another element, called a GT box, located in the upstream promoter region of *arg-1* are responsible for distinct aspects of tissue-specific expression. Specifically, E2 (*arg-1*) is required for *arg-1* expression in all three tissue types. However, the other two E boxes are required for specific spatial expression of *arg-1*. E1 (*arg-1*) is necessary for regulating expression of *arg-1* in the hmc and vms and E3 (*arg-1*) is necessary for expression in the enteric muscles. Furthermore, the GT element is important for expression in the hmc and vms, but does not influence the expression of *arg-1* in the enteric muscles (Zhao et al., 2007). However, there were no GT elements in the first intron of *hlh-8*, nor were any other specific elements identified, but there were sequences 3' of E2 that were completely conserved among nematode species. Both of the above

explanations regarding the importance of the 3' DNA, point to the dynamic nature of the tissue-specific transcriptional regulation of CeTwist and also elucidate the multiple dimensions of this transcription factor to regulate other genes. It will be important to explore this region of DNA that represents either a binding element for a CeTwist co-factor or sequences that are important for proper CeTwist dimer binding.

E1 has repressor and enhancer activity

Another observation from the minimal promoter data was the discovery of the potential repressor role of E1. This role was clearly seen in the SDM comparing the construct that contains both E1 and E2, to the SDM construct that has E1 mutated, but keeps E2 intact (pSM7(E1E2); pSM24(E1mut); Fig. 10B, E). When E1 is disrupted, the expression level in all tissue types increased. This pattern can also be seen when comparing the construct that contains both E boxes with a construct that contains E2 and lacks E1 (pSM7(E1E2); pSM10(E2a); Fig. 10B). Interestingly, the DNA directly upstream of E1 is highly conserved between all 5 organisms, and may contain a site where an additional factor may bind (Fig. 12). To explore the possibility of a co-repressor element present in the conserved DNA, the TESS program was used (http://www.cbil.upenn.edu/tess). This program searches multiple data-bases for factors that bind to consensus sequences. However, there were not any commonly defined transcription factor binding sites identified. Therefore, this conserved portion of DNA could correspond to a binding site for a new factor or represent a non-consensus site.

hlh-8 (iΔ) phenotype might be due to a decreased level of wild-type CeTwist and also disruption of E2 3' DNA

It was shown that *hlh-8 (iΔ)* animals had a less severe phenotype than the presumptive *hlh-8* null mutants. Specifically, *hlh-8 (iΔ)* animals were Con, and semi-Egl compared to the *hlh-8 (-)* animals that are Con and Egl (Fig. 15). Furthermore, a battery of *gfp* reporter constructs were used to show that some of the downstream CeTwist targets are activated in *hlh-8 (iΔ)* animals, whereas no target genes are activated in *hlh-8 (-)* animals (Table 9). Lastly, an M lineage *gfp* reporter revealed that *hlh-8 (iΔ)* animals have M lineage defects that are less severe than *hlh-8 (-)* nematodes (Table 10).

This study has revealed two important findings to explain the *hlh-8 (iΔ)* phenotype. First, RT-PCR exposed splicing defects in *hlh-8 (iΔ)* animals due to the proximity of the deletion break point to the splicing acceptor site of intron 1. Furthermore, the splicing defects caused a decrease of wild-type CeTwist mRNA and four aberrant splice products were produced. It was predicted that only one of the aberrant splice products could possibly produce a functional protein. The decreased level of wild-type CeTwist mRNA likely contributes to the *hlh-8 (iΔ)* phenotype. The second observation to explain the *hlh-8 (iΔ)* phenotype is from the minimal promoter constructs. The DNA 3' adjacent to E2 is important for *hlh-8::gfp* expression in the vms and hmc. Furthermore, the 5' deletion break point of the *hlh-8 (iΔ)* mutation is only 9 nucleotides downstream from E2 (Fig. 12). It is plausible that the disruption of the E2 flanking nucleotides interferes with the autoregulation of *hlh-8* and, thus, contributes to the decreased expression of CeTwist in *hlh-8 (iΔ)* animals to cause the phenotypes. Perhaps

the enteric muscles require a higher level of CeTwist in order to properly develop than the vms. This could explain the reason why *hlh-8 (iΔ)* animals were Con, but could still lay eggs into the environment. This argument would also explain the rescue experiments that show genomic DNA of *hlh-8* rescues the Con phenotype more efficiently than *hlh-8* cDNA, 78% from 20%, respectively (Corsi et al., 2000). The genomic DNA containing intron 1 would allow for autoregulation of CeTwist to occur in the animals; thus, increasing the level of CeTwist molecules to a threshold that would allow for proper enteric muscle development. Since the cDNA does not contain the E1 and E2 elements, autoregulation would not take place in *hlh-8 (-)* animals with this construct, resulting in an overall lower concentration of CeTwist. Similarly, it has been shown in mammalian development that the level of Twist dictates whether a homo- or heterodimer forms. It has been proposed that an environment with higher levels of Twist leads to homodimer formation, whereas, heterodimers preferentially form in situations with decreased concentrations of Twist (Connerney et al., 2006). The data from my study supports the mammalian model. For example, perhaps enteric muscles require CeTwist homodimers for proper differentiation, but CeTwist/CeE/DA heterodimers are responsible for the hmc and vm development. Thus, in *hlh-8 (iΔ)* animals, if there is a lower amount of CeTwist in the enteric muscles then the appropriate homodimers cannot form leading to enteric muscle defects.

Model for regulating *hlh-8* expression

CeTwist is expressed in the undifferentiated M lineage, M lineage-derived coelomocytes, sex muscles, enteric muscles and the hmc. The elements for the undifferentiated M lineage cells and coelomocytes have been isolated previously in the promoter region (Harfe et al., 1998b). Furthermore, it has previously been reported that Hox factors are responsible for activating the *hlh-8* promoter (Liu and Fire, 2000). This prior information about *hlh-8* fits well with the new gene regulation discovery from this study.

A model is proposed in which the Hox factors bind to the *hlh-8* promoter and are responsible for the expression of a moderate level of CeTwist in the animal. This moderate level of CeTwist molecules is sufficient for early M lineage and coelomocyte development (Fig. 17A). Once a threshold of CeTwist molecules accumulates, autoregulation of CeTwist occurs through the E boxes in the first intron, which increases the concentration of CeTwist molecules in the tissue. This higher level of CeTwist is required for proper development of the vms, enteric muscles, and hmc (Fig. 17B). This model fits well with the fact that the vms are derived from M lineage cells (Fig 7). However, if E1 and E2 are strictly autoregulatory elements, then some other elements yet to be identified must initially be responsible for expression of CeTwist in the enteric muscles and the hmc. Furthermore, it is possible that certain target genes may require a higher level of CeTwist to be expressed at all or it is possible that a greater amount of CeTwist is required to lead to enough target gene expression for tissues to properly function. Evidence for this possibility comes from *hlh-8 (i*Δ*)* animals (Fig 17C, D).

These nematodes had a lower level of CeTwist and were able to partially activate downstream targets of CeTwist. Additionally, even though vms were made in *hlh-8 (i*Δ*)* animals, they did not function properly, which caused a semi-Egl phenotype. Thus, a threshold of CeTwist was required for development of vms and certain target activation, but a higher level was needed for proper vm and enteric muscle function and the expression of other CeTwist target genes (Fig. 17). A better understanding of exactly how individual target genes are regulated may distinguish the importance of concentration of CeTwist for specific function and tissue development.

To clarify the CeTwist transcription regulation model, it will be important to find additional elements in the intron that are responsible for contributing to the spatial expression of *hlh-8* controlled by E1 and E2. Once specific elements are identified, it will be important to locate the factors that bind to these elements. Mutational and RNAi analysis of the factors will not only elucidate the transcriptional regulation of CeTwist, but also unlock CeTwist's relationship with other transcription factors. Furthermore, an understanding of CeTwist homo- and CeTwist/CeE/DA heterodimer individual activities will provide further support for the model.

This CeTwist regulation model is strengthened by the fact that autoregulation has been reported to affect the temporospatial expression of other bHLH factors. PTF1a is a non-Twist family bHLH factor that is important in the proper development of the pancreas in mammals. In particular, PTF1a is important for the early epithelium morphogenesis and later in development for acinar cell differentiation (Masui et al., 2008). Acinar cells are polarized cells of the pancreas that have an endocrine function

(Lodish et al., 2000). Interestingly, PTF1a regulates itself in a temporospatial fashion through E boxes found in the promoter region of the gene. The autoregulatory element is shown to have a maintenance role in *PTF1a* expression in the early epithelium precursors and also later in development to maintain a superinduction of PTF1a for the differentiation program of the acinar cells (Masui et al., 2008). The maintenance role of the autoregulatory elements of PTF1a is an intriguing difference compared to CeTwist autoregulation. E1 and E2 elements seem to be important for controlling expression of CeTwist in distinct cells at specific developmental time points. Specifically, elements in the promoter region control expression of CeTwist in Sex Myoblast descendants, the sex muscle precursors, up to the point of differentiation. Whereas, E1 and E2 are responsible for CeTwist expression in differentiated vms. This control mechanism could represent a way to alter the levels of CeTwist and thereby switch which target genes are regulated in undifferentiated versus differentiated cells by CeTwist.

Human impact from this research

The homolog of *hlh-8* in humans encodes for Twist, which has an important role in both cancer progression and mesoderm development. It is evident from human disease, that the concentration of Twist molecules is critical to control. Thus, understanding the regulation of CeTwist is important to elucidate in order to gain insight into the dynamic regulation of human Twist. The inappropriate up-regulation of Twist is implicated as a critical factor for the metastasis of tumors (Yang et al., 2004). On the other hand, mutations in the coding region of this gene, which leads to an overall

decrease in functional Twist molecules, cause Saethre-Chotzen syndrome, a craniosynostotic disorder. This syndrome is an autosomal dominant disorder due to haploinsufficiency of Twist, which also speaks to the importance of Twist protein level for proper function. In the human Twist gene there are three E boxes in the 2 kb upstream region, four in intron 1, one in exon 2, and four in the 2 kb downstream region. Although none of these E boxes have the specific sequence as that of E1 and E2, CATCTG, it is plausible that human Twist undergoes autoregulation through a different E box and the transcriptional regulation is conserved between *C. elegans* and humans. A recent study analyzed the coding region of the human Twist gene in patients diagnosed with Saethre-Chotzen syndrome and found that only a fourth of these patients had a mutation in the this region of DNA (Stenirri et al., 2007). Perhaps, a disruption of an element outside of the coding region of Twist could explain the phenotype of the remaining Saethre-Chotzen syndrome patients where no gene mutation has been identified. Furthermore, mutations in non-coding Twist regions could also explain unidentified craniosynostotic disorders by disrupting an autoregulation element that causes a decrease in the level of Twist.

Future directions and significance of this work

It will be important to explore the regulation of CeTwist target genes to understand the mechanism of CeTwist's function. Additionally, investigating the regulation of CeE/DA will aid in understanding the method of CeTwist dimer selection and may elucidate the tissue-specific expression of CeTwist. Currently, there is little to

nothing known about the regulation of this partner of CeTwist. Investigations can encompass understanding *in vivo* if CeTwist targets are bound by CeTwist homodimers versus CeTwist/CeE/DA heterodimers and if dimer choice is affected by more or less CeTwist availability. Furthermore, isolating co-factors that aid in CeTwist regulation will be essential to unlocking the mechanism of CeTwist function.

Twist is an interesting protein to use as a model to understand transcriptional control due to the dynamics of the homo- and heterodimers that function in a dose-dependant manner. Understanding the expression regulation of this important transcription factor was key to clarifying its function in tissue-specific roles. This research is the first time a protein in the Twist family has been shown to undergo autoregulation. More significantly, the finding from this study that a protein is differentially regulated to express in the same tissue at different time points has not been reported in the field previously and opens a new avenue of research for understanding the mechanism of tissue-specific transcriptional control.

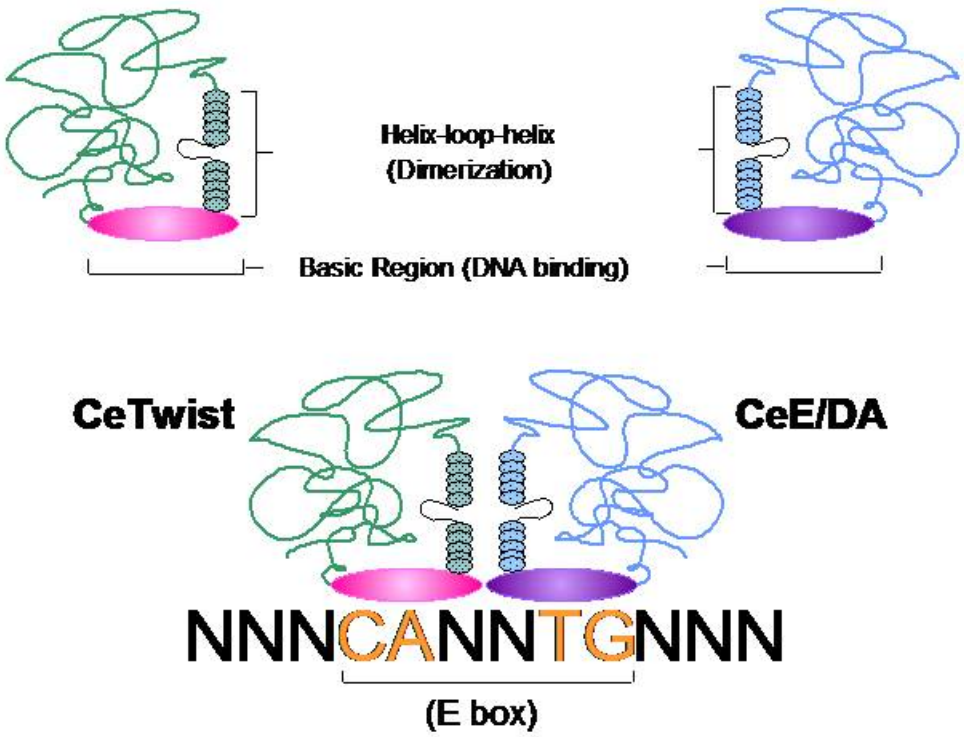


Figure 1. CeTwist is a basic helix-loop-helix transcription factor. Diagram represents individual monomers of CeTwist, on the left, and its known binding partner, CeE/DA, on the right. The helix-loop-helix domain of CeTwist is required for dimerization to itself as a homodimer or to a partner as a heterodimer, as depicted here. The basic domain is important for binding to a conserved sequence, CANNTG, called an E box.

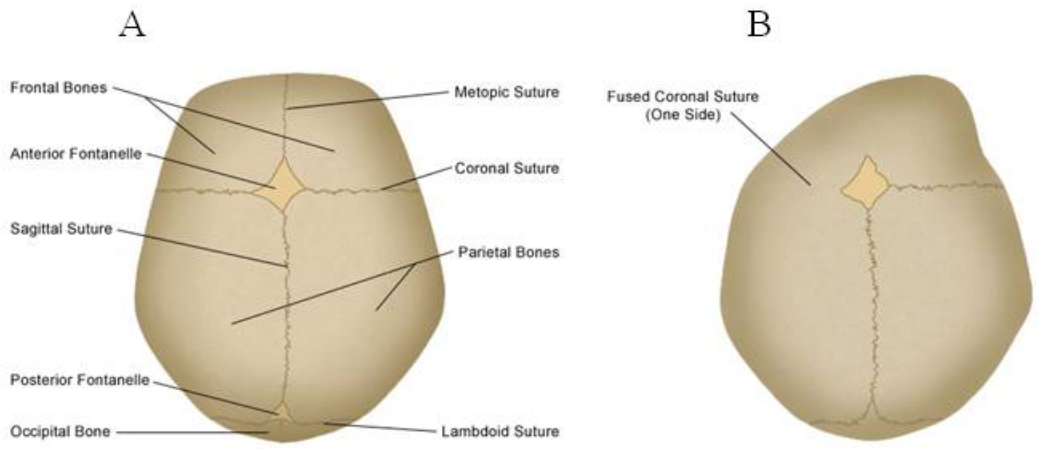


Figure 2. Saethre-Chotzen syndrome is characterized by premature closure of cranial sutures. Schematic of the top view of two newborn skulls. A) Image of a normal skull with all cranial bones and sutures labeled. B) Infant with craniosynostosis. The lack of the left coronal suture due to premature fusion is indicated. The resulting misshapen skull is also illustrated.
(http://www.lpch.org/DiseaseHealthInfo/HealthLibrary/craniofacial/cranio.html)

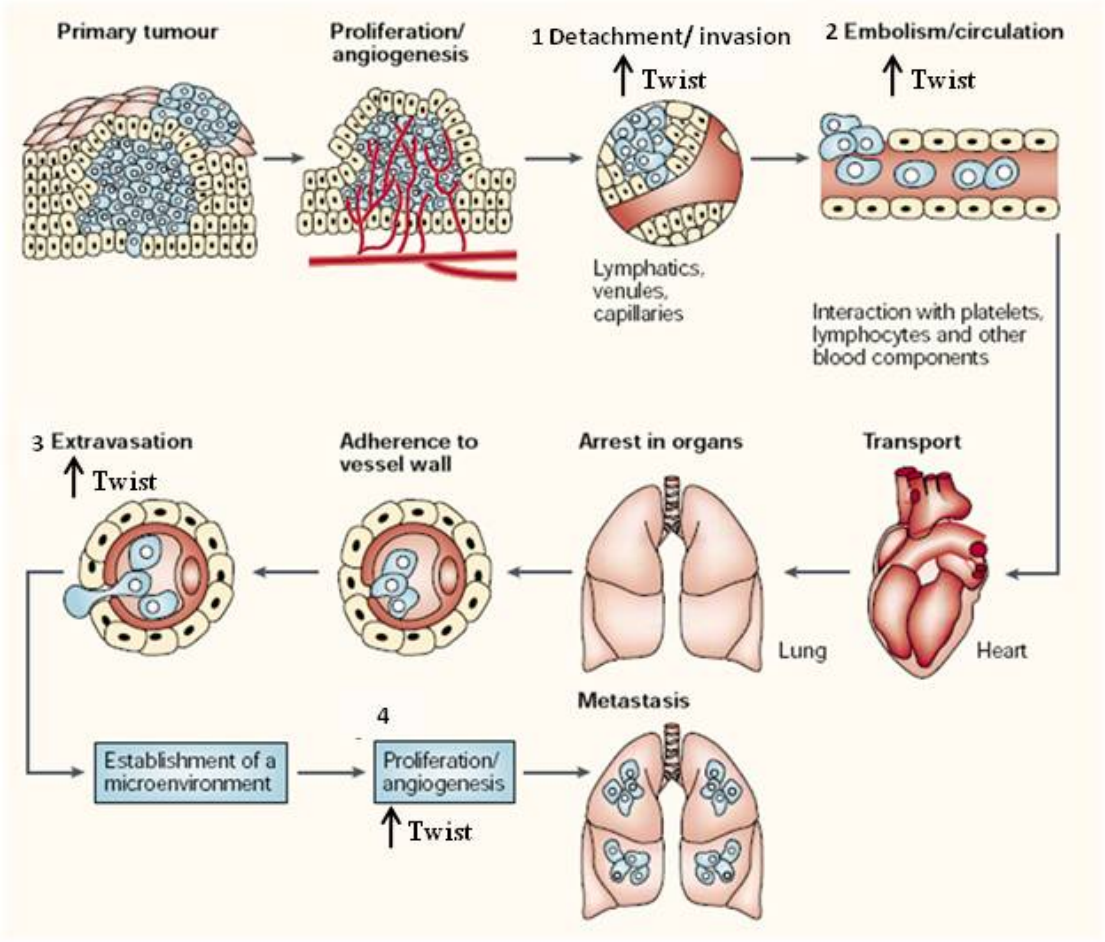


Figure 3. There are four main steps in the process of metastasis. A primary tumor is established in a specific tissue of the body. 1) Detachment/invasion is characterized by the loss of cell polarity and cell-cell adhesions. 2) Embolism/circulation is marked by the cells entering the circulatory system. 3) Some of the cells that are able to survive the circulatory system can undergo extravasation, where they leave the circulatory system to enter a new environment. 4) The last step of the formation of metastasis is the proliferation of secondary tumors in the new site. Twist is overexpressed in all four steps of metastasis (Adapted from Fidler, 2003).

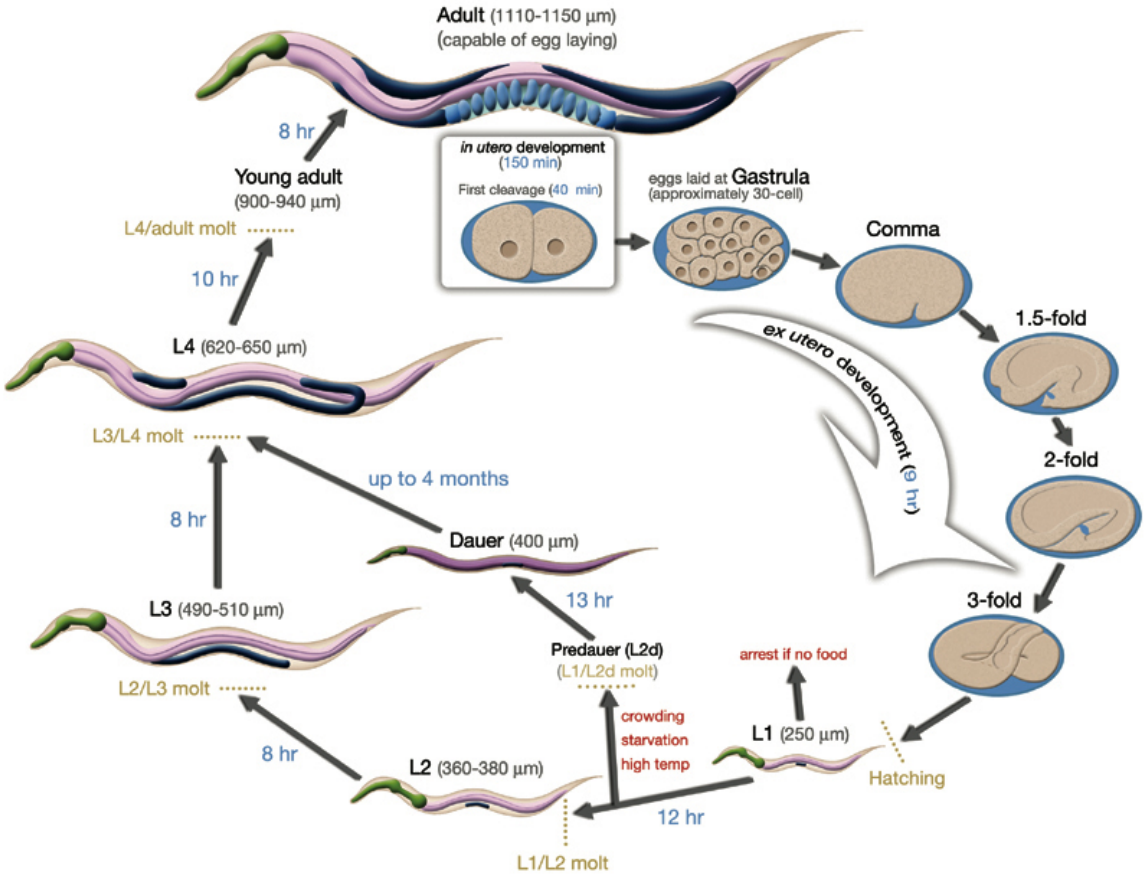


Figure 4. Lifecycle of *C. elegans*. The phases of development are written in black above each diagram. Time periods between stages are written in blue. An alternative life cycle is shown between L1 and L2 stages and is called the dauer stage. Animals will take this intermediate growth route under stressful environmental conditions. As a dauer larva, the animal can survive for months without food. Once environmental conditions improve, the dauer larva can continue developing into an adult. The entire life cycle is about 60 hours at room temperature (www.wormatlas.org).

Figure 5. *C. elegans* have transparent cells. An adult hermaphrodite is shown through Differential Interference Contrast imaging. The scale bar is 0.1mm. Embryos are seen inside and outside of the animal. Lateral view, posterior to the right. (www.wormatlas.org)

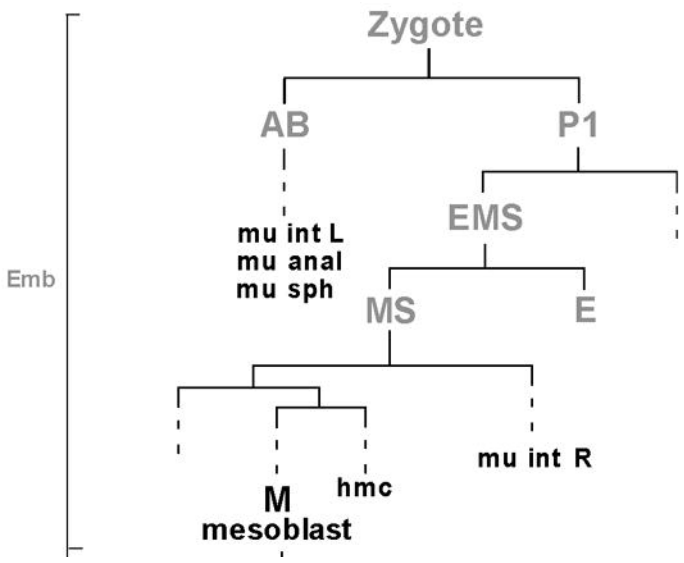


Figure 6. CeTwist is expressed in four different cell lineages. An abbreviated lineage of the embryonic phase of *C. elegans*. Founder cells are in grey and cells that express CeTwist are shown in black. Cell divisions are depicted by a horizontal line. Dashed lines correspond to cell divisions not shown. 'hmc' stands for head mesodermal cell. 'mu int L' and 'mu int R' are left and right intestinal muscles, respectively. 'mu anal' is the anal depressor and 'mu sph' is the anal sphincter cell. (adapted from Zhao et al., 2007)

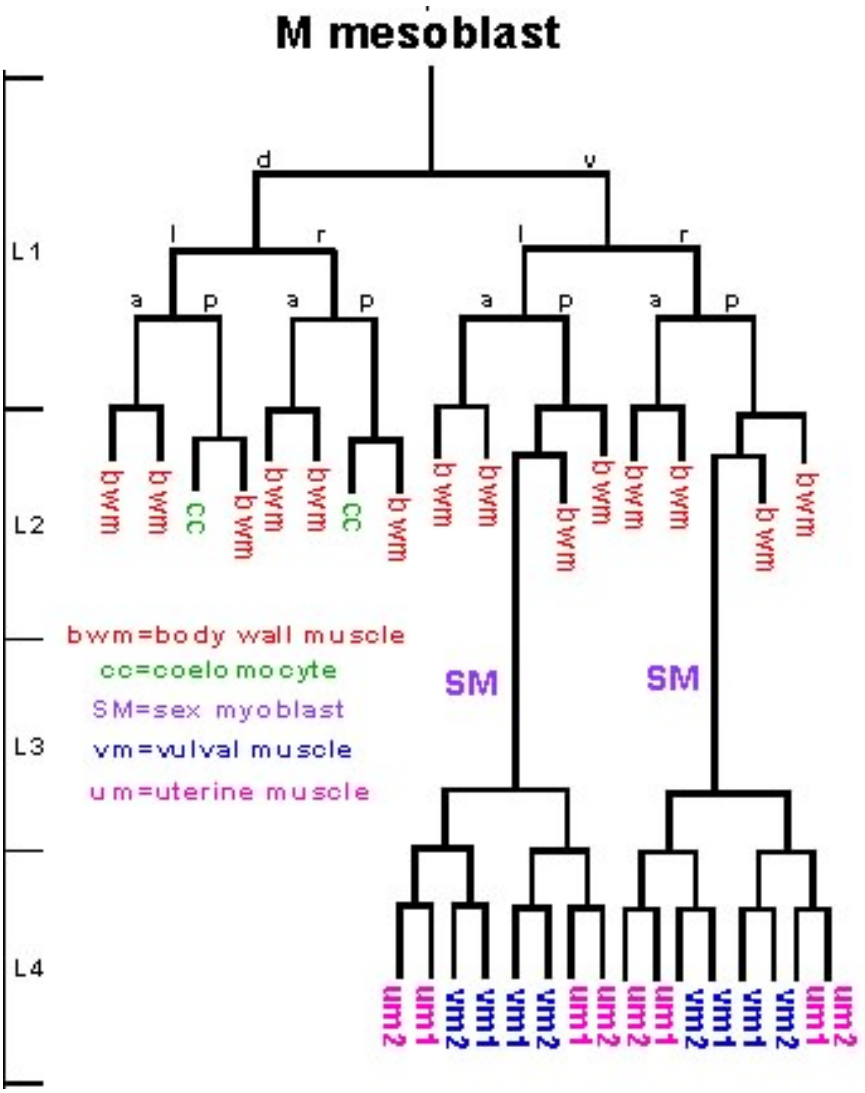


Figure 7. CeTwist has an intimate role with the M lineage. Schematic of the M lineage with timepoints during larval development indicated on the left. The M mesoblast is born embryonically during the three-fold stage and divides post-embryonically. 'd' and 'v' indicate dorsal-ventral division, 'l' and 'r' represent a left-right division, and 'a' and 'p' indicate a anterior-posterior division. The first division of the M mesoblast is a dorsal-ventral division. Successive rounds of division lead to 16 cells that will differentiate into 14 body wall muscles (bwm) and two coelomocytes (cc) in the L2 stage. Furthermore, through these divisions, two sex myoblast (SM) cells will arise and migrate toward the center of the animal where they will undergo four rounds of division leading to 16 cells that will differentiate into eight vulval muscles (vm) and eight uterine muscles (um) in the late L4 stage. (Adapted from Corsi et al., 2000)

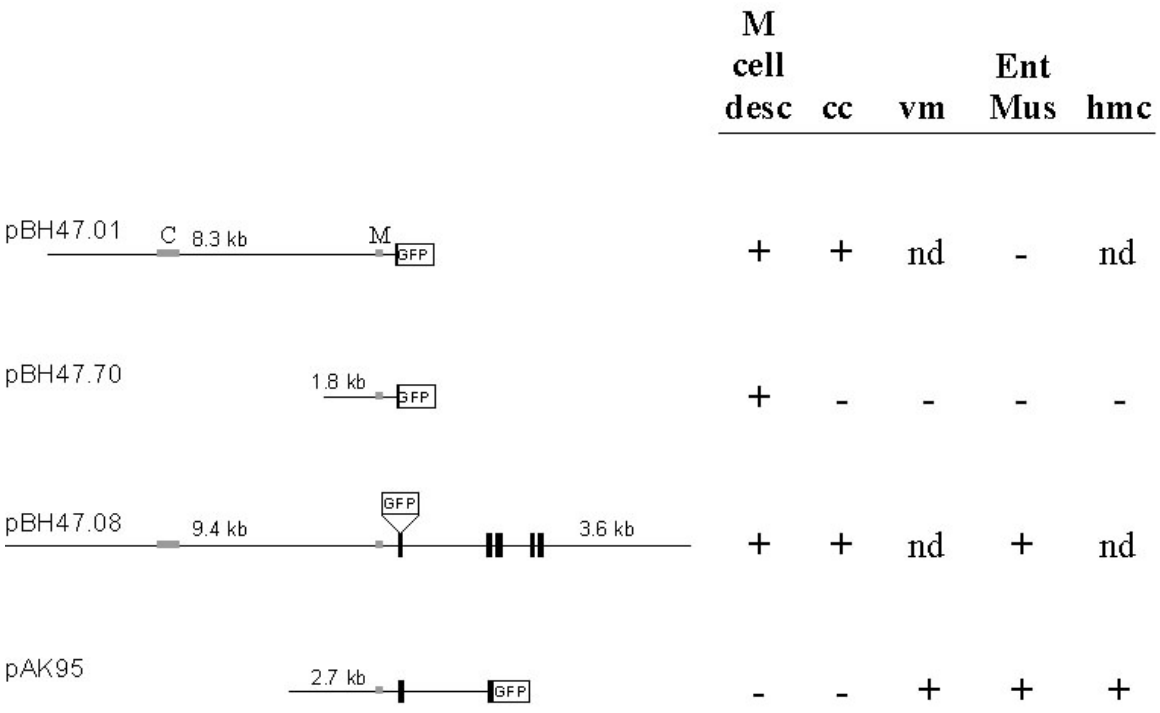


Figure 8. Summary of expression of several GFP reporter constructs of *hlh-8*. Constructs that contain intron 1 have expression in the vulval muscles and enteric muscles. 'M cell desc' refers to the undifferentiated cells that divide from the M mesoblast cell. Coelomocytes, vulval muscles, enteric muscles, and the head mesodermal cell are represented by 'cc', 'vm', 'Ent Mus', and 'hmc', respectively. The 'M' and 'C' labels in the *gfp* reporters mark the location of the elements that control expression in the undifferentiated M mesoblast cell descendents and in the coelomocytes, respectively. '+' corresponds to the construct having expression in the specific cell type, '-' means there was no GFP present in the particular cell type, and 'nd' refers to not determined. Expression pattern of pBH47.08 and pBH47.01 was reported in Harfe et al 1998b. Expression patterns of pBH47.70 and pAK95 were collected in this study.

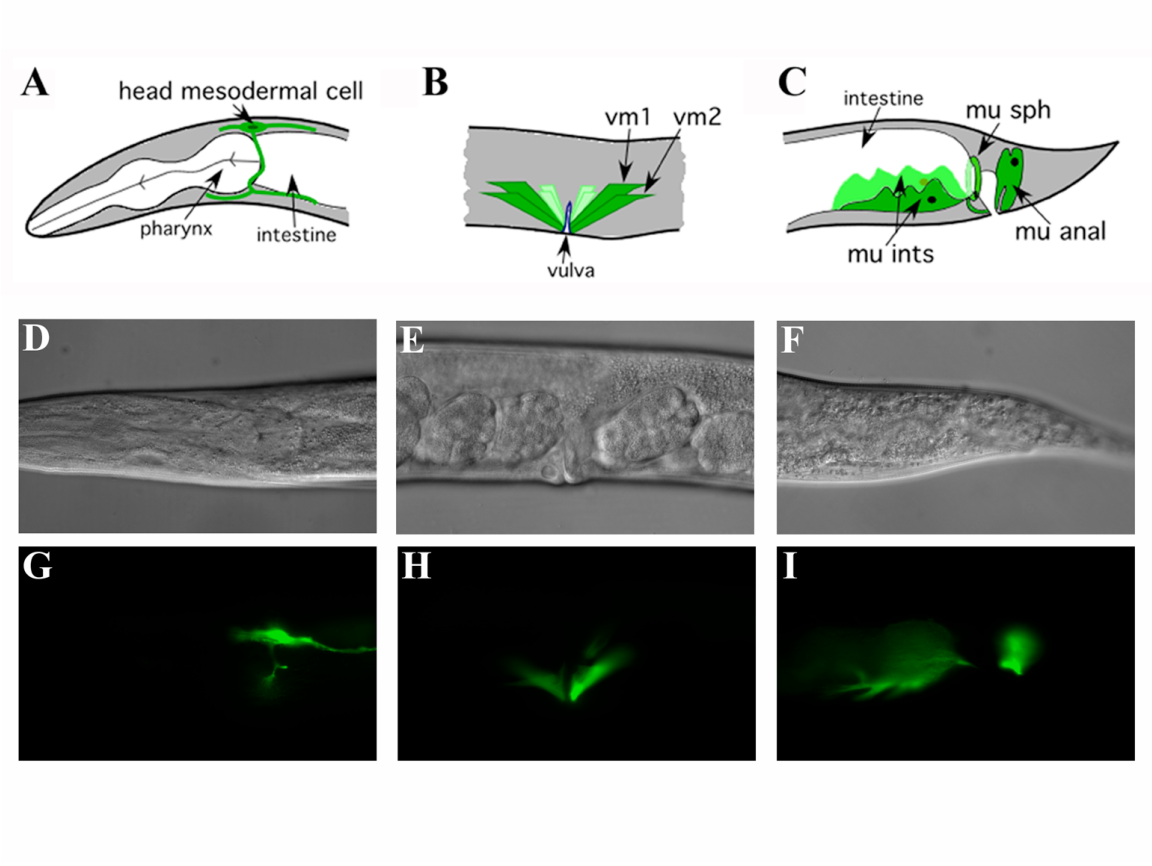


Figure 9. Intron 1 sequences control expression of *hlh-8* in differentiated mesodermal tissues. (A-C) Schematic representation of the tissues where *hlh-8* is expressed. (D-F) Nomarski and (G-I) GFP images of tissues. *hlh-8* expression is in (A, D, G) the head mesodermal cell, (B, E, H) the vulval muscles, (C, F, I) and in the four enteric muscles: left and right intestinal muscles (mu ints), anal sphincter (mu sph), and anal depressor (mu anal).

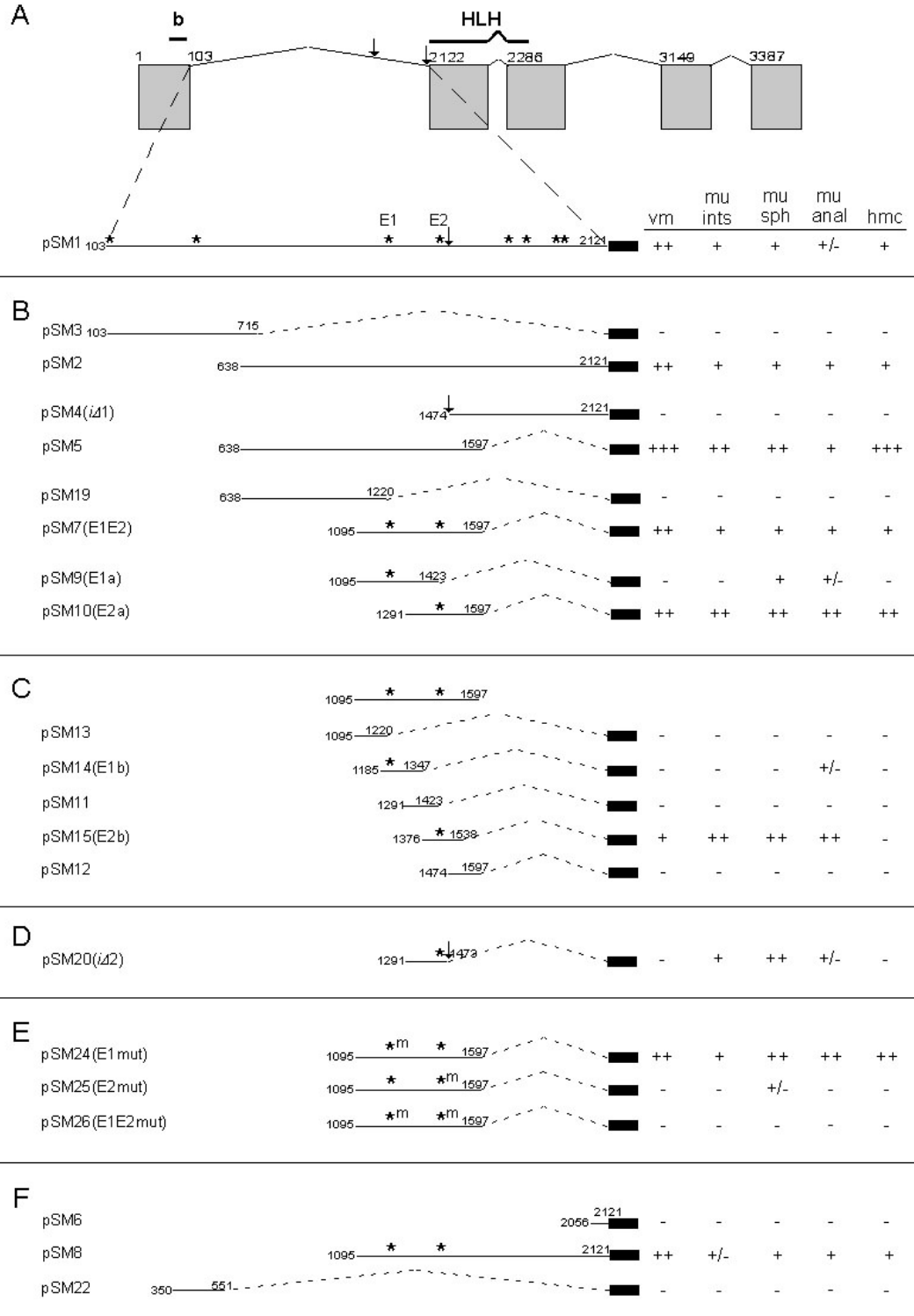

A
b
HLH
1
103
2122
2286
3149
3387
E1
E2
vm
mu ints
mu sph
mu anal
hmc
pSM1 103 2121 ++ + + +/- +
B
pSM3 103 715 - - - - -
pSM2 638 2121 ++ + + + +
pSM4(iΔ1) 1474 2121 - - - - -
pSM5 638 1597 +++ ++ ++ + +++
pSM19 638 1220 - - - - -
pSM7(E1E2) 1095 1597 ++ + + + +
pSM9(E1a) 1095 1423 - - + +/- -
pSM10(E2a) 1291 1597 ++ ++ ++ ++ ++
C
1095 1597
pSM13 1095 1220 - - - - -
pSM14(E1b) 1185 1347 - - - +/- -
pSM11 1291 1423 - - - - -
pSM15(E2b) 1376 1538 + ++ ++ ++ -
pSM12 1474 1597 - - - - -
D
pSM20(iΔ2) 1291 1473 - + ++ +/- -
E
pSM24(E1mut) 1095 m 1597 ++ + ++ ++ ++
pSM25(E2mut) 1095 m 1597 - - +/- - -
pSM26(E1E2mut) 1095 m m 1597 - - - - -
F
pSM6 2056 2121 - - - - -
pSM8 1095 2121 ++ +/- + + +
pSM22 350 551 - - - - -

Figure 10. Two E boxes in intron 1 are necessary for expression of *hlh-8::gfp* in a subset of tissues. Different regions of intron 1 were inserted into the *egl-18::gfp* minimal promoter construct (shown as a black rectangle) and each construct was scored for activity in at least two independent lines and at least 30 animals for each line. Schematics of constructs and *hlh-8* genomic DNA are drawn to scale. GFP activity was scored in the vulval muscles (vm), intestinal muscles (mu ints), anal sphincter (mu sph), anal depressor (mu anal), and head mesodermal cell (hmc). The amount of expression was determined based on the percentage of animals expressing the *gfp*: +++ (90-100%); ++ (60-89%); + (20-59%), +/- (7-19%), - (0-6%). (A) Line drawing of the *hlh-8* locus. The first nucleotide of all exons and intron 1 are numbered above. The *hlh-8 (iΔ)* 5' deletion break point is indicated by a vertical arrow. pSM1 includes the entirety of intron 1. All E boxes found in intron 1 are indicated with an asterisk. (B) Four sets of constructs were made to isolate a smaller region of DNA that retained expression. E1 and E2 are indicated with asterisks. (C) pSM7(E1E2) was divided into 5 smaller fragments to narrow the region required for expression. Expression was retained in those fragments that contain either E1, pSM14(E1b), or E2, pSM15(E2b), E box. (D) pSM20(*iΔ*2) is a 3' deletion of pSM10 where the *hlh-8 (iΔ)* deletion starts. (E) Site directed mutagenesis (SDM) of E1 and E2 to test the contribution of each E box to expression. Mutated E boxes are indicated by m. (F) Additional constructs made to test for additional enhancer elements or to show expression of a control element. pSM6 contains a small region of DNA that had high homology in nematode species. pSM8 contains the DNA that was homologous between *C. elegans* and *C. briggsae*. pSM22 contains the E box that was used as a negative control for the EMSA experiments.

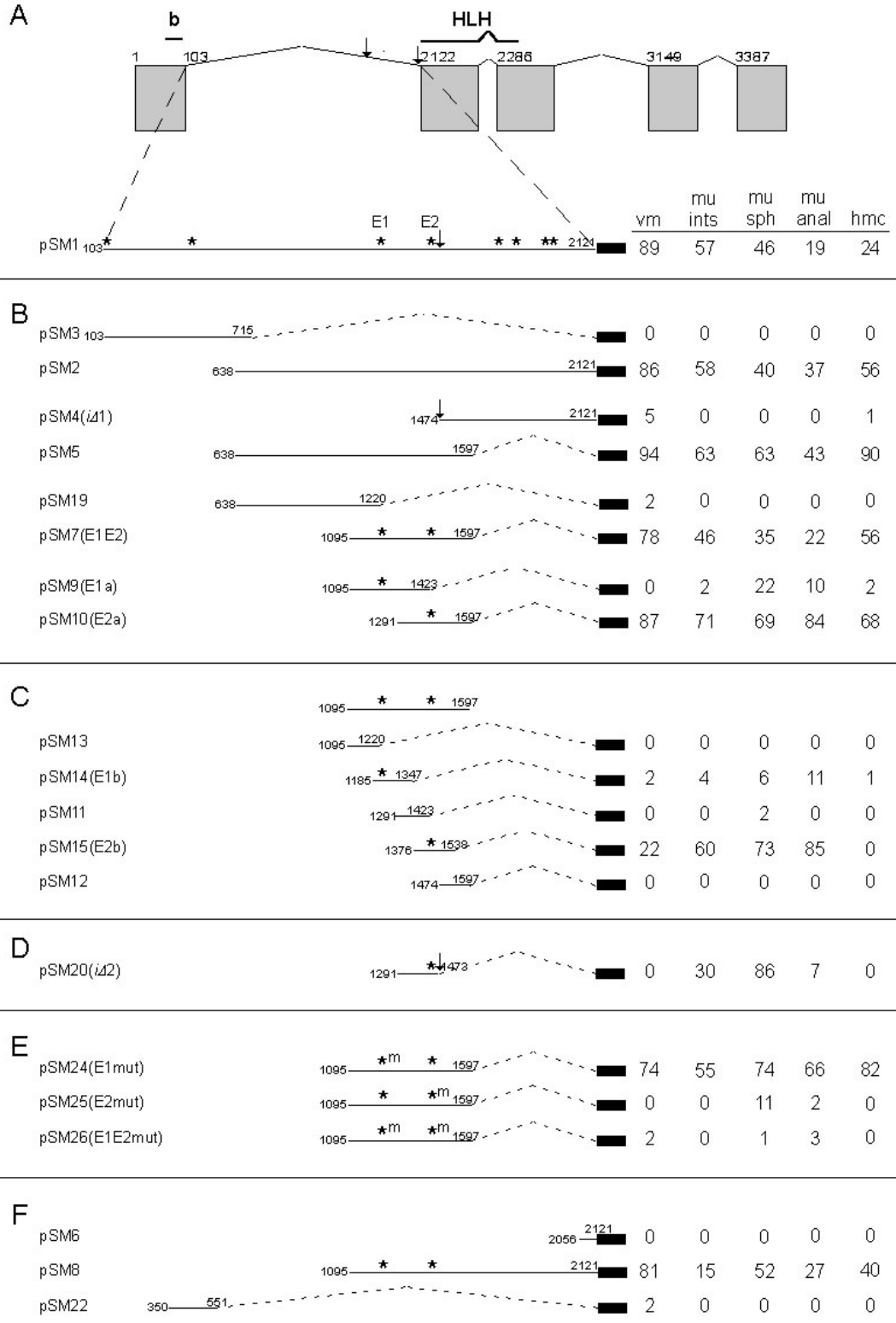

A
b
HLH
1
103
2122
2286
3149
3387
E1
E2
vm
mu ints
mu sph
mu anal
hmc
pSM1 103 2124 89 57 46 19 24
B
pSM3 103 715 0 0 0 0 0
pSM2 638 2121 86 58 40 37 56
pSM4(iΔ1) 1474 2121 5 0 0 0 1
pSM5 638 1597 94 63 63 43 90
pSM19 638 1220 2 0 0 0 0
pSM7(E1E2) 1095 1597 78 46 35 22 56
pSM9(E1a) 1095 1423 0 2 22 10 2
pSM10(E2a) 1291 1597 87 71 69 84 68
C
1095 1597
pSM13 1095 1220 0 0 0 0 0
pSM14(E1b) 1185 1347 2 4 6 11 1
pSM11 1291 1423 0 0 2 0 0
pSM15(E2b) 1376 1538 22 60 73 85 0
pSM12 1474 1597 0 0 0 0 0
D
pSM20(iΔ2) 1291 1473 0 30 86 7 0
E
pSM24(E1mut) 1095 m 1597 74 55 74 66 82
pSM25(E2mut) 1095 m 1597 0 0 11 2 0
pSM26(E1E2mut) 1095 m m 1597 2 0 1 3 0
F
pSM6 2056 2121 0 0 0 0 0
pSM8 1095 2121 81 15 52 27 40
pSM22 350 551 2 0 0 0 0

Figure 11. Expression values of *hlh-8::gfp* constructs. The same data that are found in Figure 10 are shown here with the percentage of animals that expressed *gfp* in the particular tissue indicated numerically. Each value is determined from a minimum of 30 animals.

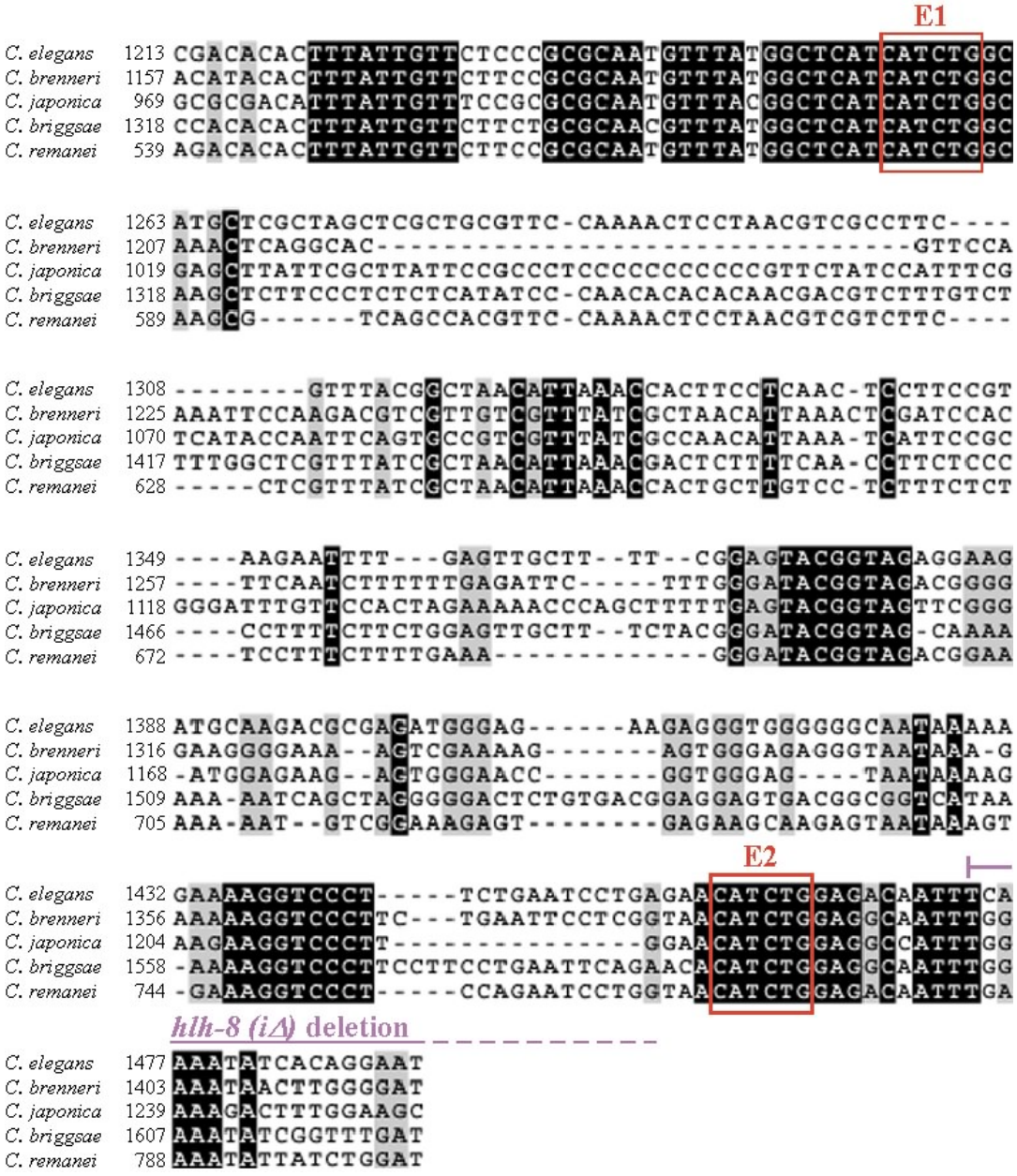


Figure 12. Conservation of intron 1 between distantly related nematodes. Alignment of a portion of intron 1 of *hlh-8* in *C. elegans* and homologs found in *C. brenneri*, *C. japonica*, *C. briggsae* and *C. remanei*. Sequences were obtained from www.wormbase.org. Black shading indicates where all five nucleotides from each species are identical. Red boxed in areas marks the location of the E1 and E2 E boxes. The nucleotides below the purple line indicate those that are deleted in the *hlh-8 (iΔ)* mutation.

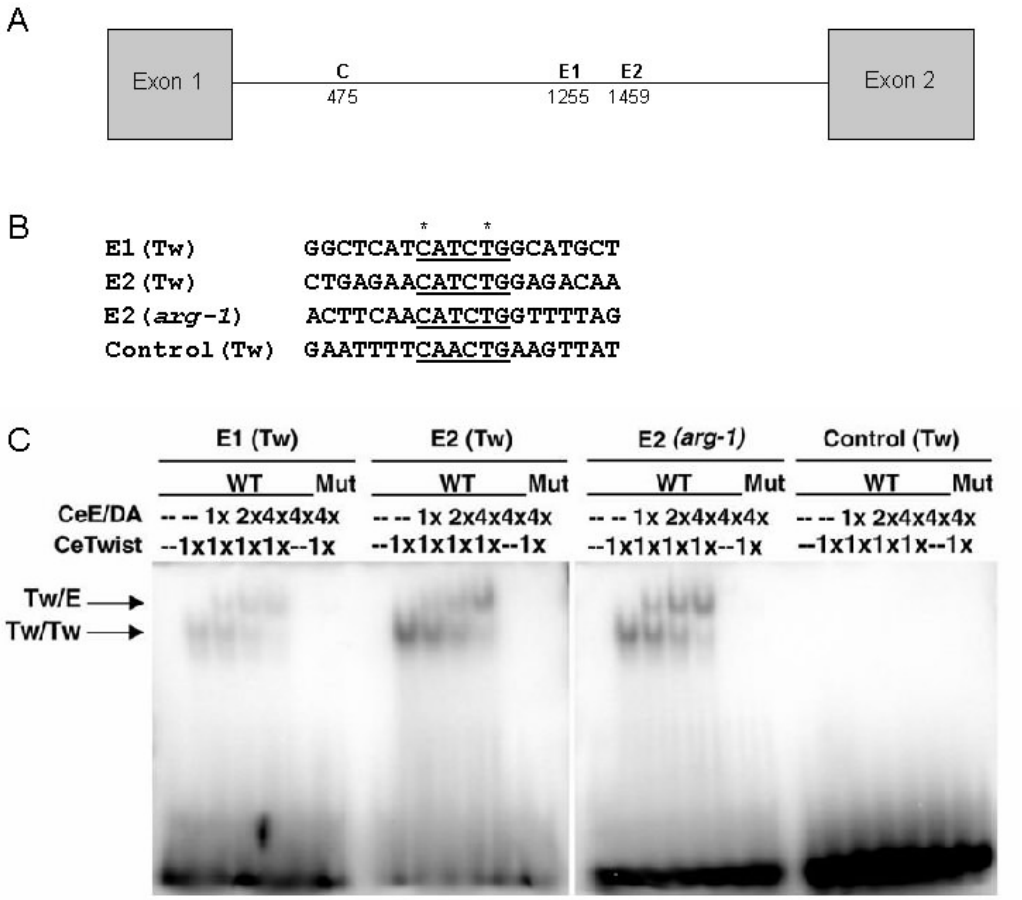


Figure 13. CeTwist dimers can bind to E1 and E2 *in vitro*. An Electrophoretic Mobility Shift Assay (EMSA) was done with CeTwist and its known binding partner CeE/DA. (A) Line drawing indicating the position of the 3 intron E boxes used in the EMSA. (B) 20mers used in shift as probes contain the E box (underlined) and 14 flanking nucleotides. E1 (Tw), E2 (Tw), and Control (Tw) are from *hlh-8* intron 1. E2 (*arg-1*) is from the promoter of *arg-1*, a downstream target of CeTwist. This E box is required for *arg-1* expression (Zhao et al., 2007). The asterisks indicate nucleotides that were mutated as a negative control similar to the previous SDM experiment in Fig. 4. (C) Native gel containing the radiolabeled probes in (B) plus varying amounts of purified protein. 2x indicates twice as much CeTwist protein was added to the reaction than in 1x and half as much as in 4x. Arrows point to the bands corresponding to CeTwist/CeE/DA heterodimers (top band) or CeTwist/CeTwist homodimers (bottom band). WT corresponds to using the probes in the unmutated form shown in (B). Mut lanes used probes with the E box changed from CANNTG to AANNAG. Twist dimers bound with higher affinity to E2 (Tw) than E1 (Tw) and did not bind to the Control (Tw) E box.

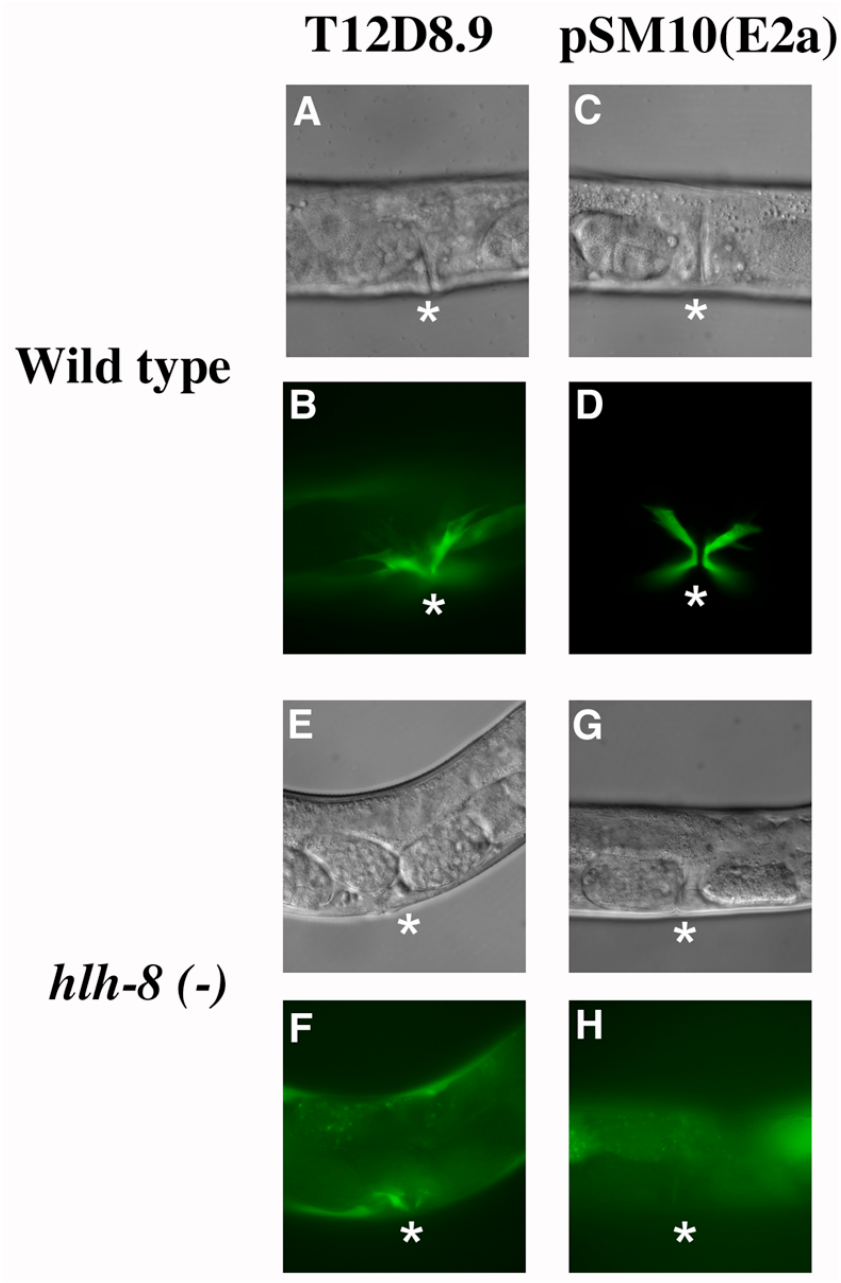


Figure 14. *hlh-8* is autoregulated through E2. Nomarski and GFP images of vulval region of (A-D) wild-type and (E-H) *hlh-8 (-)* animals with reporter constructs of (A, B, E, F) a non-target gene of CeTwist, T12D8.9 and (C, D, G, H) pSM10(E2a). GFP expression is lost in vulval muscle-like cells in *hlh-8 (-)* animals with pSM10(E2a) reporter in (H), but not with the T12D8.9 reporter in (F). Asterisks mark the vulval opening where vms or vm-like cells would be located.

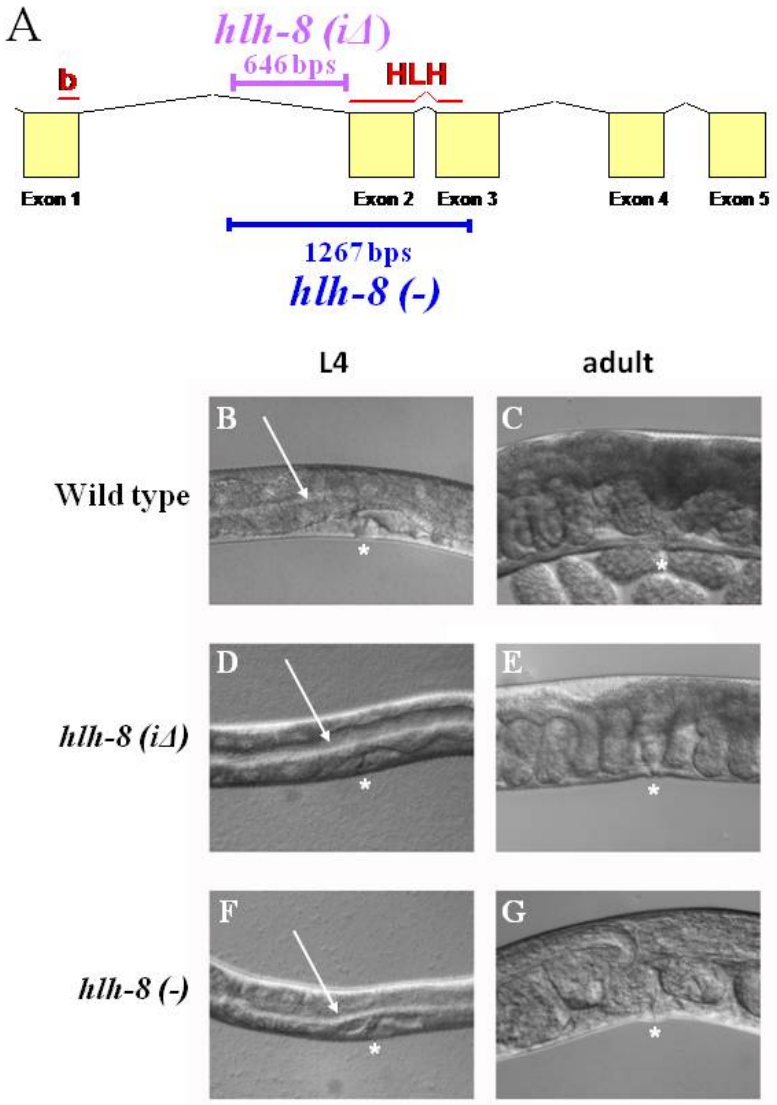


Figure 15. *hlh-8 (iΔ)* animals are constipated and partially egg-laying defective. (A) Line drawing of the genomic region of *hlh-8*. The regions that encode the basic and helix-loop-helix are indicated above. The nucleotides deleted in *hlh-8 (iΔ)* and the null allele *hlh-8 (-)* are indicated with purple and blue bars, respectively. (B-G) Nomarski images of the central region of animals with white asterisks marking the vulva location. (B, D, F) L4 larvae. Arrow indicates the lumen of the intestine. Note the expanded lumen in D and F. (C, E, G) Adults. (E) In *hlh-8 (iΔ)* animals, embryos are overlapping each other as they are backing up in the uterus. (G) In *hlh-8 (-)* animals developing late-stage embryos can be seen within the hermaphrodite. Those embryos have not been laid due to improper development of the sex muscles.

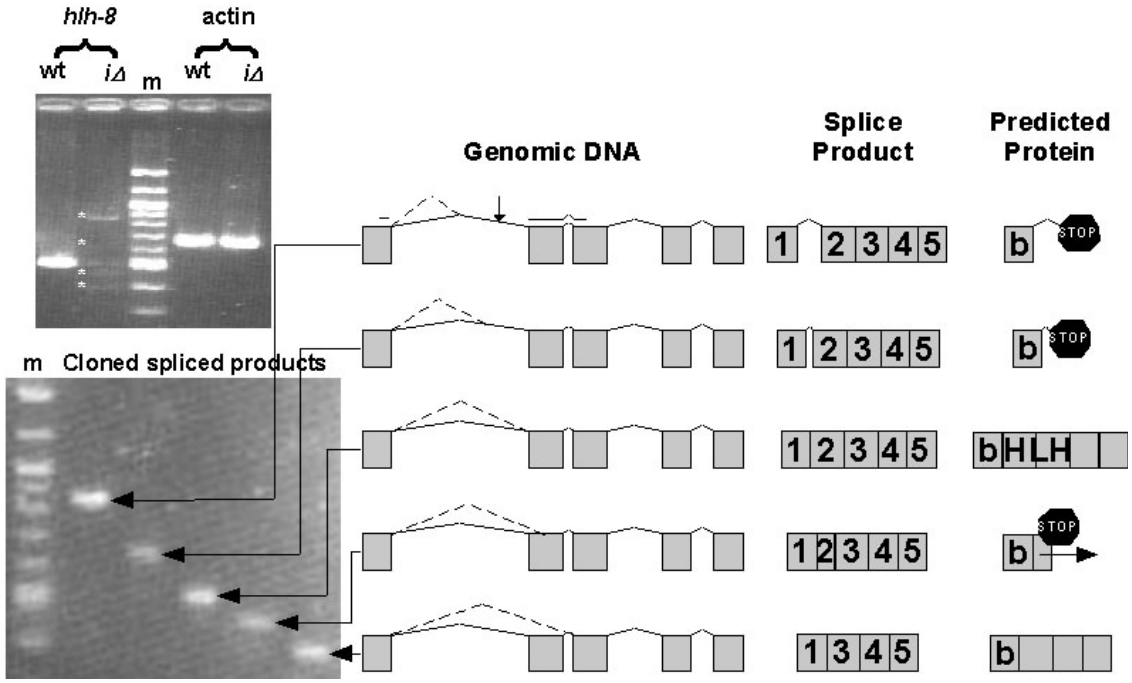


Figure 16. Splicing defects in *hlh-8 (iΔ)* animals. Reverse Transcription PCR (RT-PCR) revealed 4 alternate spliced products, marked with asterisks, of *hlh-8* in *hlh-8 (iΔ)* animals (Top Gel). The RT-PCR spliced fragments were cloned and sequenced to determine exact splicing variants. Individual clones were run out on a gel (Bottom Gel). The schematics on the right are drawn to scale. 'Genomic DNA' indicates where the sequencing results revealed splicing is occurring in each fragment by a dotted line. Exons are represented by gray boxes and introns by solid lines. A vertical arrow indicates the *hlh-8 (iΔ)* deletion 5' break point on the genomic DNA. 'Spliced Product' designates the various splice products determined from sequencing data. 'Predicted Protein' depicts the polypeptide that results from each of the spliced products. Stop signs indicate premature stop codons and a horizontal arrow indicates a frame shift due to splicing into an exon.

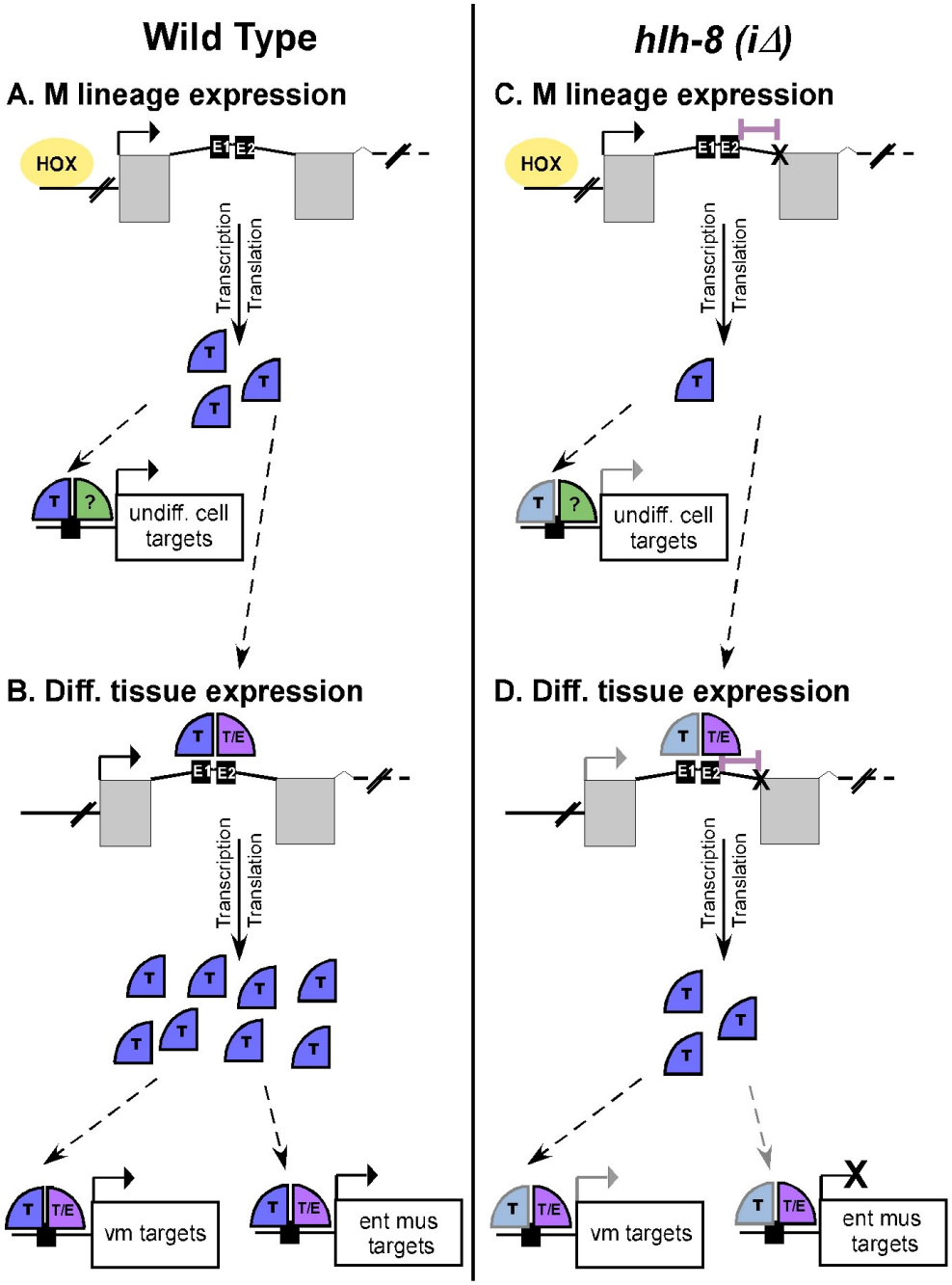


Figure 17. Model for the transcriptional regulation of *hlh-8*. (Left Panel) Wild-type animals. (A, C) Elements in the *hlh-8* promoter are bound by Hox Factors, which cause expression of CeTwist (T) earlier in development. A minimal concentration of CeTwist allows for unknown CeTwist dimer partners (?) to activate target genes in

undifferentiated cells (black arrow). (B, D) CeTwist Dimers, either homo- or CeE/DA heterodimers (T, T/E) made in step A are used for autoregulation through E1 and E2 intron elements, which results in an increased CeTwist expression. CeTwist homodimers and CeTwist/CeE/DA heterodimers are proposed to activate E1 and E2 elements based on *in vitro* EMSA and *in vivo hlh-8 (-)* and *hlh-2* RNAi data. A higher concentration of CeTwist would be required to reach the threshold to activate targets in developed vulval muscles (vm) and enteric muscles (ent mus). Proposed CeTwist homo- or CeE/DA heterodimers (T, T/E) activate vm targets. (Right Panel) *hlh-8 (iΔ)* animals. Purple bar above the *hlh-8* genomic schematic represents the region of intron 1 deleted in *hlh-8 (iΔ)* animals. Hox Factors are able to bind to the elements in the *hlh-8* promoter and activate transcription of the *hlh-8 (iΔ)* locus. However, due to the splicing defects (X), there is a decreased number of CeTwist molecules. The decrease in CeTwist molecules causes a decrease in undifferentiated cell target activation (grey arrow). The lower starting CeTwist concentration coupled with the deletion of the 3' nucleotides adjacent to E2 causes inefficient autoregulation. This results in a decrease in vm target activation as seen with the *gfp* target reporter constructs and the semi-Egl phenotype (grey arrow). No enteric muscle target *gfp* reporters were activated in *hlh-8 (iΔ)* animals (X). The hmc may fit into this model as being regulated similar to the vms. However, unlike the vms, no hmc target *gfp* reporters were activated in *hlh-8 (iΔ)* animals.

Table 1.
Insert size and DNA location used for the construction of the *hlh-8::gfp* plasmids.

Construct name	Insert size (bp)	Position of insert from the *hlh-8* first intron[a]
pSM1[b]	2019	103 to 2121
pSM2	1484	638 to 2121
pSM3	613	103 to 715
pSM4(*iΔ*1)	648	1474 to 2121
pSM5	960	638 to 1597
pSM6	66	2056 to 2121
pSM7(E1E2)	503	1095 to 1597
pSM8	1027	1095 to 2121
pSM9(E1a)	329	1095 to 1423
pSM10(E2a)	307	1291 to 1597
pSM11	133	1291 to 1423
pSM12	124	1474 to 1597
pSM13	126	1095 to 1220
pSM14(E1b)	163	1185 to 1347
pSM15(E2b)	163	1376 to 1538
pSM19	583	638 to 1220
pSM20(*iΔ*2)	183	1291 to 1473
pSM22	202	350 to 551
pSM24(E1mut)	503	1095 to 1597 (mut)[c]
pSM25(E2mut)	503	1095 to 1597 (mut)
pSM26(E1E2mut)	503	1095 to 1597 (mut)

[a] Base pair (bp) 1 refers to the first nucleotide of ATG in exon 1.
[b] pSM1 contains the entire sequence of intron 1.
[c] mut correlates with the specific E box sequence(s) being changed from CATCTG to AATCAG in the corresponding plasmid, with the remaining sequences of the plasmid staying the same.

Table 2.
The oligonucleotides used to construct the *hlh-8::gfp* plasmids.

Construct name	Primer name[a,d]	Primer sequence[b]
Preliminary PCR for all constructs[c]	AC-1	CGCGTGGTTCGAAAGAATGAAGTG
	AC-3	GTAGTGGAGGCAACTGTGAAGGTC
pSM1	AC-387 Sal I	ACGCGTCGACGTAAATGTTCATTC AATTGTTTTTGAGAGTTTCGGG
	AC-373	GAAGATCTGTAAACAAATCATAAT ACGAGTGATC
pSM2	AC-388 Sal I	CAACGCGTCGACCCCTTCATTGCA CGCTTTGG
	AC-373	GAAGATCTGTAAACAAATCATAAT ACGAGTGATC
pSM3	AC-387 Sal I	ACGCGTCGACGTAAATGTTCATTC AATTGTTTTTGAGAGTTTCGGG
	AC-389 Bgl II	GAAGATCTGCGGATTGATCTACTG AAACTGTACC
pSM4(*iΔ*1)	AC-371	CAACGCGTCGACTCAAAATATCAC AGGAATCAGAGAGGC
	AC-373	GAAGATCTGTAAACAAATCATAAT ACGAGTGATC
pSM5	AC-388 Sal I	CAACGCGTCGACCCCTTCATTGCA CGCTTTGG
	AC-372	GAAGATCTATCACTCATTTTCCCC ATC
pSM6	AC-374	CAACGCGTCGACAACATTATCTAC TGTTTATGCGCG
	AC-373	GAAGATCTGTAAACAAATCATAAT ACGAGTGATC
pSM7(E1E2)	AC-442 Sal I	CAACGCGTCGACGAACACGACTCT TTCGCATATCGC
	AC-372	GAAGATCTATCACTCATTTTCCCC ATC
pSM8	AC-442 Sal I	CAACGCGTCGACGAACACGACTCT TTCGCATATCGC
	AC-373	GAAGATCTGTAAACAAATCATAAT ACGAGTGATC
pSM9(E1a)	AC-442 Sal I	CAACGCGTCGACGAACACGACTCT TTCGCATATCGC

	AC-499 Bgl II	GAAGATCTGCCCCCCACCCTCTTC TCCCATCTCGCG
pSM10(E2a)	AC-498 Sal I	CAACGCGTCGACCTCCTAACGTCG CCTTCGTTTACGGC
	AC-372	GAAGATCTATCACTCATTTTCCCC ATC
pSM11	AC-498 Sal I	CAACGCGTCGACCTCCTAACGTCG CCTTCGTTTACGGC
	AC-499 Bgl II	GAAGATCTGCCCCCCACCCTCTTC TCCCATCTCGCG
pSM12	AC-371	CAACGCGTCGACTCAAAATATCAC AGGAATCAGAGAGGC
	AC-372	GAAGATCTATCACTCATTTTCCCC ATC
pSM13	AC-442 Sal I	CAACGCGTCGACGAACACGACTCT TTCGCATATCGC
	AC-525 Bgl II	GAAGATCTGTGTGTCGTCGAGCCT GGCGGTCAGGG
pSM14(E1b)	AC-526 Sal I	CAACGCGTCGACCACATATCACCC TGACCGCCA
	AC-527 Bgl II	GAAGATCTCGGAAGGAGTTGAGGA AGTGGTTTAATGTTAGCCG
pSM15(E2b)	AC-528 Sal I	CAACGCGTCGACCGGTAGAGGAAG ATGCAAGACGCGAG
	AC-529 Bgl II	GAAGATCTCGTCCGTGGAAAGAAG AAAAATAAG
pSM19	AC-388 Sal I	CAACGCGTCGACCCCTTCATTGCA CGCTTTGG
	AC-525 Bgl II	GAAGATCTGTGTGTCGTCGAGCCT GGCGGTCAGGG
pSM20(*iΔ*2)	AC-498 Sal I	CAACGCGTCGACCTCCTAACGTCG CCTTCGTTTACGGC
	AC-569 Bgl II	GAAGATCTCTAATTGTCTCCAGAT GTTCTCAGG
pSM22	AC-570 Sal I	CAACGCGTCGACCTGCGTACGACT TTCACAC
	AC-571 Bgl II	GAAGATCTCGTGTGAAGGGGATAT TTCATTGC
pSM24(E1mut)	AC-442 Sal I	CAACGCGTCGACGAACACGACTCT TTCGCATATCGC
	AC-372	GAAGATCTATCACTCATTTTCCCC ATC

pSM25(E2 mut)	AC-442 Sal I	CAACGCGTCGACGAACACGACTCT TTCGCATATCGC
	AC-372	GAAGATCTATCACTCATTTTCCCC ATC
pSM26(E1E2mut)	AC-442 Sal I	CAACGCGTCGACGAACACGACTCT TTCGCATATCGC
	AC-372	GAAGATCTATCACTCATTTTCCCC ATC

[a] Primer pairs used to make insert for corresponding plasmids are shown. The shaded region refers to the 3' primer. All primers are written in the 5' to 3' direction.

[b] Restriction sites are underlined. The *Sal*-I site is GTCGAC and the *Bgl*-II site corresponds to AGATCT.

[c] Primers from exon 1 and exon 3 were used in the first round of nested PCR and the product from this reaction was used as the template for the making of the insert DNA for the above constructs.

[d] The name of the primer is shown. However, primer nomenclature was changed in the middle of the experiment to indicate which restriction site is present on the primer.

Table 3.
pKKMCS primers for screening colonies and sequencing isolated constructs.

Primer name	Primer Sequence	Location
pKKMCS F	CACTCACAACGATGGATAC	pKKMCS vector 5' of Multiple Cloning Site
pKKMCS R	TAGACTGTGTGGAGACACTGC	pKKMCS vector 3' of Multiple Cloning Site

Table 4.
Primers used for Site-Directed Mutagenesis

Primer name[a]	Primer Sequence[b]
AC-574 SDM E1 F	GTTTATGGCTCATaATCaGGCATGCTCGCTAGCTCG
AC-575 SDM E1 R	CGAGCTAGCGAGCATGCCtGATtATGAGCCATAAAC
AC-576 SDM E2 F	GGTCCCTTCTGAATCCTGAGAAaATCaGGAGACAATTC
AC-577 SDM E2 R	GAAATTGTCTCCtGATtTTCTCAGGATTCAGAAGGGACC

[a] Primer name gives location of either E1 or E2 mutagenesis. The F and R of the primer name refers to the primer being in the forward or reverse direction with respect to the intron sequence, respectively. However, primers are written in the 5' to 3' direction.

[b] E box location is underlined and nucleotides that are changed are in lowercase. The canonical E box sequence is CANNTG and was changed through SDM to AANNAG.

Table 5.
Electrophoretic Mobility Shift Assay probes.

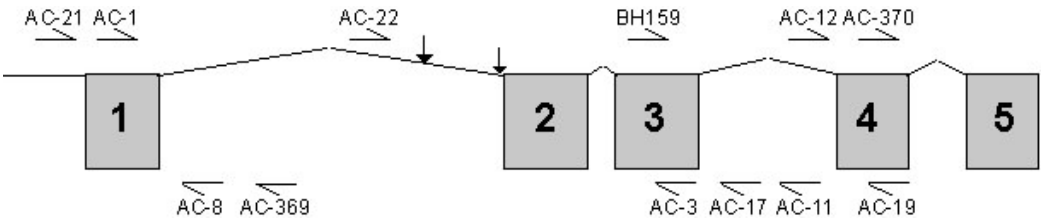
Probe name[a]	Primer name[b]	DNA sequence[c]
E1 (Tw) WT	AC-557 F EMSA	GGCTCATCATCTGGCATGCT
	AC-558 R EMSA	AGCATGCCAGATGATGAGCC
E1 (Tw) Mut	AC-565 F EMSA M	GGCTCATaATCaGGCATGCT
	AC-566 R EMSA M	AGCATGCCtGATtATGAGCC
E2 (Tw) WT	AC-559 F EMSA	CTGAGAACATCTGGAGACAA
	AC-560 R EMSA	TTGTCTCCAGATGTTCTCAG
E2 (Tw) Mut	AC-567 F EMSA M	CTGAGAAaATCaGGAGACAA
	AC-568 R EMSA M	TTGTCTCCtGATtTTCTCAG
E2 (*arg-1*) WT[d]	AC-544 F EMSA	ACTTCAACATCTGGTTTTAG
	AC-545 R EMSA	CTAAAACCAGATGTTGAAGT
E2 (*arg-1*) Mut	AC-546 F EMSA	ACTTCAAaATCaGGTTTTAG
	AC-547 R EMSA	CTAAAACCtGATtTTGAAGT
Control (Tw) WT	AC-561 F EMSA	GAATTTTCAACTGAAGTTAT
	AC-562 R EMSA	ATAACTTCAGTTGAAAATTC
Control (Tw) Mut	AC-563 F EMSA M	GAATTTTaAACaGAAGTTAT
	AC-564 R EMSA M	ATAACTTCtGATtAAAATTC

[a] Probe names indicate which E box is represented. Shaded probes correspond to the E box of the sequence being mutated from CANNTG to AANNAG.
[b] The F and R of the primer name refers to the primer being in the forward or reverse direction, respectively. M in the primer name refers to the specific mutated E box.
[c] E box of the probe pairs is underlined. Changed nucleotides of the mutated probes are in lowercase.
[d] Probes for *arg-1* were designed by J. Zhao (Zhao et al., 2007).

Table 6.
Primers used for *hlh-8 (iΔ)* confirmation via sequencing.

Primer name	Direction	Primer sequence	Location
AC-21	Forward	GCGGACCGCTGCAGAGATTCTTCG	5' Promoter
AC-1	Forward	CGCGTGGTTCGAAAGAATGAAGTG	Exon 1
AC-369	Reverse	GTGTGAAAGTCGTACGCAG	Intron 1
AC-8	Reverse	CGAGCTGCTGATCTAGGTTG	Intron 1
AC-22	Forward	GGAACTGTAGAGCCATTGCTTC	Intron 1
BH159	Forward	TACAATTTACAGTCGGCT	Exon 3
AC-3	Reverse	GTAGTGGAGGCAACTGTGAAGGTC	Exon 3
AC-12	Forward	GAAAATTCTCAGCTTCAGAG	Intron 3
AC-11	Reverse	GAAAGCGGAAAATGCTAAG	Intron 3
AC-17	Reverse	GAACATGTTCCAGCCTACGGAAC	Intron 3
AC-19	Reverse	GTTTTGGCGGCGGACAAGTCTG	Exon 4
AC-370	Forward	CAGACTTGTCCGCCGCCAAAAC	Exon 4

a

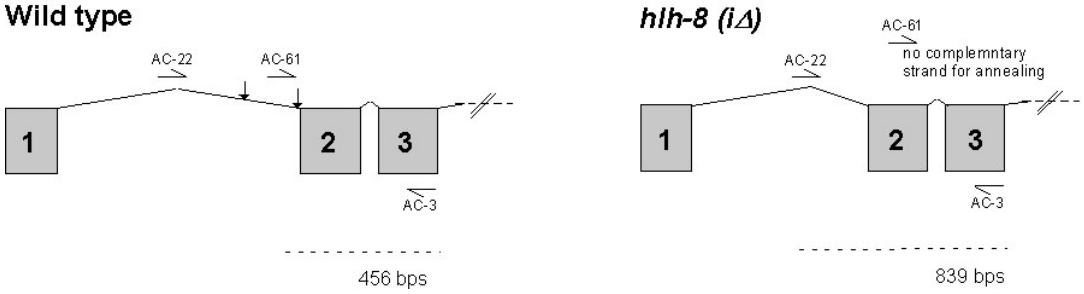


[a] Diagram of *hlh-8* genomic region illustrates the location of primers. Exons are shaded in gray and numbered. Vertical arrows identify the deletion break points of the *hlh-8 (iΔ)* locus. Half arrows show the approximate location of the primers and the arrows point in either the forward (pointing to the right, top of the diagram) or reverse (pointing to the left, bottom of the diagram) depending on which strand the primer was designed to complement.

Table 7.
Primers for *hlh-8 (iΔ)* confirmation through PCR.

Primer name	Direction	Primer Sequence	Location
AC22	Forward	GGAACTGTAGAGCCATTGCTTC	Intron 1
AC61	Forward	GCCGGAAATTGCCGAAAACTG	Intron 1
AC3	Reverse	GTAGTGGAGGCAACTGTGAAGGTC	Exon 3

a



[a] Diagram to illustrate location of primers in *hlh-8* genomic region. Wild type is shown on the left and the *hlh-8 (iΔ)* locus is on the right. Exons 1 to 3 are shown by grey boxes, the dotted line with hash marks indicates the entire gene is not shown. Half arrows indicate the location of specific primers. AC61 is located in the region of DNA that is deleted in *hlh-8 (iΔ)* animals (vertical arrows). Wild-type DNA will preferentially use AC61 and AC3 to make a product size 456 base pairs (dashed line, bottom of diagram, left) when the polymerase extension time is limited to prevent the formation of a 1102 base pair product use of AC22 and AC3. DNA extracted from homozygous *hlh-8 (iΔ)* can only be amplified using primers AC22 and AC3 to make a product of 839 base pairs (dashed line, bottom of diagram right).

Table 8.
Primers used for Reverse Transcriptase Polymerase Chain Reaction.

Primer name	Primer sequence[a]	Location[b]
AC-1	CGCGTGGTTCGAAAGAATGAAGTG	*hlh-8* Exon 1 Outside Primer
BH141	GAGCATGTGCCAACAGACGGGAACGTCAAAG GACCA	*hlh-8* Exon 1 Inside Primer
AC-400	GCCAGTGAATTGGAGTGAGTTG	*hlh-8* Exon 5 Outside Primer
AC-401	GGGATTTGGAGTTGAGATGGGACACAGTTC	*hlh-8* Exon 5 Inside Primer
AC-73	GTGACGACGAGGTTGCCGC	*act-4* Actin Primer
AC-74	CAGCTCCCGCTGTATCGTC	*act-4* Actin Primer
BH141	GAGCATGTGCCAACAGACGGGAACGTCAAAG GACCA	*hlh-8* Exon 1 sequencing
AC-401	GGGATTTGGAGTTGAGATGGGACACAGTTC	*hlh-8* Exon 5 sequencing
AC-499 Bgl II	GAAGATCTGCCCCCCACCCTCTTCTCCCATC TCGCG	*hlh-8* Intron 1 sequencing
BH143	GTAGTGTGTGGATCTTGCTCATCTTGTCTGA GGGCA	*hlh-8* Exon 2 sequencing
AC-3	GTAGTGGAGGCAACTGTGAAGGTC	*hlh-8* Exon 3 sequencing
AC-19	GTTTTGGCGGCGGACAAGTCTG	*hlh-8* Exon 4 sequencing

[a] Primers are written in the 5' to 3' direction.

[b] Nested PCR was performed to analyze the spliced products from the *hlh-8* locus in *hlh-8 (iΔ)* and wild-type animals. The outside primers were used in the first PCR amplification and the inside primers were used in the second round. Primers against *act-4*, a gene that encodes for actin, were used as a total RNA quantity control. Sequencing primers were used to sequence isolated spliced products from the *hlh-8* locus in *hlh-8 (iΔ)* animals.

Table 9.
GFP expression pattern of CeTwist downstream gene reporters and coelomocyte reporter in wild-type and *hlh-8* mutant animals.

Reporter[a]	Genotype	GFP pattern			
		VMs	Ent Mus	hmc	CC
arg-1::gfp					
	Wild type	+[c]	+	+	-
	hlh-8 (-)	-	-	-	
	hlh-8 (iΔ)	-	-	-	
egl-15::gfp					
	Wild type	+	-	-	-
	hlh-8 (-)	-			
	hlh-8 (iΔ)	+/-[d]			
Ndebox::gfp					
	Wild type	+	-	-	-
	hlh-8 (-)	-			
	hlh-8 (iΔ)	-			
Intrinsic cc::gfp					
	Wild type [b]	-	-	-	+
	hlh-8 (-)				+/-[e]
	hlh-8 (iΔ)				+

[a] Integrated reporters: *arg-1::gfp* and *egl-15::gfp* are downstream targets of *hlh-8*. *Ndebox::gfp* is a transcriptional reporter of *ceh-24*, also a downstream target of *hlh-8*. *Intrinsic cc::gfp* is expressed in all six coelomocytes including the two that arise from the M lineage.

[b] n value was over 100 for each category with the exception of Wild type (n=33) and *hlh-8 (iΔ)* (n=46) animals being scored with the *intrinsic cc::gfp* reporter.

[c] Symbols used: +, positive expression; -, no expression; +/-, non-wild-type expression

[d] 15% of animals expressed GFP; 74% of those turned off prematurely as compared to wild-type animals which had persistent *gfp* expression past day 2 of adulthood.

[e] 23% of animals did not have 6 coelomocytes as found in wild-type.

Table 10.
M lineage descendants in wild-type and *hlh-8* mutant animals.

Genotype	M lineage pattern in animals[a]		
	D/V division of M cell	2 SM-like cells	Division of SM-like cells
Wild type	100% (72)[b]	100% (22)	100% (26)
hlh-8 (-)	30% (98)	32% (38)	100% (27)
hlh-8 (iΔ)	79% (97)	48% (44)	92% (53)

[a] Animals expressed an integrated *hlh-8::gfp* that contained a promoter and no coding sequences.
[b] n values of animals scored are presented in parentheses.

REFERENCES

Alexander, N. R., Tran, L. N., Rekapally, H., Summers, C. E., Glackin, C. and Heimark, R. L. (2006). N-cadherin gene expression in prostate carcinoma is modulated by integrin-dependent nuclear translocation of Twist1. *Cancer Res.* **7**, 3365-3369.

Aroian, R. V., Levy A. D., Koga, M., Ohshima, Y., Kramer, J. M. and Sternberg, P. W. (1993). Splicing in *Caenorhabditis elegans* does not require an AG at the 3' splice acceptor site. *Mol. Cell Biol.* **13**, 626-637.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71-94.

***C. elegans* Sequencing Consortium.** (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012-2018.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805.

Chen, Z. F. and Behringer, R. R. (1995). *Twist* is required in head mesenchyme for cranial neural tube morphogenesis. *Genes Dev.* **9**, 686-699.

Connerney, J., Andreeva, V., Leshem, Y., Muentener, C., Mercado, M. A. and Spicer, D. B. (2006). Twist1 dimer selection regulated cranial suture patterning and fusion. *Dev. Dyn*. **235**, 1345-1357.

Corsi, A. K., Kostas, S. A., Fire, A. and Krause, M. (2000). *Caenorhabditis elegans* Twist plays an essential role in non-striated muscle development. *Development* **127**, 2041-2051.

Corsi, A. K., Brodigan, T. M., Jorgensen, E. M. and Krause, M. (2002). Characterization of a dominant negative *C. elegans* Twist mutant protein with implications for human Saethre-Chotzen syndrome. *Development* **129**, 2761-2772.

Culetto, E. and Sattelle, D. B. (2000). A role for *Caenorhabditis elegans* in understanding the function and interactions of human disease genes. *Hum. Mol. Genet.* **9**, 869-877.

Fackler, M. J., McVeigh, M., Evron, E., Garrett, E., Mehrotra, J., Polyak, K., Sukumar, S. and Argani, P. (2003). DNA methylation of *RASSF1A, HIN-1, RAR-β, Cyclin D2* and *Twist* in *in situ* and invasive lobular breast carcinoma. *Int. J. Cancer* **107**, 970-975.

Fidler, S. (2003). The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Reviews: Cancer* **3**, 1-6.

Fox, R. M., Watson, J. D., Von Stetina, S. E., McDermott, J., Brodigan, T. M., Fukushige, T., Krause, M. and Miller D. M. III. (2007). The embryonic muscle transcriptome of *Caenorhabditis elegans*. *Genome Biology* **8**, 188.

Gilbert, S. F. (2006). *Developmental Biology 8th ed.* Sinauer Associates, Inc: Sunderland, Mass.

Gort, E. H., Van Haaften, G., Groot, A. J., Plasterk, R. H., Shvarts, A., Suijkerbuijk, K. P., Van Laar, T., Van der Wall, E., Raman, V., Van Diest, P. J., Tijsterman, M. and Vooijs, M. (2007). The *TWIST1* oncogene is a direct target of hypoxia-inducible factor-2α. *Oncogene* 1-10.

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

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., Vas 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.

Ishii, M., Merrill, A. E., Chan, Y. S., Gitelman, I., Rice, D. P. and Sucov, H. M. (2003). Msx2 and Twist cooperatively control the development of the neural crest-derived skeletogenic mesenchyme of the murine skull vault. *Development* **130**, 6131-6142.

Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser A. G. and Ahringer, J. (2000). Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans. Genome Biology* **2**, 1-10.

Karp, X. and Greenwald, I. (2003). 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.* **17**, 3100-3111.

Karp, X. and Greenwald, I. (2004). Multiple roles for the E/Daughterless ortholog HLH-2 during *C. elegans* gonadogenesis. *Dev. Biol*. **272**, 460-469.

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

Kwok, W. K., Ling, M. T., Lee, T. W., Lau, T., Zhou, C., Zhang, X., Chua, C. W., Chan, K. W., Chan, F. L., Glackin, C., Wong, Y. C. and Wang, X. (2005). Up-regulation of TWIST in prostate cancer and its implication as a therapeutic target. *Cancer Res.* **65**, 5153-5162.

Leptin, M. and Grunewald, B. (1990). Cell shape changes during gastrulation in *Drosophila*. *Development* **110**, 73-84.

Lewis, J. A. and Fleming, J. T. (1995). Basic culture methods. In: Epstein, H. F., Shakes, D. C. (Eds.), Methods in Cell Biology. *Caenorhabditis elegans*: Modern Biological Analysis of an Organism. Academic Press, San Diego, pp. 4-29.

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.

Lodish, H., Berk, A., Zipursky, L. S., Matsudarira, P., Baltimore, D. and Darnell, J. (2000). Molecular Cell Biology 4th ed. New York:: W. H. Freeman.

Masui, T., Swift, G. H., Hale, M. A., Meredith, D. M., Johnson, J. E. and MacDonald, R. J. (2008). Transcriptional autoregulation controls pancreatic *Ptf1a* expression during development and adulthood. *Mol. Cell Biol.* **28**, 5458-5468.

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.

Peinado, H., Olmeda, D. and Cano, A. (2007). Snail, ZEB and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nature Reviews: Cancer* **7**, 415-428.

Plasterk, R. H. (1995). Reverse genetics: from gene sequence to mutant worm. In: Epstein, H. F., Shakes, D. C. (Eds), Methods in Cell Biology. *Caenorhabditis elegans:* Modern Biological Analysis of an Organism. Academic Press: San Diego, pp. 59-80.

Ratisoontorn, C. and Cunningham, M. L. (2007). Saethre-Chotzen Syndrome: Acrocephalosyndactyly Type III. *GeneReview*.

Rosivatz, E., Becker, I., Specht, K., Fricke, E., Luber, B., Busch, R., Höfler, H. and Becker, K. F. (2002). Differential expression of the epithelial-mesenchymal transition regulators Snail, SIP1, and Twist in gastric cancer. *Am. J. Pathol.* **161**, 1881-1891.

Sahlin, P., Windh, P., Lauritzen, C., Emanuelsson, M., Grönberg, H. and Stenman, G. (2007). Women with Saethre-Chotzen syndrome are at increased risk of breast cancer. *Genes, Chromosomes, and Cancer* **46**, 656-660.

Satoh, K., Hamada, S., Kimura, K., Kanno, A., Hirota, M., Umino, J., Fujibuchi, W., Masamune, A., Tanaka, N., Miura, K., Egawa, S., Motoi, F., Unno, M., Vonderhaar, B. K. and Shimosegawa, T. (2008). Up-regulation of MSX2 enhances the malignant phenotype and is associated with Twist 1 expression in human pancreatic cancer cells. *Am. J. Pathol.* **172**, 926-939.

Simpson, P. (1983). Maternal-zygotic gene interactions during formation to the dorsoventral pattern in *Drosophila* embryos. *Genetics* **105**, 615-632.

Stasinopoulos, I. A., Mironchik, Y., Raman, A., Wildes, F., Winnard, P. and Raman, V. (2005). HOXA5-Twist interaction alters p53 homeostasis in breast cancer cells. *J. Biol. Chem.* **280**, 2294-2299.

Stenirri, S., Restagno, G., Ferrero, G. B., Alaimo, G., Sbaiz, L., Mari, C., Genitori, L., Maurizio, F., and Cremonesi, L. (2007). Integrated strategy for fast and automated molecular characterization of genes involved in craniosynostosis. *Clinical Chemistry* **53:10,** 1767-1774.

Sulston, J. E. and Horvitz, H. R. (1977). Post-embryonic lineages of the nematode *Caenorhabditis elegans. Nucl. Acids Res.* **18**, 2033-2036.

Thisse, B., Messal, M. E. and Perrin-Schmitt, F. (1987). The twist gene: isolation of a *Drosophila* zygotic gene necessary for the establishment of dorsoventral pattern. *Nucl. Acids Res.* **15**, 3439-3453.

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, P., Zhao, J. and Corsi, A. K. (2006). Identification of novel target genes of CeTwist and CeE/DA. *Dev. Biol*. **293**, 486-498.

Wilkie, A. O. (1997). Craniosynostosis: genes and mechanisms. *Hum. Mol. Genet.* **6**, 1647-1656.

Wilson-Rawls, J., Rhee, J. M. and Rawls, A. (2004). Paraxis is a basic helix-loop-helix protein that positively regulates transcription through binding to specific E-box elements. *J. Biol. Chem*. **279**, 37685-37692.

Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A. and Weinberg, R. A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927-939.

Yang, J. and Weinberg, R. A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev. Cell*. **14**, 818-829.

Yang, M. H., Wu, M. Z., Chiou, S. H., Chen, P. M., Chang, S. Y., Liu, C. J., Tengs, S. C. and Wu, K. J. (2008). Direct regulation of *TWIST* by HIF-1α promotes metastasis. *Nature Cell Biology* **10**, 295-305.

Yanowitz, J. L., Shakir, M. A., Hedgecock, E., Hutter, H., Fire, A. and Lundquist, E. A. (2004). UNC-39, the *C. elegans* homolog of the human myotonic dystrophy-associated homeodomain protein Six5, regulates cell motility and differentiation. *Dev. Biol*. **272**, 389-402.

Zhao, J., Wang, P. and Corsi, A. K. (2007). The *C. elegans* Twist target gene, *arg-1,* is regulated by distinct E box promoter elements. *Mech. Dev*. **124**, 377-389.