UNIVERSITY OF CALGARY

Identifying transcriptional regulators of pharyngeal gene expression in *Caenorhabditis elegans*

by

Benjamin Shaun Grintuch

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

Calgary, Alberta

August, 2007

© Benjamin Shaun Grintuch 2007

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Determining transcriptional regulators of pharyngeal development in *Caenorhabditis elegans*" submitted by Benjamin Shaun Grintuch in partial fulfillment of the requirements for the degree of Master of Science.

_____

Supervisor, Dr. Jeb Gaudet
Department of Biochemistry and Molecular Biology

_____

Dr. Paul Mains
Department of Biochemistry and Molecular Biology

_____

Dr. William Brook
Department of Biochemistry and Molecular Biology

_____

Dr. Jillian Parboosingh
Department of Medical Genetics

_____

Date

Abstract

Organ development is a complex process that has not been thoroughly described. In *C. elegans* the FoxA transcription factor, PHA-4, functions as a "master regulator" of pharyngeal development. However, PHA-4 alone cannot account for the diversity of temporal and spatial patterns of gene expression in the pharynx. Thus, analysis of the promoter of the pharyngeal gene of *M05B5.2* should identify other regulators of pharyngeal development.

Analysis of the *M05B5.2* promoter identified four regulatory elements necessary for expression: PIG1 through PIG4 (Pharyngeal, Intestinal and Gonadal regulatory elements 1 through 4). PIG2-4 are each necessary for expression of *M05B5.2*, however none is sufficient to activate expression in enhancer assays. PIG1, in contrast, is both necessary and sufficient for expression in the pharynx. PIG4 does not resemble any known transcription factor binding site, while PIG3 appears to be a general element found in a diverse group of promoters and PIG2 is a perfect match to the binding site of the AP-1 class of bZIP transcription factors.

Interestingly, PIG1 contains both a PHA-4 binding site and an overlapping sequence that resembles a HNF6/ONECUT binding site. PIG1 functions as a negative regulatory element that inhibits hypodermal expression of *M05B5.2*. Enhancer assays also demonstrate that both the PHA-4 and ONECUT sites of the PIG1 element are required in overlap or in tandem to activate expression in the pharynx. The gene *ceh-39*, which encodes a predicted ONECUT factor, is necessary for normal activation through PIG1, suggesting that CEH-39 may act directly on PIG1. Thus, CEH-39, a novel pharyngeal regulator, was identified in this study.

Table of Contents

Approval Page.....ii

Abstract.....iii

Table of Contents.....iv

List of Tables.....vii

List of Figures.....viii

List of Abbreviations.....x

Introduction.....1

Transcriptional Regulation of Gene Expression.....2

Caenorhabditis elegans as a model system.....4

The pharynx of *C. elegans*.....6

Factors involved in pharyngeal specification, development and gene expression..7

PHA-4.....7

CEH-22.....9

PEB-1.....10

HLH-6.....10

TBX-2.....11

PHA-2.....11

Pharyngeal expression of M05B5.2 depends on factors in addition to PHA-4.....12

Materials and Methods.....20

General *Caenorhabditis elegans* maintenance and nomenclature.....20

Plasmid Construction.....21

Deletion analysis of M05B5.2.....21

Constructs created for mutation analysis.....21

Mutation of the GATA site.....21

Mutation of the PIG3 site.....22

Mutation of the PIG2 site.....22
Specific mutation of the PHA-4 site of PIG1.....22
Specific mutation of the ONECUT site of PIG1.....23
C. briggsae reporter.....23
Enhancer constructs.....24
Repressor assays with dpy-7 constructs.....25
Ectopic hypodermal expression construct.....25
Microinjections.....26
RNAi injection.....27
C. elegans lysis for PCR.....27
let-527 rescue.....28
Results.....34
RNAi to M05B5.2.....34
Analysis of the coding region of M05B5.2.....37
l*et-527* may be a mutation in M05B5.2.....38
Promoter analysis.....40
Expression of the homologous CBG21925 gene, in C. briggsae.....41
Deletion analysis.....42
Point mutations identify three discrete cis-regulatory elements.....44
Enhancer assays identify one sufficient regulatory site.....46
Inhibition of hypodermal constructs by PIG1 and its component sites.....48
Overexpression of M05B5.2 in hypodermal tissue.....49
Identifying the upstream regulators of M05B5.2 expression.....50
Identification of Candidate factors that may act through PIG4.....50
Identification of Candidate factors that may act through PIG3.....51
Identification of Candidate factors that may act through PIG2.....52
Identification of Candidate factors that may act through PIG1.....53
Forkhead factors.....53
HNF6/ONECUT factors.....54
The model of M05B5.2 regulation.....56

Discussion.....85

The Function of M05B5.2.....85

Regulation of expression of *M05B5.2*.....87

PIG1 is a complex regulatory element.....88

Other examples of overlapping binding sites.....90

ceh-39 encodes a ONECUT factor that may directly regulate PIG1.....92

Sufficiency of the PIG elements.....94

Conservation of regulation of foregut genes.....95

References.....97

List of Tables

Table 1. Worm strains used in this study.....29

Table 2. Description of the deletion plasmids and their construction.....30

Table 3. Primers used to create the enhancer constructs.....32

Table 4. The timing of the lethality in M05B5.2(RNAi) and *let-527*.....58

Table 5. A summary of the four *cis*-acting regulatory sequences.....59

Table 6. The bZIP genes of C. elegans.....60

Table 7. The forkhead factors of *C. elegans*.....61

Table 8. The method used to test the forkhead factors as inhibitory candidates.....62

Table 9. The ONECUT candidate factors in *C. elegans*.....63

List of Figures

Figure 1. The pharyngeal morphology.....15

Figure 2. The cells of the pharynx.....16

Figure 3. The lineage of the pharyngeal cells.....17

Figure 4. The regulators of pharyngeal specification, development and function.....18

Figure 5. The expression of *M05B5.2*.....19

Figure 6. Multimerization protocol.....33

Figure 7. The pharyngeal phenotypes due to injection of dsRNA to M05B5.2.....64

Figure 8. The transmembrane prediction and signal peptide prediction of M05B5.2.....65

Figure 9. The chromosomal region to which *let*-527 is mapped.....66

Figure 10. The expression of Cb-M05 in *C. briggsae*.....67

Figure 11. Expression of the 5` deletion constructs of *M05B5.2*.....68

Figure 12. Expression of the 3` deletion constructs of *M05B5.2*.....69

Figure 13. Expression of the *M05B5.2* reporters with Region I or PIG4 deleted.....70

Figure 14. The *M05B5.2*::*GFP*::*HIS2B* reporter showing both Region I and PIG4.....71

Figure 15. Expression of enhancer constructs of Regions I and PIG4.....72

Figure 16. Comparison between M05B5.2 and its homologous gene in *C. Briggsae*.....73

Figure 17. Mutation analysis of the conserved sites in the promoter of M05B5.2.....74

Figure 18. The expression patterns of the mutated PIG1-3 sites.....75

Figure 19. Comparison of the PIG1, PHA-4 and HNF6/ONECUT binding sites.....76

Figure 20. Expression of *M05B5.2* with mutations in the PHA-4 and ONECUT sites....77

Figure 21. Expression of PIG1 enhancer constructs 78

Figure 22. The expression of the *dpy-7* repressor constructs.....79

Figure 23. Reporter expression is not affected in *daf-4* mutant animals.....80

Figure 24. M05 expression is not affected in *atf-2* mutant animals.....81

Figure 25. Enhancer expression in *ceh-39* animals.....82

Figure 26. 5`Δ270M05 expression in *ceh-39* animals.....83

Figure 27. Transcriptional regulation of M05B5.2.....84

List of Abbreviations

atf	activating transcription factor family
BLAST	Basic Local Alignment Search Tool
bp	base pair
ceh	C. elegans homeobox
daf	abnormal dauer formation
DNA	DeoxyriboNucleic Acid
Dpy	Dumpy
DsRNA	double stranded RiboNucleic Acid
elt	erythroid-like transcription factor family
fkh	forkhead transcription factor family
FoxA	Forkhead box A
GFP	Green Fluorescent Protein
HIS2B	HIStone 2B
hlh	helix loop helix
HNF3	Hepatocyte Nuclear Factor
Kb	kilobase
let	lethal
myo	myosin heavy chains
peb	pharyngeal enhancer binding
PCR	Polymerase Chain Reaction
PFAM	Protein FAMily database
pha	defective pharynx development
PIG	Pharyngeal Intestinal and Gonadal regulatory element
rol	roller
PSI-BLAST	Position-Specific Iterative - Basic Local Alignment Search Tool
RNA	RiboNucleic Acid
RNAi	RiboNucleic Acid interference
tbx	t box family
Unc	Uncoordinated
YFP	Yellow Flourescent Protein

A	Adenosine
C	Cytidine
D	A, G or T
G	Guanine
H	A, C or T
R	puRine (A or G)
T	Thymidine
W	Nucleic acids with a Weak interaction (A or T)
Y	pYrimidine (C or T)

Introduction

Animal organs are composed of multiple cell types that must form coordinately in the right place and in the right sequence during development. This complex developmental process, organogenesis, requires very precise control to ensure proper organ formation and function. Previous work on organ development has found that this process requires a complex network of components, many of which encode transcription factors (Cereghini, 1996; Chen et al., 2006; Schrem, Klempnauer and Borlak, 2004). Parts of these regulatory pathways have been described in a few organs including the heart, eye and pancreas (Davis-Silberman and Ashery-Padan, 2007; Kume, 2005; Srivastava, 2006). The developmental pathways of many other organs have not been thoroughly characterized and we do not fully understand these pathways in those organs that have been previously studied. Therefore, our knowledge of organ development needs to be expanded to work out all of the processes involved in specification, differentiation and function of the different organs.

A full description of the developmental pathway(s) involved in foregut organ formation has been hindered by the complexity of organs such as the liver and the pancreas. The study of a relatively simple foregut organ, the pharynx of *Caenorhabditis elegans*, allows for a better understanding of these regulatory pathways. This organism has many characteristics that make it ideal for genetic and molecular studies to identify factors in its developmental network. Although simpler than other foregut organs, the pharynx still faces many of the same challenges that these more complex organs face. For example, during development, different pharyngeal cell types are specified that perform different functions in the organ. As well progenitor cells from different embryonic

lineages are incorporated into the adult organ. These challenges make the pharynx of *C. elegans* an ideal model to study the developmental pathway.

Transcriptional Regulation of Gene Expression

Eukaryotic genes encode a large variety of proteins that have many different functions. Interestingly though, the arrangement of the coding sequences and regulatory regions have been generally conserved. The regulatory regions of eukaryotic genes include a core promoter immediately adjacent to the transcriptional start site as well as a regulatory promoter upstream of the core promoter and the enhancer regions including the upstream promoter, introns and 3` untranslated regions (Garcia et al., 1986; Griffiths et al., 1999). The core promoter functions to bind and assemble the pre-initiation complex including RNA polymerase II (Chandler et al., 1983; Griffiths et al., 1999; Pelham, 1982). The main function of this core promoter is to ensure that the RNA polymerase II enzyme transcribes from the correct start site in the correct orientation (Breathnach and Chambon, 1981). The core promoter does not function to determine which genes are expressed but to respond to transcription factors that are recruited by the regulatory promoters and enhancer regions (Griffiths et al., 1999).

The regulatory promoter is generally located very close to the core promoter, much closer than the enhancer regions, which can be found more than 50 kb away from the core promoter (Muller et al., 1988). The regulatory promoter contains *cis*-acting elements that bind proteins to facilitate the binding of RNA polymerase II (Griffiths et al., 1999). In comparison the enhancer regions contain *cis*-acting elements that bind other proteins known as transcription factors that either enhance or silence expression of the

gene from a distance. Enhancers are *cis*-acting sequences that are bound by activating transcription factors that can greatly increase the transcription rate of a gene or multiple genes that have the same regulatory regions (Griffiths et al., 1999). Silencers are *cis*-acting elements that are bound by repressive transcription factors that inhibit gene expression by reducing transcription of a gene or multiple genes that have the same regulatory regions (Griffiths et al., 1999).

There are two main models that have been developed to explain the interaction between transcription factors that bind to the enhancers and silencers possibly far from the transcriptional start site and the basal promoter. The first model, the looping model, suggests that DNA between the enhancer or silencer and the core promoter loops out to allow for a direct interaction between these regulatory regions and the core promoter. It is likely that the transcription factors bound to the enhancers and silencers interact with the primed basal transcription machinery at the core promoter to initiate or inhibit transcription (Griffiths et al., 1999; Tolhuis et al., 2002). The second model, the tracking model, proposes that the transcription factors that bind to the enhancers and silencers form a complex that migrates (tracks) towards the core promoter to interact with the basal transcriptional machinery to activate or inhibit transcription (Blackwood and Kadonga, 1998; Griffiths et al., 1999). However, it is unknown how the transcriptional complex would track along the regulatory regions to the core promoter region.

Identification of the *cis*-acting elements within the regulatory regions of a gene will enable one to determine the necessary sites that activate or inhibit expression of a specific gene. The discovery of these *cis*-acting regions will also allow one to identify the

transcription factors that bind to the regulatory regions of a gene to determine the inputs necessary to modulate expression of the gene.

C. elegans as a model system

C. elegans have many features that make it ideal for genetic and molecular studies. The features include a short life span, transparent body and organs and both hermaphroditic and dioecious modes of reproduction. As well, the physiology, anatomy, genome and cell lineage of this nematode have been well characterized, making it an ideal model organism for developmental research.

The easy handling and fast generation time of *C. elegans* make this organism ideal for laboratory research. Adult *C. elegans* are only about 1 mm long and can be fed on lawns of *E. coli* on petri dishes, making them very easy to care for and store in a laboratory. These animals have a very short life cycle of four days when grown at 20 °C . Embryogenesis of *C. elegans* takes about 14 hours from fertilization to hatching of the first larval stage. This process follows a specific pattern of cell division and development to give rise to a 558 cell larval hermaphrodite (Wood, 1988a). This larva then develops through four larval stages, L1-L4, with a molt punctuating the transition between these different stages. After the L4 molt, the worm is a fertile adult containing fewer than 1000 somatic cells. *C. elegans* are transparent and individual cells can be readily visualized using NDIC (Nomarski Differential Interference Contrast) optics. This allows for the study of development at the level of individual cells .

C. elegans has a relatively simple anatomy that makes it useful for the study of organ development (reviewed by Wood, 1988b). The outer layer of this organism is made

up of an extracellular cuticle that is secreted by the underlying hypodermal cells. Four strips of body-wall muscles run the length of this animal, two ventrally and two dorsally, for its locomotion. The simple nervous system of *C. elegans* is made up of just 302 cells, with most of them found surrounding the pharynx, in the dorsal and ventral nerve cords and in the tail. The bilobed pharynx found at the anterior end of this animal pumps and crushes food as it passes it along to the intestine. The gonad of the hermaphrodite is made up of two U-shaped arms in which a syncytium of germ line nuclei develop to pass through the spermatheca for fertilization before being excreted through the vulva. In contrast, the male gonad is made up of a single U-shaped structure in the body and has a specialized tail for mating.

The short life cycle and self-fertilization of these worms make them well suited to genetic analysis. The number of progeny that *C. elegans* can produce is restricted by the amount of sperm; with unmated hermaphrodites able to lay about 300 fertilized eggs compared to the more than 1000 fertilized eggs that a mated hermaphrodite can produce (Wood, 1988b). Another benefit to analysis of this animal is that its genome has been sequenced and is annotated online for public access through Wormbase (www.wormbase.org). The genomes of two related nematodes, *C. briggsae* and *C. remanei*, have also been sequenced and annotated online at Wormbase (www.wormbase.org). The sequencing and annotation of the genomes of these three closely related nematodes allows for fast and easy comparison, which can be readily used for phylogenetic footprinting.

The pharynx of *C. elegans*

The morphology and function of the pharynx of *C. elegans* has been well characterized. The pharynx is a neuromuscular organ that draws in food from the environment and transports it to the intestine (Figure 1; Albertson and Thompson, 1976). The pharynx initiates digestion through mechanical grinding in the posterior bulb prior to food entering the intestine. The pharyngeal-intestinal valve separates the pharynx and the intestine.

The pharynx is a simple foregut organ that is made up of only 80 cells in adults. However, similar to more complex foregut organs like the liver and pancreas, the pharynx is made up of multiple cell types that perform different functions: muscle, neuronal, marginal, epithelial and gland cells (Figures 2; Albertson and Thomson, 1976; Lavon and Benvenisty, 2005). Also, similar to more complex foregut organs like the pancreas, pharyngeal cells arise from different cell lineages (Sulston et al., 1983; Murtaugh and Melton, 2003). In general, the anterior pharynx is derived from descendants of the ABa blastomere while the posterior pharynx is derived from descendents of the MS blastomere (Figure 3).

By the end of gastrulation, approximately 330 minutes after fertilization, 78 of the 80 pharyngeal cells are already born and contained within a ball-like primordium (Albertson and Thompson, 1976). Shortly after gastrulation, with the exception of the connection between the pharynx and the mouth, the pharyngeal cells become encapsulated in the specialized basal lamina that partitions them from the rest of the embryo and throughout their life (Portereiko and Mango, 2001).

Factors involved in pharyngeal specification, development and gene expression

The intricacy of the *C. elegans* pharynx suggests that its development requires a complex network of transcription factors, a few of which are known. The cells of the pharynx are specified from two different embryonic lineages to differentiate into five different cell types, requiring regulatory factors to specify pharyngeal cells at the correct developmental stage and to ensure each cell develops into the correct type in its proper place and orientation. The pharyngeal transcription factors that have been identified include *pha-4, ceh-22, peb-1, hlh-6, tbx-2* and *pha-2* (Figure 4). However, these factors do not account for the specification and differentiation of all pharyngeal cells and thus, other factors in the pharyngeal regulatory network are yet to be identified.

PHA-4

The specification of pharyngeal cell fate requires PHA-4, the *C. elegans* homolog of the HNF-3/*forkhead* transcription factor. *pha-4* is expressed very early in the digestive tract during embryogenesis and throughout the adult life of this animal (Azzaria et al., 1996; Gaudet and Mango, 2002). Inactivation of this gene in *C. elegans* leads to a loss of foregut cells and a conversion of these cells into ectodermal cell types (Horner et al., 1998; Mango, Lambie and Kimble, 1994). Overexpression of *pha-4* is sufficient to convert otherwise non-pharyngeal cells to a pharyngeal fate (Horner et al., 1998). As well as being necessary and sufficient for pharyngeal cell fate, this gene is necessary for normal development and function of the pharynx. The loss of PHA-4 activity at any point in the life of *C. elegans* results in lethality, which suggests that PHA-4 is required not

only for pharyngeal specification during embryogenesis but surprisingly also throughout the whole life of the animal (Gaudet and Mango, 2002).

HNF-3/FoxA2 binds to DNA through the forkhead/winged helix domain (Clark et al., 1993). The consensus-binding site for this DNA binding domain as determined by Overdier, Porcella and Cost (1994) is TRTTKRY (R = A or G, K = T or G and Y = T or C). PHA-4 has also been shown to bind *in vitro* to at least a subset of possible TRTTKRY sequences, suggesting that PHA-4 and mammalian FoxA have similar binding site preferences (Gaudet and Mango, 2002).

Gaudet and Mango (2002) also found that it is possible that different affinities of the PHA-4 binding site may have an effect on the timing of target gene expression, with higher affinity resulting in earlier expression. However, this affinity binding theory does not explain the entire pharyngeal gene expression pattern. For example, the pharyngeal muscle gene *myo-2* is expected by the affinity-binding hypothesis to express early in development, whereas it is in fact activated relatively late in development (Gaudet and Mango, 2002; Okkema et al., 1993). Further research has found that other DNA regulatory elements may be responsible for early and late expression of pharyngeal genes, suggesting that it is more than simply affinity binding that is responsible for temporal expression (Gaudet et al., 2004).

Interestingly, the *pha-4* gene is expressed in more than just the pharynx. It is also expressed in the rest of the digestive tract, including the intestine and hindgut, and in the adult somatic gonad (Azzaria et al., 1996). The mutation of *pha-4* leads to a loss of pharynx, which leads to the developmental arrest of *C. elegans* embryos so that the role of PHA-4 in the rest of the digestive tract or the somatic gonad has not been examined.

Recent data has found that PHA-4 may have a function outside of pharyngeal development. This gene may play a role in diet-restriction mediated longevity of *C. elegans* (Panowski et al., 2007).

Because PHA-4 is expressed in all pharyngeal cells, it must work in concert with other factors to specify the different cell types within the pharynx. I will now discuss the other factors that have been identified to affect the development of the other pharyngeal cell types.

CEH-22

ceh-22, a homeobox gene homologous to Drosophila *tinman* and vertebrate Nkx2.5, has a clear role in activating the expression of the pharyngeal muscle gene *myo-2* (Okkema and Fire, 1994; Okkema et al., 1997). c*eh-22* is a downstream target of PHA-4 that is expressed in most of the pharyngeal muscle cells starting early during embryogenesis (Okkema and Fire, 1994). CEH-22 is not required for muscle cell differentiation, however the loss of this protein leads to weaker than normal muscle contraction and a thinner than normal pharynx (Okkema et al., 1997). Animals that lack CEH-22 frequently arrest as L1 and die, likely from starvation (Okkema et al., 1997). Ectopic expression of CEH-22 in the pharynx is able to convert pharyngeal cells to pharyngeal muscle cells. However, ectopic expression of CEH-22 outside of the pharynx does not affect cell specification, suggesting that more than CEH-22 is required to specify the pharyngeal muscle fate (Okkema et al., 1997).

PEB-1

peb-1 is expressed throughout the pharynx in all cell types except neurons as well as in non-pharyngeal cells including the hypodermis, hindgut and vulva (Fernandez, Gibbons and Okkema, 2004; Thatcher et al., 2001). The loss of *peb-1* by RNAi (RNA interference) or loss of function mutants suggests that this gene is required for normal molting and feeding of the animal possibly caused by defective pharyngeal gland cells (Fernandez, Gibbons and Okkema, 2004). This protein appears to inhibit PHA-4 binding to the *myo-2* promoter *in vitro* suggesting that PEB-1 inhibits PHA-4's ability to activate transcription during the formation of the *C. elegans* pharynx (Kalb et al., 2002). PEB-1 is not required for differentiation of the different pharyngeal cell types, suggesting that it is not necessary for normal pharyngeal specification but instead for pharyngeal function (Fernandez, Gibbons and Okkema, 2004). The loss of this factor also leads to defects in the other tissues in which it is expressed.

HLH-6

A basic Helix-Loop-Helix transcription factor, HLH-6, has been identified as a regulator of a sub-set of pharyngeal gland genes and is likely necessary for normal gland cell differentiation (Smit and Gaudet, in preparation). HLH-6 is directly regulated by PHA-4 and LAG-1, the effector of Notch signaling in *C. elegans* (Raharjo and Gaudet, 2007; Vikas Ghai, personal communication). The loss of HLH-6 leads to a loss of some, but not all, pharyngeal glands, while the remaining glands have an abnormal appearance. Therefore, HLH-6 functions in specification and/or differentiation of pharyngeal glands (Smit and Gaudet, in preparation). HLH-6 is also likely required for protection from

bacteria as *hlh-6* mutants die when fed live bacteria but not when they are fed an abiotic media (Smit and Gaudet, in preparation).

TBX-2

The T-box factor TBX-2 is required for the formation of the ABa-derived pharyngeal muscle cells (Chowdhuri et al., 2006; Smith and Mango, 2007). *tbx-2* mutants lack pharyngeal muscle cells derived from the ABa blastomere, whereas the muscle cells derived from the MS blastomere and ABa derived pharyngeal marginal cells are still present (Chowdhuri et al., 2006). This gene is expressed early in pharyngeal precursor cells at around the 100-cell stage, while later expression is seen in body-wall muscle and pharyngeal neurons (Chowdhuri et al., 2006). Interestingly, the ABa pharyngeal muscle cells appear to be correctly specified during early embryogenesis. However, these cells appear to arrest during later development and do not correctly form pharyngeal muscle cells (Smith and Mango, 2007). *tbx-2* expression requires PHA-4 activation and possible repression by components of *glp-1*/Notch signaling (Smith and Mango, 2007). Thus, it is likely that ABa derived pharyngeal muscle cells require both PHA-4 and TBX-2 for proper specification of the pharyngeal muscle fate.

PHA-2

The gene *pha-2* was first identified in mutant screens by Avery (1993) as animals that have a feeding defect. These animals showed a thicker and shorter than normal pharyngeal isthmus. Further study by Morck et al. (2004) cloned *pha-2* and showed that it encoded a homeobox transcription factor homologous to vertebrate Hex. Expression of

PHA-2 begins very early in the pharyngeal primordium and it is expressed primarily in pm5 pharyngeal muscle cell precursors and then adult pm5 cells; pm5 cells are the cells that make up the pharyngeal isthmus (Morck et al., 2004). Thus it is likely that the loss of PHA-2 leads to a defect in pm5, likely an elongation or cell-fate defect, which in turn causes the morphological defects in animals lacking this gene. *pha-2* is auto regulated and activated by PHA-4, similar to the other pharyngeal factors (Morck et al., 2004).

Previous research has identified three muscle specific transcription factors (TBX-2, PHA-2 and CEH-22) one gland specific transcription factor (HLH-6) and two broadly-acting pharyngeal transcription factors, PHA-4 and PEB-1 (Figure 4). However, factors involved in the development of other pharyngeal cell types have not been identified, including epithelial, marginal and neuronal cells. As well, all of the factors involved in the different temporal and spatial expression patterns of pharyngeal genes have not been identified. Therefore, further study of regulation of other pharyngeal genes should help to fully describe the factors involved in *C. elegans* foregut organogenesis.

Pharyngeal expression of *M05B5.2* depends on factors in addition to PHA-4

PHA-4 is required for the expression of most pharyngeal genes; however, the removal of PHA-4 binding sites from the promoters of at least two genes, *myo-2* and *M05B5.2*, does not completely abolish pharyngeal expression (Gaudet and Mango, 2002). The promoter of *myo-2* has been thoroughly studied, which has resulted in the identification of multiple regulatory sites and two regulatory transcription factors, CEH-22 and PEB-1. Further characterization of *M05B5.2*, which shows an expression pattern

differing from that of *myo-2*, should identify other regulators of pharyngeal gene expression.

A reporter construct for *M05B5.2* is expressed throughout the pharynx beginning shortly after specification of pharyngeal precursors but before formation of the primordium and continuing through adulthood (Figure 5; Gaudet and Mango, 2002). This reporter is also expressed in the rest of the digestive tract and the somatic gonad, which is the same pattern of expression as PHA-4. However, removal of the PHA-4 sites does not abolish expression of this gene, suggesting that other regulators are activating expression of *M05B5.2*.

Other evidence also confirms the pharyngeal expression of *M05B5.2*. The Nematode Expression Database (http://nematode.lab.nig.ac.jp) is an online resource that attempts to characterize the expression of all annotated *C. elegans* genes through *in situ* hybridization. *in situ* hybridization for M05B5.2 has been studied in embryos but not adults, and indicates expression of the gene in pharyngeal cells of developing embryos (Figure 5). A microarray experiment by Gaudet and Mango (2002) has also found that this gene is probably expressed in the pharynx. Multiple lines of evidence have shown that this gene is expressed in the pharynx of *C. elegans*, suggesting that further study should find novel regulators of pharyngeal gene expression.

Interestingly, some unpublished analysis of the *M05B5.2* promoter by Jeb Gaudet suggests that there are both positive and negative regulators acting on this gene that should be identifiable through further analysis. The mutation of a specific PHA-4 site in this promoter results in ectopic expression in the hypodermis while leaving the expression of this gene in the digestive tract and somatic gonad unaffected (Figure 5).

This suggests that a negative regulator present in the hypodermis is restricting the expression of this gene to achieve the normal pattern of expression. Therefore, further analysis of the promoter of this gene should identify both positive and negative regulatory elements required for normal expression of *M05B5.2* in *C. elegans.*

In this study I characterized the promoter of *M05B5.2* to identify the regulatory inputs of this gene that are necessary to drive its specific expression pattern. Through this analysis I have found four critical regulatory sites, PIG1-4 (Pharyngeal Intestinal and Gonadal regulatory elements 1-4) that control expression of *M05B5.2*. Further study of these regulatory elements identified a *trans*-acting factor, CEH-39, that activates expression of *M05B5.2* through the PIG1 site.

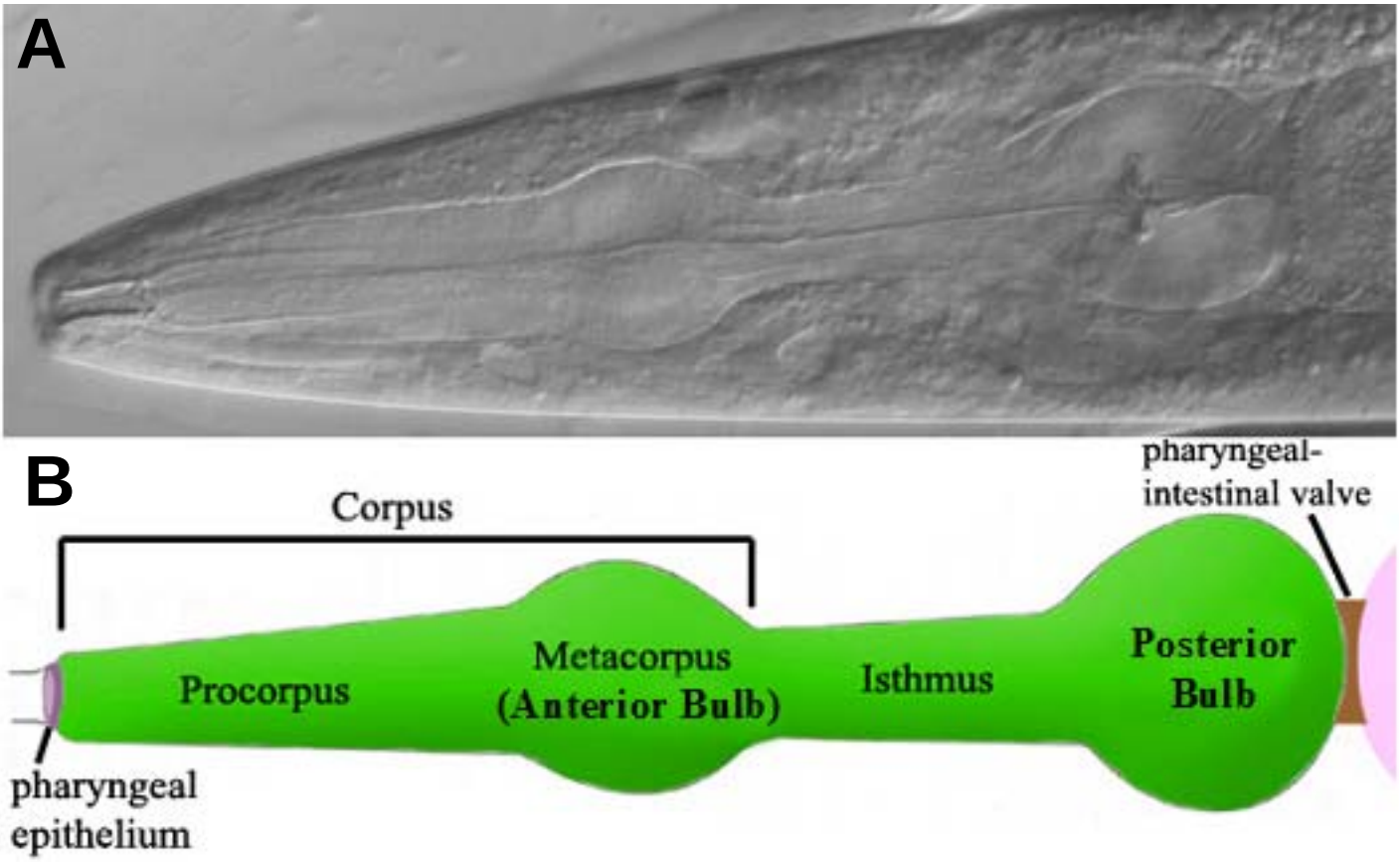


Figure 1. The pharynx of *C. elegans*. A) A Nomarski Differential Interference Contrast image of the pharynx. B) A cartoon drawing of the pharynx showing the different regions of this organ. Figures from Altun & Hall [2005].

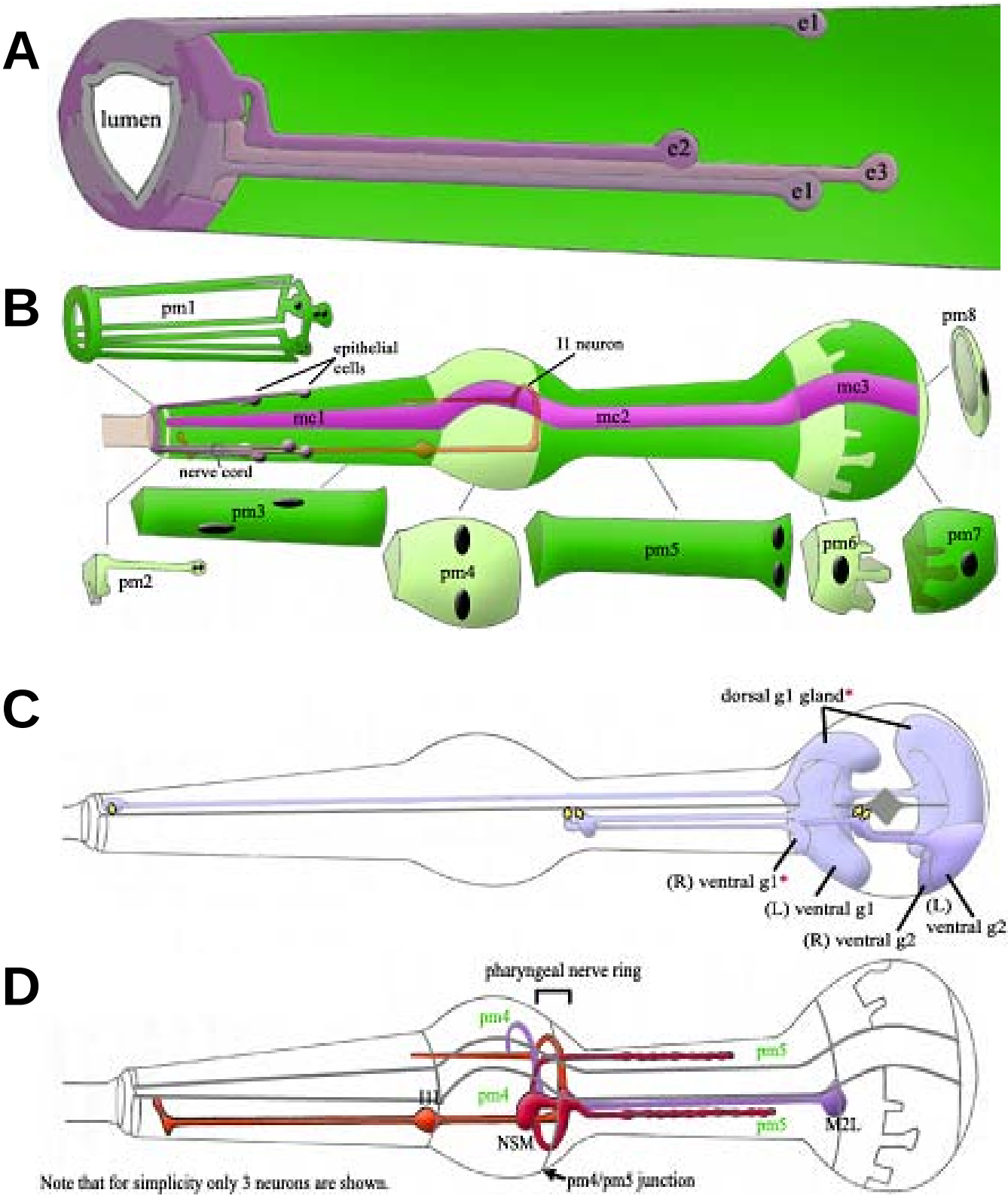


Figure 2. The cells of the pharynx. A) The epithelial cells e1-3. B) The marginal, mc1-3 and muscle cells, pm1-8. C) The gland cells g1-2. D) Some of the pharyngeal neurons, IL1, NSM, M2L. Figures from Altun and Hall [2005].

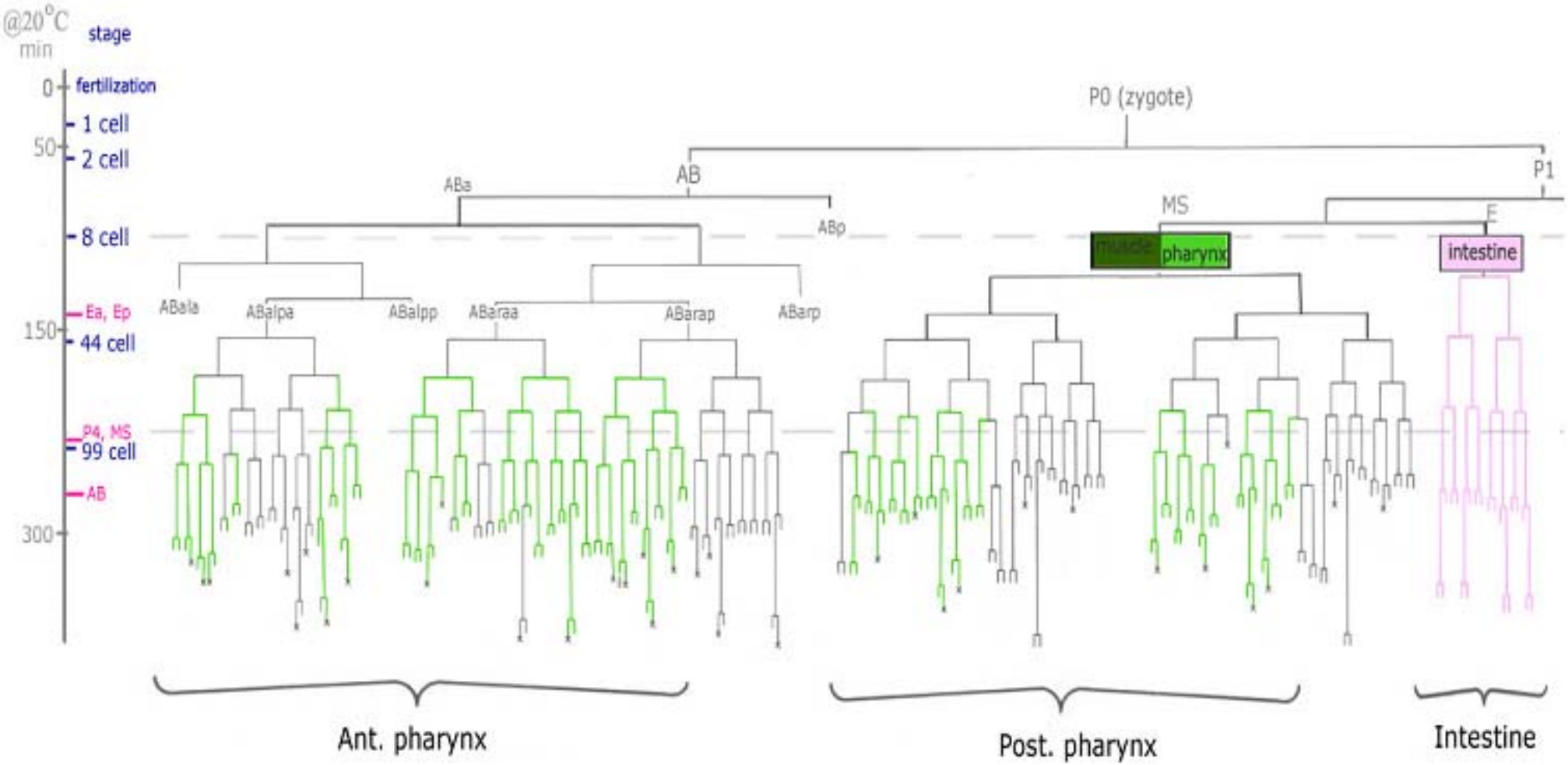


Figure 3. The lineage of the pharyngeal cells. The pharyngeal cells are highlighted green. The blue writing indicates the stage at which cells are born and the pink writing indicates when the cells migrate into the embryo. Figure from Altun & Hall [2005].

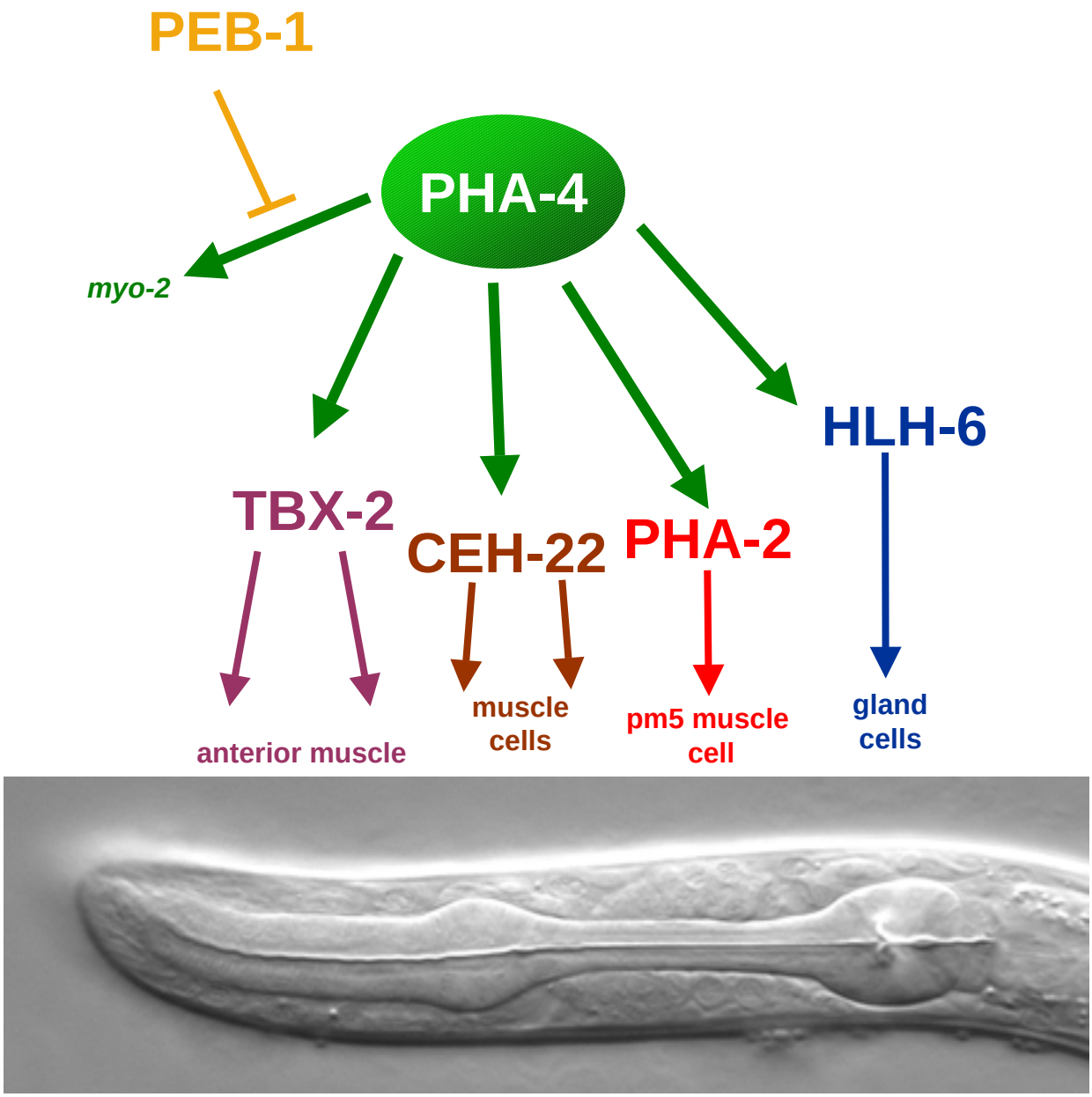


Figure 4. The known regulators of pharyngeal specification, development and function. This cartoon represents the known developmental regulators and their interactions. The organ identity factor PHA-4 regulates CEH-22, PHA-2, TBX-2 and HLH-6 expression. CEH-22, PHA-2 and TBX-2 are all required for normal muscle cell development and specification while HLH-6 is required for normal gland cell development and specification. PEB-1 likely regulates pharyngeal gene expression in the epithelial, marginal, gland and muscle cells.

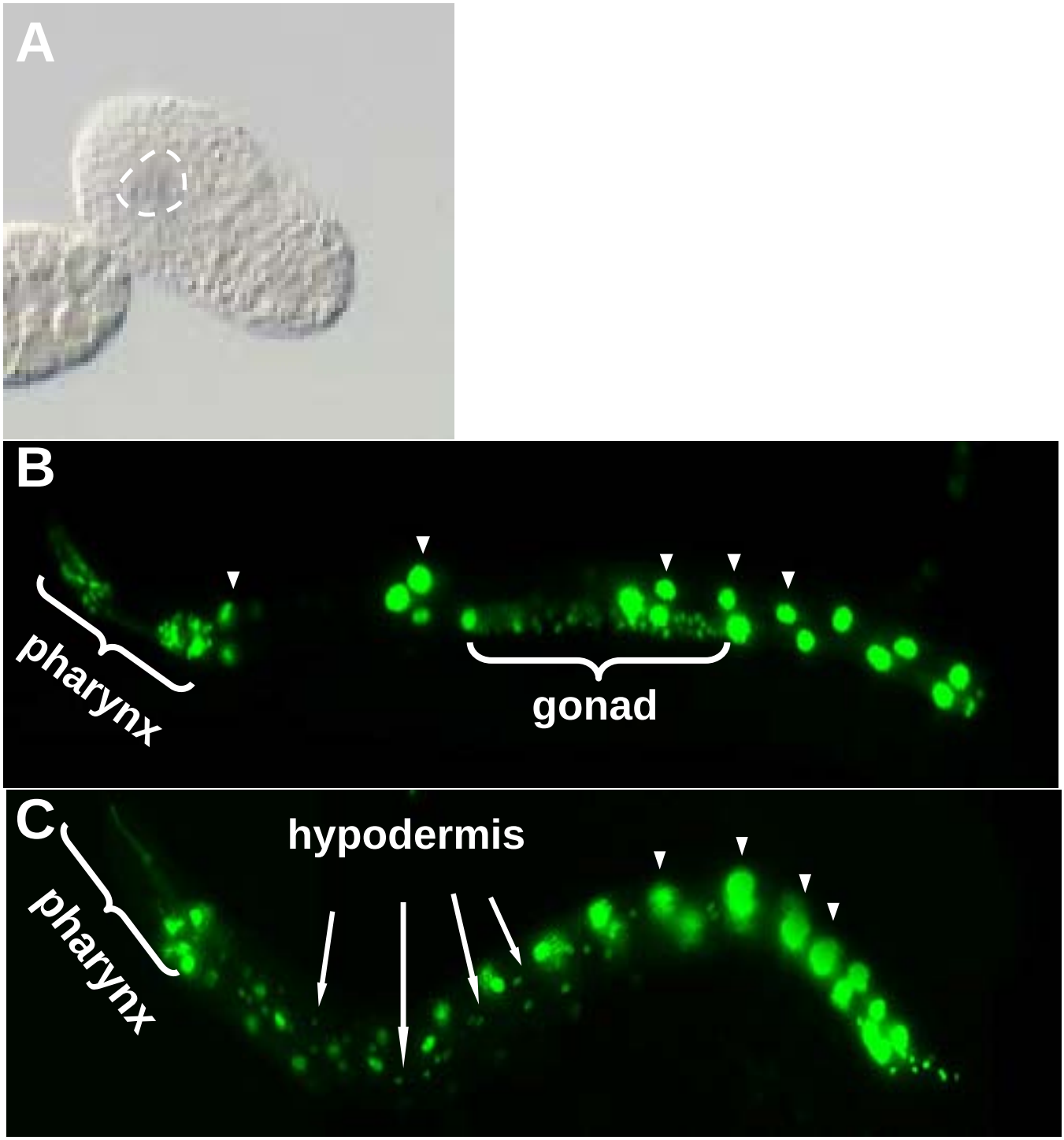


Figure 5. The expression of *M05B5.2*. A) The *in situ* hybridization from the Nematode Expression Database [http://nematode.lab.nig.ac.jp] showing *M05B5.2* embryonic expression in the pharynx, highlighted by the white dashed line. B) Expression of the *M05B5.2*::*GFP*::*HIS2B* reporter in the pharynx, intestine (arrowheads), hindgut and somatic gonad. C) Expression of mutM05::*GFP*::*HIS2B* in the pharynx, intestine (arrowheads), hindgut, somatic gonad and hypodermis (arrows).

Materials and Methods

General *C. elegans* maintenance and nomenclature

The *C. elegans* Bristol strain, referred to as N2, was used as the wildtype strain for analysis. Animals were maintained as described by Brenner (1974) on nematode growth medium plates (NGM) that were seeded with the OP50 strain of *E. coli*. These plates were then kept at either 15ºC or 20ºC during experimental manipulation.

The nomenclature for *C. elegans* follows the guidelines outlined by Horvitz and colleagues (1979). Genes are assigned a three or four letter gene name that is related to some aspect of the mutant phenotype or to the predicted product. As well as the name, these genes are also assigned a number based on how many previous genes with that name exist. The names of genes are written as lower case italicized letters. For example, the gene *pha-4* is the fourth gene with a defective PHArynx development phenotype. In addition, when referring to a protein the entire gene name is capitalized, PHA-4 and when referring to a phenotype the first letter of the gene is capitalized, Pha. The source laboratory for a given strain is indicated by a two or three letter designation followed by a number; the Gaudet lab strain designation is "GD". Similarly, alleles are designated by a separate one to three letter label followed by a number; the Gaudet lab allele designation is "*iv*". More information about the genes of *C. elegans* and their names can be found online at www.wormbase.org. The strains of nematodes used during this study, including their strain name and genotype, are listed in Table 1.

Plasmid Construction

Deletion analysis of M05B5.2

Plasmids were constructed using standard molecular biology techniques (Sambrook and Russel, 2001). All deletion constructs were cloned using the parent plasmids pSEM544, a plasmid containing 1kb of *M05B5.2* promoter cloned in front of a translational fusion of green fluorescent protein (GFP) and histone 2B (HIS2B; from the *C. elegans his-48* gene) and pSEM461, a derivative of pSEM544 that carries a disrupted/mutated PHA-4 binding site in the *M05B5.2* promoter (Gaudet and Mango 2002). Deletions of the *M05B5.2* promoter were generated by PCR amplification and subsequent cloning into the parent plasmids, using restriction sites present in the PCR product or built in to the oligonucleotide primers, as indicated in Table 2.

Constructs created for mutation analysis

The mutation constructs were created by PCR amplification of mutant sequence from the parent plasmids, pSEM544 and pSEM461, using oligonucleotide primers that alter the original sequence and introduce a new restriction endonuclease site (Ho et al., 1989). These sites are then used to ligate the fragments back into the parent plasmid and to verify the identity of the clones.

Mutation of the GATA site

The conserved GATA site was mutated to an *Fsp* I restriction site by PCR fragment amplification using two oligonucleotides, oGD5-GGAGACAAGTTTGGTGC GCAGATGGAAAAATGCG and oGD6-CGCATTTTTCCATCTGCGCACCAAACTT

GTCTCC. These oligonucleotides along with oGD3 and oGD4 amplified two PCR fragments that were then ligated back into the parent plasmids at the *Kpn* I and *Pst* I restriction sites.

Mutation of the PIG3 site

Plasmid mutPIG3, in which the conserved PIG3 site was mutated to an *Nsi* I restriction site, was generated by PCR amplification using two oligonucleotides, oGD79-TTTAATGCATGTTATGTGCGTATGAGTCAGTGAAG and oGD80-TAAATGCATG TTTTGGTGCGCAGATGG. These oligonucleotides along with oGD3 and oGD4 amplified two PCR fragments that were then ligated back into the parent plasmids at the *Kpn* I and *Pst* I restriction sites.

Mutation of the PIG2 site

Plasmid mutPIG2, in which the conserved PIG2 site was mutated to a *Spe* I restriction site, was generated by PCR amplification using two oligonucleotides oGD115-TCTTCACTGACTAGTACGCACATAGGAGA and oGD116- TCTCCTATGTGCGTA CTAGTCAGTGAAGA. These oligonucleotides along with oGD3 and oGD4 amplified two PCR fragments that were then ligated back into the parent plasmids at the *Kpn* I and *Pst* I restriction sites.

Specific mutation of the PHA-4 site of PIG1

To mutate the PHA-4 site in the PIG1 element while leaving the ONECUT site intact (PIG1-PHA-4(-)), I used site directed PCR mutagenesis using two oligonucleotides, oGD270-AAAATGTAACCCC*ACTAGT*TTCTCTATGTTTGCAGCA

CACTGACTCATAC and oGD271- GTATGAGTCAGTGTGCTGCAAACATAGAGA A*ACTAGT*GGGGTTACATTTT that mutate the PHA-4 site to the underlined sequence and add the italicized *Spe* I restriction site for cloning. These oligonucleotides along with oGD3 and oGD4 amplified two PCR fragments that were then ligated back into the parent plasmids at the *Kpn* I and *Pst* I restriction sites. These constructs were sequenced to check for the presence of the mutated PIG1 site.

Specific mutation of the ONECUT site of PIG1

To mutate the ONECUT site in the PIG1 element while leaving the PHA-4 site intact (PIG1-ONECUT(-)), I used site directed PCR mutagenesis using two oligonucleotides, oGD297-AAAATGTAACCCC*ACTAGT*TTCTCTAAGTTGACTCTTC ACTGACTCATAC and oGD298-GTATGAGTCAGTGAAGAGTCAACTTAGAGAA*A CTAGT*GGGGTTACATTTT that mutate the ONECUT site to the underlined sequence and add the italicized *Spe* I restriction site for cloning. These oligonucleotides along with oGD3 and oGD4 amplified two PCR fragments that were then ligated back into the parent plasmids at the *Kpn* I and *Pst* I restriction sites. These constructs were sequenced to check for the presence of the mutated PIG1 site.

C. briggsae reporter

I created a reporter (Cb-M05) containing the promoter of the homologous *C. briggsae* gene to *M05B5.2*, CBG21925, using a method similar to that of Gaudet and Mango (2002) in their production of the *C. elegans* M05B5.2 reporter. I PCR amplified 1000 bp of sequence upstream of the predicted start site in CBG21925 from *C. briggsae* (AF16) genomic DNA. The

first oligonucleotide primer oGD75- CGCGTCGACTTCCGAAACATTCTTCCAAA introduced a *Sal* I site and the second primer oGD74- CGGGGTACCTTTGAAACGGAAAGGGTAAATC introduced a *Kpn* I site at the ends of the amplified genomic DNA fragment. This promoter fragment was then cloned into the pAP.10 plasmid, which contains the *GFP::HIS2B* translational fusion described above, at the *Kpn* I and *Sal* I sites. Finally, a restriction digest was performed to confirm that the cloning was successful. For this study the *C. elegans* M05B5.2 reporter will be described as M05B5.2, whereas the C. briggsae M05B5.2 reporter will be described as Cb-M05.

Enhancer constructs

Some of the enhancer constructs were created using oligonucleotides containing three tandem copies of the site under investigation. Complementary oligonucleotides were annealed to create double stranded fragments with short overhangs to facilitate cloning. These double stranded fragments were then cloned into the *Sph* I and *Sal* I restriction sites of a “promoter-less” GFP reporter plasmid, pPD95.77 (kindly provided by Andrew Fire) to create the 3xPIG1, 3xPIG2, 3xPIG3, 3xPHA-4 and 3xONECUT constructs. The combinatorial 3xPIG2::3xPIG3 enhancer was created by ligating together restriction fragments from the 3xPIG2 and 3xPIG3 constructs. The oligonucleotides used to create these enhancer constructs are listed in Table 3.

The 3xRegion I and 3xPIG4 enhancers were created by an enzyme based multimerization protocol adapted from Robinett et al. (1996) (Figure 6). The primers were designed with *Hind* III and *Nhe* I restriction sites on one end and *Bam* HI and *Avr* II restriction sites on the other. These primers were annealed to generate a double stranded fragment with overhangs for cloning; these fragments were ligated to the pPD95.77 “promoter-less” GFP reporter . Next, two independent digests of this single copy enhancer plasmid were performed, the first with *Eco* RI and *Nhe* I and

the second with *Eco* RI and *Avr* II. *Avr* II and *Nhe* I overhangs are compatible, but their ligation does not recreate either restriction site. A ligation at the *Eco* RI and *Avr* II*/Nhe* I sites was done to recapitulate the full-length reporter that now contains two copies of the enhancer fragment. This was repeated to get three copies of the enhancer in front of GFP. The primers that were used to create these enhancers are listed in Table 3.

Repressor assays with dpy-7 constructs

Oliver Hobert provided the minimal *dpy-7* reporter plasmid, which was originally cloned by John Gilleard as pdp7.12 (Gilleard et al., 1997). This reporter contains the minimal *dpy-7* promoter in front of GFP that is expressed in the hypodermis of *C. elegans* from comma stage embryos through adulthood. The 3xPIG1::*dpy-7*::*GFP*, 3xPHA-4::*dpy-7*::*GFP*, 3xONECUT::*dpy-7*::*GFP* and 3xPHA-4::3xONECUT::*dpy-7*::*GFP* constructs were created by restriction digest and ligation of restriction fragments. The *dpy-7*::*GFP* construct was cut with *Hind* III, blunt-ended by treatment with Klenow and then cut with *Eco* RI. Next, the 3xPIG1, 3xPHA-4, 3xONECUT and 3xPHA-4::3xONECUT constructs were cut with *Bam* HI, overhang blunted and then cut with *Eco* RI. Fragments of these two restriction digests were ligated at the blunt ends and the *Eco* RI site to ensure proper orientation of the fragments after ligation. The correct structure of these constructs was confirmed by restriction digest.

Ectopic hypodermal expression construct

A construct was created to determine the effects of expression of *M05B5.2* in the hypodermis by cloning the *M05B5.2* gene downstream of the mutM05 promoter, which is active in the hypodermis. The construct includes the genomic sequence of *M05B5.2* from the predicted

start codon to 500 bp 3’ of the predicted stop codon. The genomic DNA was amplified using the two oligonucleotide primers oGD516- GGGGTACCATG GAAATTCGACTAAGGCAC, which introduced a *Kpn* I restriction site and oGD517- T TTCCCGGGGATCATGGAATTTGGGTCG, which introduced an *Apa* I site. Next, this fragment was cloned into the mutM05 plasmid at the *Kpn* I and *Apa* I restriction sites to replace the coding region of the fluorescent protein.

Microinjections

C. elegans transformations were performed by microinjection following standard procedures (Jin, 1999; Mello et al., 1991). The injection mixes for the deletion and mutational analysis were made up of the reporter construct of interest at 10 ng/µL, 10 ng/µL of the co-injection marker *B0507.1::YFP* (pGD28), which is expressed in the pharyngeal glands, and 80 ng/µL of pBS II(SK+) plasmid as carrier DNA. Enhancer constructs were injected at 50 ng/µL along with 50 ng/µL of pRF4, which carries the dominant Roller mutation *rol-6(su1006)* (Mello et al., 1991). Enhancers containing PHA-4 binding sites are toxic at high concentrations (Raharjo and Gaudet, 2007); in these cases, constructs were injected at 20 ng/µL with 50 ng/µL of pRF4 injection marker and 30 ng/µL of pBS II(SK+).

For each construct at least three independent transgenic lines were examined. These lines were scored as embryos, larvae and adults to look for expression of the reporter or enhancer constructs. Transgenics were established in an N2 background except where noted.

RNAi injection

Double stranded RNA (dsRNA) injections were performed using a modified version of the protocol described by Fire et al. (1998). The RNAi clones were obtained from the Ahringer/Geneservice Ltd. Library (Fraser et al., 2000; Kamath et al., 2003) and *in vitro* transcribed using Promega's T7 polymerase. The dsRNA was injected at a concentration of 1 μg/μL into both syncytial arms of the gonad of the reporter strain of interest. When two dsRNA constructs were injected simultaneously each double stranded RNA construct was injected at 0.5 μg/μL.

Injected hermaphrodites were placed onto a recovery plate for 24 hours at 20ºC to allow time for the RNAi to take effect. Injected animals were then transferred individually to plates and transferred every 24 hours thereafter to new plates for the next 3 days. The progeny of these injected worms were observed as embryos, larvae and adults to look for changes in reporter expression, morphology or survival.

C. elegans **lysis for PCR**

The protocol for *C. elegans* lysis and PCR was similar to the protocol used by Williams et al. (1992) with modifications. Single animals were placed in lysis buffer in a PCR tube. Next, these samples were frozen at –80ºC for 30 minutes followed by incubation at 60ºC for 2 hours and then at 95ºC for 15 minutes. The PCR mixture was added to the single animal lysates and the desired DNA fragment was amplified through PCR.

***let-527* rescue**

The *M05B5.2* rescuing fragment was amplified from N2 genomic DNA with oligonucleotides oGD252-TCTAGATTGGCAAACAATCTGAAA and oGD253- GAGC TCGAAAACTCCCGTAACAAC. The amplification product includes 1000 bp 5' of the predicted start codon and 500 bp 3' of the predicted stop codon. This fragment was injected directly into N2 and *let-527* worms at a variety of concentrations from 5-50 ng/µL. The injection mixtures also contained pBS II (SK+) to a final concentration of 100 ng/µL and two injection markers: 10 ng/µL of the pharyngeal gland-specific *B0507.1*::*YFP* reporter and 30 ng/µL of *elt-2*::*GFP*::*lacZ*, an intestine-specific reporter (Fukushige et al., 1999).

Table 1. Worm strains used in this study.

Strain	Genotype
N2	Wildtype *C. elegans*
AF16	Wildtype *C. briggsae*
GD73	*rrf-3(pk1426)* II; *mIs11* IV [*pes-10::GFP myo-2::GFP*]
SM469	*pxIs6* IV [*pha-4::GFP::HIS2B rol-6(su1006)*]
KR523	*let-527(h209) dpy-5(e61) unc-13(e450)* I; *sDp2(I;f)*
DR63	*daf-4(m63)* III
FX*	*atf-2(tm467)* II
DR26	*daf-16(m26)* I
PY1133	*unc-130(oy10)* II
FX*	*fkh-5(tm317)* III
FX*	*ceh-38(tm321)* II
VC626	*ceh-39(gk296)* X
VC766	*ceh-39(gk329)* X
FX*	*ceh-49(tm1149)* V

*Strains created by the Mitani lab with allele designations *tm* do not have identifying FX strain numbers.

Table 2. Description of the deletion plasmids and their construction.

Description of the plasmid	Insert	Vector	Oligonucleotide primers*
5`Δ400M05 (pGD32)	400bp *Pst* I-*Kpn* I PCR-amplified promoter fragment	pSEM544	oGD39-AACTGCAGAGGTGAGCACATACA + oGD4-GGACAACTCCAGTGAAAAG
5` Δ400mutM05 (pGD33)	400bp *Pst* I-*Kpn* I PCR-amplified promoter fragment	pSEM461	oGD39 + oGD4
5`Δ270M05 (pGD35)	270bp *Pst* I-*Kpn* I PCR-amplified promoter fragment	pSEM544	oGD40- AACTGCAGGTAACCCCGTTATGT + oGD4
5`Δ270mutM05 (pGD34)	270bp *Pst* I-*Kpn* I PCR-amplified promoter fragment	pSEM461	oGD40 + oGD4
5`Δ230M05 (pGD36)	230bp *Pst* I-*Kpn* I PCR-amplified promoter fragment	pSEM544	oGD41-AACTGCAGCATACGCACATAGGA + oGD4
5`Δ200M05 (pGD37)	200bp *Pst* I-*Kpn* I PCR-amplified promoter fragment	pSEM544	oGD42-AACTGCAGTGGTGATAAGATGGA + oGD4
Δ50bpM05 (pGD53)	GATA site mutated to *Fsp* I restriction site and then *Fsp* I-*Sac* II restriction digest, blunt-ended and ligated	pSEM461	GATA site mutation with: oGD5-GGAGACAAGTTTGGTGCGCAGATGGAAAAATGCG + oGD6-CGCATTTTTCCATCTGCGCACCAAACTTGTCTCC
3`Δ200M05 (pGD208)	800bp *Pst* I-*Kpn* I PCR-amplified promoter fragment	pSEM544	oGD263-GGTACCTTATCACCAAACTTGTCTCC + oGD3-GCTGAAATCACTCACAACGATGG
3`Δ150M05 (pGD209)	Ligation of two PCR-amplified promoter fragments: 50bp *Nsi* I-*Kpn* I and 800bp *Nsi* I-*Pst* I	pSEM544	oGD264-ATGCATTCACTTTTTTGTGTTTTTTAT + oGD4 oGD267-TGCATTTATCACCAAA + oGD3
3`Δ100M05 (pGD210)	Ligation of two PCR-amplified promoter fragments: 100bp *Nsi* I-*Kpn* I and 800bp *Nsi* I-*Pst* I	pSEM544	oGD265-ATGCATCAGTTTAAGATTTCATTTACTTC + oGD4 oGD267 + oGD3
3`Δ100mutM05 (pGD236)	Ligation of two PCR-amplified promoter fragments: 100bp *Nsi* I-*Kpn* I and 800bp *Nsi* I-*Pst* I	pSEM461	oGD265 + oGD4 oGD267 + oGD3
3`Δ50M05 (pGD211)	Ligation of two PCR-amplified promoter fragments: 150bp *Nsi* I-*Kpn* I and 800bp *Nsi* I-*Pst* I	pSEM544	oGD266-ATGCATGTTTTCTTGACATTTCTATTTT + oGD4 oGD267 + oGD3

3`Δ50mutM05 (pGD276)	Ligation of two PCR-amplified promoter fragments: 150bp *Nsi* I-*Kpn* I and 800bp *Nsi* I-*Pst* I	pSEM461	oGD266 + oGD4 oGD267 + oGD3
3`Δ25M05 (pGD277)	Ligation of two PCR-amplified promoter fragments: 175bp *Nsi* I-*Kpn* I and 800bp *Nsi* I-*Pst* I	pSEM544	oGD404-GACAGCAACTTGTCTCCTATGTGC + oGD4 oGD267 + oGD3
3`midΔ100M05 (pGD278)	Ligation of two PCR-amplified promoter fragments: 50bp *Nsi* I-*Kpn* I and 850bp *Nsi* I-*Pst* I	pSEM544	oGD406-ATGCATCACAAAAAAGTGAAGAAAATC + oGD4 oGD407-TATTTTATTTTTTAAGTTTTCTTGACAT + oGD3
3`midΔ100mutM05 (pGD279)	Ligation of two PCR-amplified promoter fragments: 50bp *Nsi* I-*Kpn* I and 850bp *Nsi* I-*Pst* I	pSEM461	oGD406 + oGD4 oGD407 + oGD3

* The sequences of the oligonucleotide primers are all listed in the 5` to 3` direction

Table 3. Primers used to create the enhancer constructs.

Oligonucleotide name	Element in multiple copies in the enhancer constructs	Oligonucleotide sequence*
OGD250	PIG1	CTCTATGTTGACTCTTCCTCTATGTTGACTCTTCCTCTATGTTGACTCTTC
OGD251	PIG1	TCGACAAAAGAGTCAACATAGAGGAAGAGTCAACATAGAGGAAGAGTCAACATATAGCATG
oGD125	PIG2	CTAACTGACTCATACGCACTGACTCATACGCACTGACTCATATTG
oGD126	PIG2	TCGACAATATGAGTCAGTGCGTATGAGTCAGTGCGTATGAGTCAGTTAGCATG
oGD94	PIG3	CTAGGAGACAAGTTTAGGAGACAAGTTTAGGAGACAAGTTG
oGD95	PIG3	TCGACAACTTGTCTCCTAAACTTGTCTCCTAAACTTGTCTCCTAGCATG
oGD395	PHA-4	CTTATGTTTGCAGCCTATGTTTGCAGCCTATGTTTGCAGTTG
oGD396	PHA-4	TCGAGGCTGCAAACATAGGCTGCAAACATAGGCTGCAAACATAGGCATG
oGD391	ONECUT	CTCTAAGTTGACTCTTCCTCTAAGTTGACTCTTCCTCTAAGTTGACTCTTTTG
oGD392	ONECUT	TCGAGAAGAGTCAACTTAGAGGAAGAGTCAACTTAGAGGAAGAGTCAACTTAGAGCATG
oGD461	PIG4	AGCTTGCTAGCGCTGTCTTCATTTTTTCCCCACTATTTTCGATTTTCTTCACTTTTCCTAGGCTGCA
oGD462	PIG4	GCCTAGGAAAAGTGAAGAAAATCGAAAATAGTGGGGAAAAAATGAAGACAGCGCTAGCA
oGD437	Region I	AGCTTGCTAGCTCTATGTTGACTCTTCACTGACTCATACGCACATAGGAGACAAGTTTCCTAGGCTGCA
oGD438	Region I	GCCTAGGAAACTTGTCTCCTATGTGCGTATGAGTCAGTGAAGAGTCAACATAGAGCTAGCA

*The element in multiple copy in each enhancer is underlined.

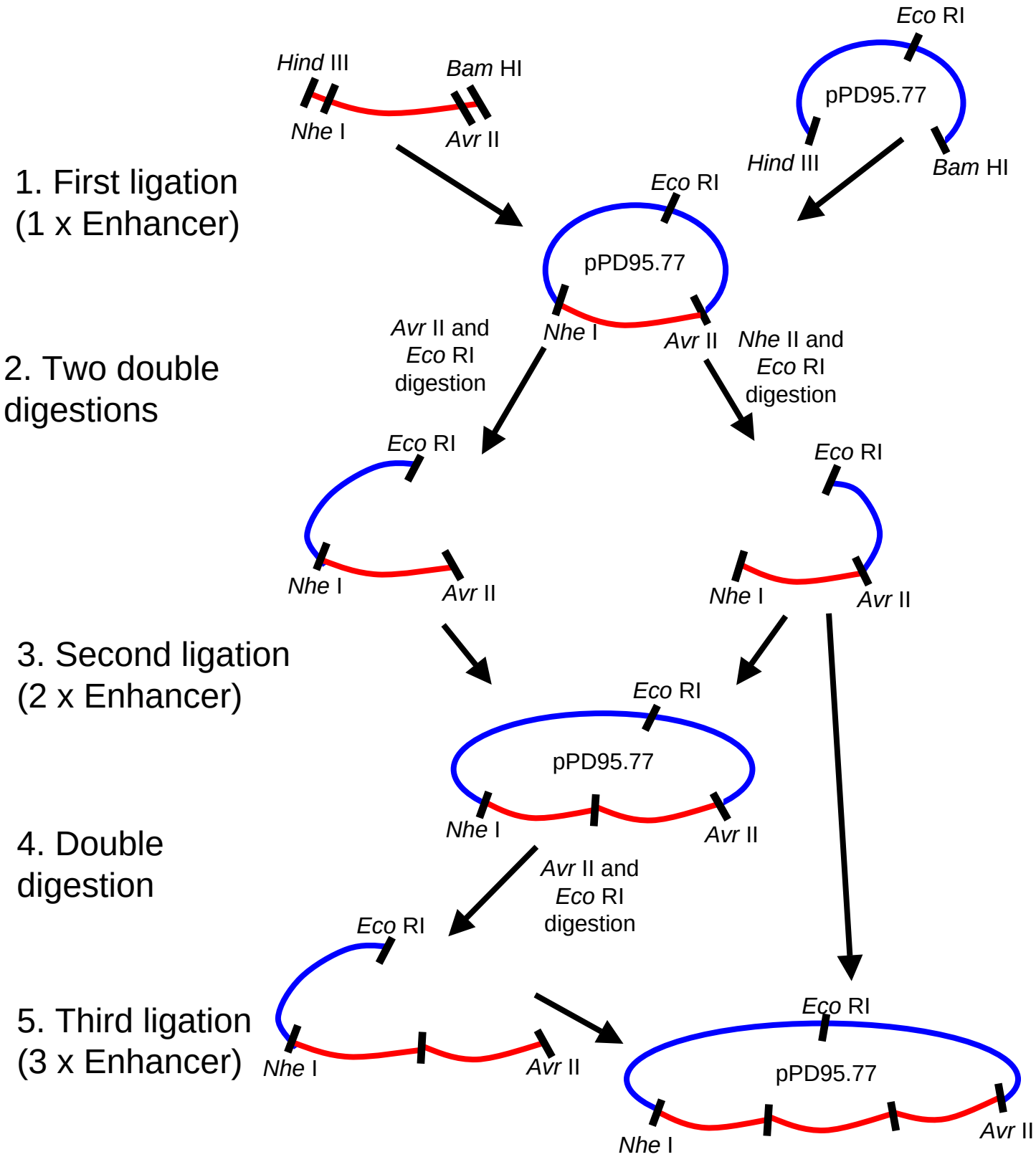


Figure 6. Multimerization protocol adapted from Robinett et al. [1996].

I performed all of the following experiments except for the initial 5` deletion analysis that was undertaken by Sarah Fung. The constructs created and tested by Sarah Fung were the 5`Δ400M05, 5`Δ400mutM05, 5`Δ270M05, 5`Δ270mutM05 and 5`Δ200M05 reporters.

Results

M05B5.2 was previously identified as a pharyngeally-expressed gene (Gaudet and Mango, 2002). It is one of only two characterized pharyngeal genes whose expression is not wholly dependent on the presence of PHA-4 binding sites. The removal of the PHA-4 binding sites from the promoters of both *M05B5.2* and *myo-2* weakens but does not abolish pharyngeal expression (Gaudet and Mango, 2002). Analysis of the promoter of *myo-2* has led to the identification of two transcription factors, PEB-1 and CEH-22. Investigation of *M05B5.2* should similarly identify other transcription factors involved in pharyngeal specification, development and/or gene expression. Also, the mutation of one of the PHA-4 sites in the promoter of *M05B5.2* leads to ectopic expression in the hypodermis of *C. elegans*, suggesting that analysis of the promoter of this gene should identify both positive and negative regulators of expression.

RNAi to M05B5.2

M05B5.2 is expressed in the pharynx, based on *in situ* hybridization data, reporter expression, and a microarray experiment (http://nematode.lab.nig.ac.jp; Gaudet and Mango, 2002), suggesting that the loss of this gene may lead to a pharyngeal phenotype. To test this, I injected M05B5.2 dsRNA into three different *C. elegans* strains: GD73 (*rrf-3(pk1426)* II; *mIs11* IV [*myo-2::GFP pes-10::GFP*]), SM469 (*pxIs6* IV [*pha-*

4::*GFP*::*HIS2B rol-6(su1006)*]) and an N2 strain containing an extrachromosomal *M05B5.2*::*GFP*::*HIS2B* reporter. The *rrf-3(pk1426)* mutation in GD73 leads to an increased sensitivity to RNAi (Simmer et al., 2002) and the integrated *myo-2*::*GFP* reporter marks the pharyngeal muscles, which makes any pharyngeal phenotype easier to identify. The SM469 strain has an integrated *pha-4*::*GFP* translational fusion that is expressed in all pharyngeal nuclei (as well as other *pha-4* expressing cells), facilitating characterization of any pharyngeal defects. Finally, the *M05B5.2*::*GFP*::*HIS2B* strain was used to look at the effects of RNAi on the tissues in which *M05B5.2* is expressed, as this reporter is expressed in the intestine, hindgut and somatic gonad as well as the pharynx. *fkh-3 ds*RNA was injected as the control as it has been shown to have no effect on *C. elegans* (Hope et al., 2003).

The injection of M05B5.2 dsRNA into GD73 and SM469 caused pharyngeal defects and death in *C. elegans.* I observed 85% lethality in the progeny of *rrf-3; Is[myo-2::GFP]* animals injected with M05B5.2 dsRNA and 89% lethality in similar animals from the *Is[pha-4::GFP]* strain, confirming previous data from large-scale RNAi screens, which indicated embryonic and/or larval lethality for *M05B5.2(RNAi)* (Fraser et al., 2000; Sonnichen et al., 2005). Most of the *M05B5.2(RNAi)* animals arrested as L1 larvae and later died, with a small percentage dying as embryos (Table 4).

Further analysis of the *M05B5.2(RNAi)* phenotype argues that this gene is functioning in the pharynx. The pharynx of the GD73 and SM469 animals were misshapen with the anterior bulb showing the most penetrant defects (Figure 7). In addition, *M05B5.2(RNAi)* animals frequently (68% of progeny) appeared to have pharyngeal cells lying outside of the pharyngeal basement membrane. In my RNAi experiments, I observed both *myo-2*::*GFP* and *pha-4*::*GFP* expressing cells that were not

part of the pharynx and were not present in uninjected or control animals (Figure 7). The cells that express these two reporters outside of the pharyngeal membrane may have separated from the pharynx during development or may be cells that inappropriately express these markers.

The pharyngeal defects caused by injection of M05B5.2 dsRNA suggest that the RNAi lethality may be due to a feeding defect that leads to the starvation of these animals. Interestingly, some of the *M05B5.2(RNAi)* injected animals die as embryos, similar to the effect of the loss of PHA-4 in *C. elegans*, which also causes embryonic lethality (Mango, Lambie and Kimble, 1994). However, the loss of *M05B5.2* by dsRNA injection still produces pharyngeal tissue unlike the loss of PHA-4. Thus, it is possible that the loss of *M05B5.2* in *C. elegans* leads to pharyngeal defects that cause embryonic death as well as larval arrest.

The lethality due to injection of M05B5.2 dsRNA may also be caused by defects in the other tissues in which this gene is expressed, including the intestine, hindgut and somatic gonad. Therefore, I injected M05B5.2 dsRNA into an N2 transgenic strain carrying the *M05B5.2*::*GFP*::*HIS2B* reporter, which is expressed in the pharynx, intestine, hindgut and somatic gonad. I looked for defects in these other tissues, though the early lethality of the RNAi prevented scoring of gonadal expression. This RNAi injection did not cause any visible morphological defects in the intestine or hindgut. Similar to the GD73 and SM469 strains, the only defect in these animals were ectopic GFP expressing cells in the head. It is likely that these ectopic GFP expressing cells are the same cells that ectopically expressed the *myo-2*::GFP and *pha-4*::GFP markers. The lethality associated with *M05B5.2(RNAi)* is probably due to pharyngeal defects that result

in starvation. The reporter expression and RNAi data suggest that *M05B5.2* is critical for normal pharynx development and viability.

Analysis of the coding region of *M05B5.2*

The loss of M05B5.2 by dsRNA injection suggests that *M05B5.2* has a role in pharyngeal development and/or function. To try to understand the role that this gene plays in the pharynx, I looked at the predicted protein that it encodes. *M05B5.2* is predicted to encode a novel protein that analysis suggests may be a single-pass transmembrane protein with no known homologs outside of *Caenorhabditis*. BLAST searches with the predicted protein sequence of M05B5.2 did not identify any related proteins and searches using PFAM did not reveal any recognizable protein motifs. Similarly, BLAST and PSI-BLAST searches performed with the predicted extracellular, intracellular and transmembrane domains did not identify any related proteins. To determine where the predicted product of M05B5.2 may be localized in the cell, I used a transmembrane prediction program, TMHMM2 (Krogh et al., 2001; http://www.cbs.dtu.dk/services/TMHMM/) and a signal peptide prediction program, SignalIP 3.0 (Dyrlov et al., 2004; http://www.bioinformatics.leeds.ac.uk/cgi-bin/sig_pred/signal.cgi). Results from these searches reveal that the product of M05B5.2 contains a probable signal peptide, signal peptide cleavage site and transmembrane domain (Figure 8) suggesting that M05B5.2 may be a transmembrane protein with a function at the cell surface.

let-527 may be a mutation in *M05B5.2*

Given the lethal phenotype of *M05B5.2(RNAi)*, I searched for possible lethal mutants that might correspond to mutations affecting *M05B5.2*. Interestingly, the lethal mutation *let-527(h209)* maps to a small region of the genome that contains five genes, including *M05B5.2* (Figure 9). *let-527* was found in a screen for lethal mutants on chromosome I in the region between *dpy-14* and *unc-29* (McKim, Starr and Rose, 1992). *let-527* was mapped by duplication and deficiency mapping to chromosome I at 1.82 +/- 0.024 cM (McDowall and Rose, 1997). Of the five genes in the *let-527* region, only *M05B5.2* is reported to have an RNAi phenotype in large-scale RNAi screens (Fraser et al., 2000; Rual et al., 2004; Simmer et al., 2003; Sonnischen et al., 2005). Given that *M05B5.2(RNAi)* results in lethality, *let-527* may correspond to a mutation in *M05B5.2*. As well, flanking genes on either side of the *let-527* region do not have reported RNAi phenotypes. To determine whether *let-527* is a mutation in *M05B5.2*, I sequenced *M05B5.2* from the *let-527* strain and tested whether M05B5.2(+) could rescue the *let-527* phenotype.

The *let-527* mutation is homozygous lethal and is maintained as a homozygous strain rescued by a free duplication (strain KR523: *let-527(h209) dpy-5(e61) unc-13(e450)* I *; sDp2(I;f)*). The duplication *sDp2* also rescues *dpy-5* so that viable animals are Unc (uncoordinated) while dead animals are Dpy (dumpy) and Unc.

Initial observation of *let-527* animals showed that they die at the same stage as *M05B5.2(RNAi)* animals, but *let-527* animals do not have pharyngeal defects. In the *let-527* strain 81.2% (65/80) of animals fail to make it to adulthood as they arrest as L1 larvae and die 3-4 days later, which is similar to the lethality in the GD73 and SM469 strains (Table 4). In contrast, the pharynx of *let-527* animals did not show any defects,

even when marked with the *myo-2::GFP* reporter. There were no cells expressing *myo-2::GFP* outside of the pharynx, in contrast to the results of *M05B5.2(RNAi)*. Thus, *let-527* and *M05B5.2(RNAi)* produce similar but non-identical phenotypes.

To explicitly test whether *let-527* was a mutation in *M05B5.2*, I attempted to rescue the *let-527* lethality with a genomic fragment carrying wild-type *M05B5.2* sequence. This fragment contained the entire predicted *M05B5.2* gene with 1 kb of sequence upstream of the predicted start codon and 500 bp of sequence following the predicted stop codon. I first injected this rescuing construct into N2 animals with two GFP markers to establish transgenic lines. Next, I crossed M05B5.2(+) carrying animals to the *let-527* strain to get +/*let-527dpy-5 unc-13; Ex[M05B5.2(+)]* animals. From these heterozygous transgenics, I isolated surviving Dpy Unc animals (ie, that *let-527* was indeed present in the strain) where M05B5.2(+) rescued *let-527*. To verify the genotype of the survivors as *let-527 dpy-5 unc-13; Ex[M05B5.2]*, I outcrossed this strain with N2 males to obtain heterozygous progeny (*+/let-527 dpy-5 unc-13; Ex[M05B5.2(+)]* animals). The hermaphrodite progeny were allowed to self-fertilize and the resultant progeny were scored. The wildtype progeny consisted entirely of non-Dpy non-Unc animals with and without the transgene and Dpy Unc animals only with the transgene. All of the Dpy Unc animals that survived to adulthood contained the transgene, suggesting that M05B5.2(+) is required for the survival of *let-527* animals.

I therefore sequenced *M05B5.2* from *let-527* animals to search for possible molecular lesions. I picked presumptive *let-527 dpy-5 unc-13* animals that arrested as larvae and so should not carry the *let-527(+)* on the *sDp2* duplication. However, I could not be certain that the duplication was lost in these animals and it is possible that the rescued copy of *M05B5.2* from the duplication may have been sequenced. Therefore, I

also sequenced *M05B5.2* from *let-527 dpy-5 unc-13; Ex[M05B5.2(+)]* animals using primers that solely amplify the chromosomal copy of M05B5.2 to ensure that I sequenced the genomic copy of *M05B5.2* and not the rescuing fragment. In addition, I sequenced *M05B5.2* in animals that were outcrossed with N2 to remove the *sDp2* duplication. I crossed *let-527* animals with N2 males to get *+/let-527 dpy-5 unc-13* animals and then let these hermaphrodites self-fertilize. Next, I picked F2 animals that arrested as L1 that were presumed to be *let-527 dpy-5 unc-13/ let-527 dpy-5 unc-13*. *M05B5.2* sequence was examined from 250 bp upstream of the start codon through to the stop codon in each of these strains, but in no case did I find a molecular lesion. Thus, I could not confirm the rescuing data with a mutation in *M05B5.2*.

RNAi analysis determined that *M05B5.2* is an essential gene that functions in the pharynx of *C. elegans.* The pharyngeal phenotype corroborates the *in situ* hybridization, reporter and microarray data. Therefore, further examination of the promoter of this gene should identify critical pharyngeal regulatory elements and *trans*-acting factors.

Promoter analysis

Previous work with *M05B5.2* has demonstrated that it is expressed in the same cells as PHA-4: those of the pharynx, intestine, hindgut and somatic gonad (Gaudet and Mango, 2002). This result suggested a simple model in which PHA-4 is the primary regulator of *M05B5.2* expression. However, removal of predicted PHA-4 sites from the *M05B5.2* regulatory region, although it decreased the level of expression, did not eliminate expression in the pharynx and intestine, suggesting that other regulators are also acting on this promoter. Furthermore, mutation of a single PHA-4 binding site in the

promoter of *M05B5.2* resulted in ectopic expression in the hypodermis. These results suggest that there are both positive and negative regulators, in addition to PHA-4, that control *M05B5.2* expression. Further analysis of the *M05B5.2* promoter was therefore undertaken to identify the relevant *cis*-acting regulatory elements.

Expression of the homologous CBG21925 gene, in C. briggsae

One way to identify critical promoter elements is to examine conservation of sequences between related species. I therefore compared the promoter of *M05B5.2* to the promoter of its homologous gene in *Caenorhabditis briggsae,* CBG21925. To ensure that this comparison would identify likely regulators of gene expression I compared the expression pattern of the *M05B5.2*::*GFP*::*HIS2B* reporter with the expression pattern of a CBG21925 reporter (Cb-M05). The Cb-M05 reporter contained 1 kb of upstream promoter sequence from CBG21925 cloned in front of a translational fusion between green fluorescent protein (GFP) and histone-2B (HIS2B). This was injected into the wildtype *C. briggsae* strain, AF16. However, I was unable to obtain lines of AF16 transgenic animals so the expression of the F1 progeny of injected adults was directly examined. I found that the Cb-M05 reporter had a very similar expression pattern to that of the *M05B5.2*::*GFP*::*HIS2B* reporter in *C. elegans*, with expression throughout the digestive tract in the pharynx, intestine and hindgut, but it lacked somatic gonad expression (Figure 10). The expression of the homologous *C. briggsae* gene in the digestive tract indicates that comparison between these two promoters could identify critical regulatory elements.

Deletion analysis

Two constructs were used to identify the necessary regulatory elements in the promoter of *M05B5.2*. The first construct, *M05B5.2*::*GFP*::*HIS2B*, contains 1 kb of promoter sequence upstream of the predicted *M05B5.2* ATG cloned in front of a translational fusion between GFP and HIS2B. This construct is expressed in the pharynx, gut, hindgut and somatic gonad (Figures 5 and 11). The second construct, mutM05, is the same as *M05B5.2*::*GFP*::*HIS2B* but with a mutation in one of the predicted PHA-4 sites that results in ectopic expression of the reporter in the hypodermis (Figures 5 and 11). Studying the promoter using these two constructs has enabled me to identify the *cis*-regulatory sites of *M05B5.2* that coordinate its expression *in vivo*.

Progressive 5` deletion analysis using the *M05B5.2*::*GFP*::*HIS2B* (reporter containing 1 kb of promoter sequence) and mutM05 (reporter containing 1 kb of promoter sequence with the PHA-4 site mutation) constructs identified a 70 bp region that is critical for M05B5.2 expression (Figure 11). The 5`Δ400M05, 5`Δ400mutM05, 5`Δ270M05 and 5`Δ270mutM05 constructs are still expressed in all of the same tissues as *M05B5.2*::*GFP*::*HIS2B* and mutM05, whereas the 5`Δ200M05 construct does not express in any tissues in *C. elegans* (Figure 11). Therefore, the critical region for expression in the pharynx, intestine, hindgut and somatic gonad falls between 270 bases and 200 bases upstream of the predicted ATG.

An internal deletion of a 50 bp portion of this region (Region I; between 250 and 200 base pairs upstream of the ATG) in the context of the full-length 1 kb promoter, Δ50bpM05, also abolished expression in the intestine, hindgut and somatic gonad and decreased pharyngeal expression (Figures 11 and 13). This result suggests that this 50 bp region is required for normal expression of *M05B5.2*. There are still two predicted

upstream PHA-4 sites present in the full-length promoter after deletion of Region I that may be responsible for the low level of pharyngeal expression that is seen.

Progressive deletions from the 3` end identified a second region (PIG4) that is also required for expression in the pharynx, intestine, hindgut and somatic gonad (Figure 12). The expression in the hindgut and somatic gonad is abolished while the pharyngeal and intestinal expression is significantly reduced in animals containing the 3`Δ200M05, 3`Δ50M05 and 3`Δ50mutM05 reporters (Figure 12). In contrast, the expression of the 3`Δ25M05 is the same as the *M05B5.2*::*GFP*::*HIS2B* reporter (Figure 12). Deletion of the region between 175 bp and 150 bp upstream of the predicted start codon (ΔPIG4) also abolishes hindgut, somatic gonad expression while reducing pharyngeal and intestinal expression (Figures 12 and 13). The full length *M05B5.2*::*GFP*::*HIS2B* reporter with both critical regions Region I and PIG4 highlighted can be seen in Figure 14.

I have found that Region I, but not PIG4, is sufficient to activate GFP expression in some tissues. I placed three copies of either Region I or PIG4 in a "promoter-less" reporter, pPD95.77 (Wenick and Hobert, 2004) to determine if these elements can activate expression in any tissues. Three copies of PIG4 did not result in any expression (Table 5), whereas three copies of Region I drove expression in the hindgut and throughout the pharynx of *C. elegans* (Figure 15). Surprisingly, Region I is only sufficient for pharyngeal and rectal expression, while *M05B5.2* is expressed in other tissues (the intestine and somatic gonad). This suggests that any smaller regulatory element within Region I that is sufficient to drive expression will only drive expression in the pharynx and hindgut.

After these two critical regions were identified, I looked for transcription factor binding sites to determine possible *cis*-regulatory elements. In order to identify individual

sites I compared these critical regions from *C. elegans* to the promoters from the homologous genes in two related nematodes, *C. briggsae* and *C. remanei*. I compared these regions using the linear alignment program, ClustalW (Higgins et al., 1994; http://www.ebi.ac.uk/clustalw/) and the diagonal plot program, Dotlet using a variety of different settings (Junier and Pagni, 2000; http://myhits.isb-sib.ch/cgi-bin/dotlet). The alignment of PIG4 using ClustalW and Dotlet did not find any conserved elements that suggest binding sites for further analysis, while the alignments of Region I found four conserved sequences. Each of these sequences resembles binding sites for known transcription factors or previously identified elements. From 5' to 3' the sequences are a predicted PHA-4 binding site (PIG1), a possible bZIP site (PIG2), a sequence resembling the "Early-1" promoter element (PIG3) and a possible GATA factor binding site (Figure 16). I performed further analysis of these conserved elements by point mutation to determine what role they played in regulation of M05B5.2 expression.

Point mutations identify three discrete* cis*-regulatory elements

Mutations of the conserved elements within Region I have identified three discrete *cis*-acting regulatory elements that are critical for the normal expression of the *M05B5.2* reporter. As these elements are required for pharyngeal, intestinal and gonadal expression I have named them PIG1-3 (Pharyngeal, Intestinal and Gonadal regulatory elements 1-3). PIG1 is the predicted PHA-4 binding site (Gaudet and Mango, 2002), PIG2 is the possible bZIP binding site (Patel, 1999) and PIG3 is the site that resembles the “Early-1” promoter element (Gaudet et al., 2004). Mutating the conserved GATA-like site (from TGATAA to the *Fsp* I restriction site TGCGCA) did not affect expression in

either mutGATAM05 or mutGATAmutM05, thus this site is not necessary for expression of M05B5.2 (Figure 17).

Mutation of either the Early-1-like PIG3 element (from GGAGACAAG to the *Nsi* I restriction site GGATGCATG) or the bZIP-like PIG2 element (from TGACTCA to the *Spe* I restriction site TGACTAG) reduced or abolished GFP expression in all tissues. Mutation of these sites in plasmids mutPIG3M05 and mutPIG2M05 resulted in a loss of somatic gonad and hindgut expression and a decrease in intestinal and pharyngeal expression (Figures 17 and 18; Table 5). As well, mutation of PIG2 or PIG3 in the context of the hypodermally expressed reporter mutM05 abolished the ectopic hypodermal expression.

Interestingly, these mutations did not completely abolish pharyngeal expression, possibly due to the presence of additional promoter elements in the 1 kb promoter (such as two predicted PHA-4 sites upstream of the critical Region I). Therefore, I tested the effect of the PIG2 and PIG3 mutations in a truncated reporter lacking the upstream PHA-4 sites to determine their effects on pharyngeal expression. Mutation of either the PIG2 or PIG3 sites in a truncated promoter that lack all PHA-4 sites (mutPIG2truncM05 and mutPIG3truncM05) abolished expression in the pharynx (Figures 17 and 18). Therefore, both PIG2 and PIG3 are required for pharyngeal expression in the absence of upstream promoter sequence, arguing that PIG2 and PIG3 are necessary for expression of *M05B5.2* in all tissues in which the gene is expressed.

Finally, I looked at the fourth conserved site in Region I, the PHA-4 site (PIG1). As described above, mutation of this site leads to decreased pharyngeal expression and ectopic hypodermal expression (Figure 5, 17 and 18). Mutation of this site in a truncated promoter (mutPIG1truncM05) reduced pharyngeal expression, but did not abolish it,

suggesting that other elements (such as PIG2 and PIG3) can function in the absence of PHA-4 sites, but at a decreased level (Figure 17). Thus, PIG1 is involved in normal pharyngeal expression as well as inhibition of hypodermal expression (Table 5).

Enhancer assays identify one sufficient regulatory site

After finding that PIG1-3 are each necessary for expression, I tested these elements to determine if they are also sufficient for activation of transcription and found that only PIG1 is sufficient to drive expression. I tested each of these elements for sufficiency by placing three copies in front of the same "promoter-less" reporter with which I tested the sufficiency of Region I and PIG4.

Neither PIG2 nor PIG3 were sufficient to drive expression in any tissues in *C. elegans*. Placing three copies of PIG3 and the three bases flanking this site (ATAGGAGACAAGTTT; 3xPIG3) or three copies of PIG2 and the flanking bases (CACTGACTCATAC; 3xPIG2) in the 'promoter-less' reporter did not drive fluorescent expression in any tissue. Next, I created an enhancer with a combination of PIG2 and PIG3 sites since previous work has shown that combinations of sequences can act synergistically to activate pharyngeal expression (Gaudet et al. 2004; Okkema et al., 1993; Raharjo and Gaudet 2007). However, the combination of 3xPIG2 and 3xPIG3 placed upstream of the "promoter-less" GFP reporter did not result in any observable expression. Therefore, neither PIG2 nor PIG3 are sufficient alone or in combination to activate reporter expression (Table 5).

Finally, I tested three copies of PIG1 (TATGTTGACTCTT; 3xPIG1) and found that it was sufficient to activate expression in the pharynx and hindgut (Figure 21; Table 5). The expression of this enhancer is identical to that of the multimerized Region I

enhancer, suggesting that this site is solely responsible for the activity of the larger enhancer. It has been shown that the predicted PHA-4 site contained within the PIG1 element is able to bind PHA-4 *in vitro*, so that it is likely that PHA-4 is binding to activate expression of *M05B5.2* (Gaudet and Mango, 2002).

The expression of three copies of the PIG1 site is much stronger than that of other PHA-4 sites of other genes (Gaudet Lab, personal communication; Kalb et al., 2002), suggesting that this site may be more than simply a PHA-4 site. As well, analysis of the site has identified a second potential binding site for an HNF6/ONECUT homeodomain class of transcription factors that overlaps the PHA-4 site. PIG1 contains a predicted PHA-4 site (TRTTKRY) and an overlapping ONECUT-like site (DHWATTGAYTWWD) (Figure 19; Gaudet and Mango, 2002; Overdier, Porcella and Cost, 1994; Samadani and Costa, 1996). To test whether both sites are important, I generated separate mutations that affected only the predicted PHA-4 (PIG1-PHA-4(-)) or only the HNF6/ONECUT binding site (PIG1-ONECUT(-)).

Only the ONECUT-like site of PIG1 is required for repression of *M05B5.2* expression in the hypodermis. The specific mutation of PIG1 to TATGTTTGCAGCA, which removed the ONECUT-like binding site while leaving the PHA-4 site intact (PIG1-ONECUT(-)), turned on ectopic hypodermal expression (Figure 20). In contrast, a mutation of PIG1 to TAAGTTGACTT that disrupts the PHA-4 site while leaving the ONECUT-like site intact (PIG1-PHA-4(-)) had no effect on expression (Figure 20). Neither of these mutations affected intestinal, hindgut or somatic gonadal expression (Figure 20).

Given that PIG1 appears to be composed of two overlapping transcription factor binding sites (for PHA-4 and a possible ONECUT family member), I next tested whether

the individual sites were sufficient for activation of reporter expression in the pharynx. Neither three copies of the PIG1-ONECUT(-) nor three copies of the PIG1-PHA-4(-) sites activated GFP expression (Figure 21). Therefore, neither one of these two sites alone were sufficient to drive fluorescent expression. One interpretation of these results is that neither site alone is sufficient for activation of expression. Alternatively, the altered sequences used in these experiments may have a lower relative affinity for their respective binding factors compared to the wild type sequences in the *M05B5.2* promoter and are thus less active as enhancers. To distinguish between these two possibilities, I placed the 3xPIG1-ONECUT(-) and 3xPIG-1-PHA-4(-) sequences adjacent to one another in an enhancer construct (3xPIG1-ONECUT(-)::3xPIG1-PHA-4(-)); if the sites function in combination then this construct should recapitulate the activity of the 3xPIG1 construct. Interestingly, the combination of sites was able to activate robust pharyngeal expression of GFP (Figure 21), which indicated that both sites are required for pharyngeal activation. These results suggest that both PHA-4 and a factor binding to the ONECUT-like site are required for activation through PIG1.

Inhibition of other hypodermal constructs by PIG1 and its component sites

The analysis of PIG1 suggests that it contains two overlapping sites, both a PHA-4 and an ONECUT-like binding site. To further understand how the hypodermal repression of M05B5.2 occurs, I tested this site in a repressor assay with another hypodermally expressed reporter, *dpy-7* (Gilleard, Barry and Johnstone, 1997). *dpy-7* is a collagen gene that is expressed specifically in the hypodermis and its minimal promoter is the smallest portion of upstream sequence that is sufficient to drive GFP expression in the hypodermis. This minimal promoter has also been found to respond to added

elements: addition of a neuronal *cis*-element to the *dpy-7* promoter results in expression in both hypodermal and neuronal cells (Wenick and Hobert, 2004). To test the PIG1 site for inhibition of hypodermal expression I placed three copies of it in front of the minimal promoter of *dpy-7*. The addition of the 3xPIG1 fragment to the minimal promoter was able to abolish the hypodermal expression of the *dpy-7* reporter (Figure 22B).

Next, to determine whether either the PHA-4 or ONECUT-like site alone was responsible for the repression, I placed the 3xPIG1-ONECUT(-) or 3xPIG1-PHA-4(-) sequences in front of the *dpy-7* reporter (3xPIG1-ONECUT(-)::*dpy*-7::*GFP* or 3xPIG1-PHA-4(-)::*dpy*-7::*GFP*). Neither of these sites appeared to affect hypodermal expression (Figure 22D and E), suggesting that neither the ONECUT-like nor the PHA-4 site alone is sufficient to inhibit hypodermal expression of *dpy-7*. Finally, I found that placing the 3xPIG1-ONECUT(-)::3xPIG1-PHA-4(-) enhancer fragment containing tandem 3xPIG1-ONECUT(-) and 3xPIG1-PHA-4(-) sites in front of the *dpy-7* minimal promoter (3xPIG1-ONECUT(-)::3xPIG1-PHA-4(-)::*dpy-7*::*GFP*) abolished hypodermal expression (Figure 22C). Therefore, both the ONECUT-like and PHA-4 binding sites are required for inhibition of the hypodermal expression of the *dpy-7* minimal promoter. This result, that both the PHA-4 and ONECUT-like sites are required for hypodermal inhibition of *dpy-7*, contrasts the previous result that mutating only the ONECUT-like site led to ectopic hypodermal expression while mutation of only the PHA-4 site did not.

Overexpression of M05B5.2 in hypodermal tissue

Finally, to study whether repression of M05B5.2 in the hypodermis is biologically important, I placed the M05B5.2 coding region under the control of the mutM05 promoter, which is active in the hypodermis (as well as the pharynx, intestine, hindgut

and somatic gonad). One possibility is that expression of M05B5.2 in the hypodermis is detrimental and must be repressed; if so, then the mutM05::M05B5.2(+) transgene may result in a phenotype. However, this construct had no observable effect on wild-type animals: transgenics were viable and healthy with no apparent defects or lethality. This suggests that the hypodermal expression may not be detrimental to the worm's survival, though it is also possible that the product is not made in the hypodermis, leaving it unclear whether hypodermal inhibition of *M05B5.2* expression is critical in *C. elegans.*

Identifying the upstream regulators of *M05B5.2* expression

Having identified four critical regulatory sites in the 1 kb of *M05B5.2* promoter, I next used a candidate approach to identify transcription factors that may be binding to these sites. I found that an HNF6/ONECUT factor, CEH-39, is likely acting through the PIG1 to activate expression in the pharynx.

Identification of Candidate factors that may act through PIG4

Previous analysis of the PIG4 region through linear alignment and diagonal alignment software programs comparing this region from *C. elegans* to the entire promoter from *C. briggsae* and *C. remanei* did not reveal any obvious conservation. As well, the PIG4 region does not resemble any known transcription factor binding sites through analysis with ProScan (www-bimas.cit.nih.gov/molbio/proscan/), Transfac (http://www.gene-regulation.com/pub/databases.html#transfac) and Tfsitescan (www.ifti.org). This does not suggest any obvious candidate factors that might act through PIG4, so no further analysis was performed.

Identification of Candidate factors that may act through PIG3

PIG3 closely resembles a previously identified element named "Early-1" (Gaudet et al. 2004) that appears to function as an enhancer in pharyngeal genes that are expressed relatively early in development. The *M05B5.2*::*GFP*::*HIS2B* reporter is expressed very early during embryogenesis, which fits with the model that the PIG3 element is the same as the Early-1 element. The Early-1/PIG3 element has also been found in promoters of neuronal genes (Ruvinsky et al., 2007), muscle genes (D. Miller, personal communication), heat shock response genes (GuhaThakurta et al., 2002) and ethanol stress response gene (Kwon et al., 2004). This suggests that PIG3 is a general promoter element present in a multitude of promoters and is not a pharynx specific regulatory element.

Molecular and genetic screens to identify the correct transcription factor binding to the Early-1 site have not yet been successful. Yeast-1-hybrid screens for a factor binding to the Early-1 site identified the repressor SMAD, DAF-3, as a candidate factor that binds Early-1 (M. Walhout, personal communication). The SMADs are a group of proteins that usually act as transcription factors to regulate gene expression in response to TGFβ signalling ligands (Savage-Dunn, 2001). However, work by Jeb Gaudet had found that Early-1 activity is not altered in *daf-3* mutants, which suggests that DAF-3 is either not the relevant *trans*-acting factor or acts redundantly with some other factor. One possibility is that Early-1 is a binding site for a SMAD other than DAF-3.

To examine the possibility that a SMAD other than DAF-3 may be the Early-1/PIG3 transcription factor, I tested my reporters in *daf-4* mutants, in which all TGF-β signalling is defective. TGF-β signalling acts through two receptor serine/threonine protein kinases (receptor types I and II) and a family of receptor substrates, the Smad

proteins (Massague and Chen, 2000). In *C. elegans* there is only one type II receptor, DAF-4, which is required for all TGF-β signalling events (Savage-Dunn, 2001). However, the *daf-4* mutant did not affect expression of *M05B5.2*::*GFP*::*HIS2B* or mutM05 (Figure 23). Thus, the PIG3/Early-1 element is probably not a Smad-binding site and the identity of the relevant factor remains unknown.

Identification of Candidate factors that may act through PIG2

The PIG2 element (TGACTCA) is a perfect match to the binding site of AP-1, a member of the basic leucine zipper (bZIP) family of transcription factors (Patel, 1999). There are 32 genes in *C. elegans* (Table 5) that encode predicted bZIP factors (Reece-Hoyes et al., 2005). Six of the bZIP factors are good candidates for acting through PIG2 because they are expressed throughout the digestive tract based on expression and *in situ* data (*atf-2, zip-2, sknr-1,* C27D6.4, ZC376.7 and T24H10.7). Nine of the bZIP factors are excluded as candidates based on expression and *in situ* data because they are not expressed in the digestive tract, whereas the expression of the other 17 factors is either broad or unknown.

I found that the first candidate factor *atf-2* is not the general regulatory factor required for *M05B5.2* activation in all tissues. Injection of *atf-2* dsRNA into the strains carrying the *M05B5.2*::*GFP*::*HIS2B* and mutM05 reporters did not affect expression of either of these two reporters (Figure 24). As well, the expression of *M05B5.2*::*GFP*::*HIS2B* and mutM05 was not affected when placed in a mutant *atf-2* strain (Figure 24).

Further tests of the other five likely and 18 possible candidate factors will need to be done to determine which may be activating expression through the PIG2 site.

However, there are many possible candidates and they could be redundant so that further analysis will require a lot more work.

Identification of Candidate factors that may act through PIG1

The PIG1 site is necessary to inhibit expression in the hypodermis and thus an inhibitory factor(s) must bind this site. This site (ONECUT-like and PHA-4) is also sufficient to drive expression in the pharynx and in a truncated promoter it is required for full activation of pharyngeal expression. Therefore, a positive regulator must be acting through this site to turn on expression. The similarity of PIG1 to the binding site of both PHA-4/forkhead family members and the ONECUT family of transcription factors suggested that such factors are candidates for acting through PIG1. My previous data (Figure 21) has shown that both portions of PIG1 (the ONECUT-like site and the PHA-4/forkhead site) are required for PIG1 repressor and activator functions, suggesting that at least two factors function through PIG1.

Forkhead factors

First, I looked at the forkhead factors in *C. elegans* to test whether one of these functions as the negative regulator of M05B5.2 hypodermal expression, acting through PIG1. There are seventeen forkhead genes in *C. elegans,* shown in Table 7 (Reece-Hoyes et al., 2005). Of these genes, seven are good candidates (*unc-130, daf-16, fkh-3, fkh-4, fkh-5, fkh-6, fkh-7* and *fkh-8*) whereas the other ten forkhead factors, including PHA-4, are ruled out based on previous SAGE, *in situ* and expression data.

Based on my data, none of the candidate forkhead factors appear to act as inhibitors of hypodermal expression. To test for negative regulation, I looked for induced ectopic expression in the hypodermis of the *M05B5.2*::GFP::HIS2B reporter due to the

loss of the forkhead candidate factor(s). I used both RNAi and mutant strains to test the genes alone and in combination (Table 8). During this analysis I tested some combinations of forkhead factors as they have been hypothesized to act redundantly due to similarity in protein sequence and/or expression (Hope et al., 2003). I found that none of the forkheads alone or in combination led to ectopic hypodermal expression of the *M05B5.2*::*GFP*::HIS2B reporter.

HNF6/ONECUT factors

Next, I looked at members of the HNF6/ONECUT family of transcription factors as possible regulators of M05B5.2 acting through the PIG1 site. There are five possible inhibitory candidate HNF6/ONECUT transcription factors in *C. elegans* as shown in Table 9. The genes *ceh-21, ceh-38, ceh-39* and *ceh-49* are all expressed in the hypodermis (Cassata et al., 1998; http://nematode.lab.nig.ac.jp/dbest/keysrch.html) and the expression of c*eh-48* has not been previously determined. The three likely candidates for pharyngeal activation are also listed in Table 9, they are the only ones that are expressed in the pharynx (Cassata et al., 1998; http://nematode.lab.nig.ac.jp/dbest/keysrch.html). To test these candidates I used mutant strains or RNAi injections. For positive regulation I tested the effect of the loss of a ONECUT factor on the expression of the 3xPIG1 enhancer, expecting that the knockout of the correct activating factor would abolish enhancer expression. For negative regulation, I looked for induced ectopic expression in the hypodermis of the *M05B5.2*::*GFP*::*HIS2B* reporter due to the loss of the ONECUT candidate factor(s).

I was not able to identify the negative regulator of hypodermal expression by testing the ONECUT candidate factors. By injecting double stranded RNA to *ceh-21* or

ceh-48, I found that neither factor is the negative regulator of hypodermal expression (data not shown). As well, analysis of the mutant strains of *ceh-38, ceh-39* and *ceh-49* found that none of these factors are the negative regulator (data not shown). *ceh-39, ceh-21* and *ceh-41* encode very similar proteins and are found in an operon in the *C. elegans* genome, which suggests that they may be acting redundantly. To ensure that I knocked out any redundant activity of these factors, I injected *ceh-21* and *ceh-41* double stranded RNA into a *ceh-39* mutant strain expressing the *M05B5.2*::*GFP*::*HIS2B* transgene. The removal of all three of these ONECUT factors did not turn on ectopic hypodermal expression (data not shown).

While I was unable to identify a factor affecting hypodermal repression of M05B5.2, I found that the positive regulatory activity of PIG1 requires the HNF6/ONECUT factor *ceh-39*, suggesting that the CEH-39 product may act directly through PIG1. Interestingly, the expression of the 3xPIG1 enhancer construct in *ceh-39* mutant strains was abolished in adult animals (Figure 25). Testing this enhancer in two different deletion alleles *ceh-39(gk329)* and *ceh-39(gk296)* had the same effect in adult animals, however the *ceh-39(gk329)* allele was still weakly expressed in the pharynx of 41% (37/90) of embryos and larvae, whereas animals homozygous for the *ceh-39(gk296)* did not show expression in animals at any developmental stage. The expression difference between these two strains indicates that there may be a difference in the activity of the two alleles. However, the loss of adult and most earlier expression of the 3xPIG1 enhancer suggests that CEH-39 is critical for normal PIG1 activity. To verify that the loss of expression was due to the *ceh-39* mutation, I outcrossed these strains to N2 animals. The progeny of heterozygous parents showed no expression of the transgene in the pharynx of 24% (ie 1/4) of animals indicating that only homozyogous mutant CEH-

39 animals lose expression of the 3xPIG1 enhancer. Expression of the 3xPIG1-ONECUT(-)::3xPIG1-PHA-4(-)::*GFP* enhancer construct was also abolished in *ceh-39* mutant adults (Figure 25). This result shows that the isolated sites behave much like the "composite" site and that the division of the PIG1 element into its two component sites is able to recapitulate activity of the original PIG1 element. As before, outcrossing these animals to wild-type animals restored expression of the enhancer construct.

To test whether *ceh-39* was required for PIG1 activity in the context of the *M05B5.2* promoter I tested the expression of a truncated *M05B5.2* promoter construct, 5`Δ270M05, in *ceh-39* mutants. The truncated promoter lacks upstream PHA-4 sites that may function redundantly with PIG1. I found that the pharyngeal expression of this transgene was lost in 94% of the expressing animals in the *ceh-39* mutant background (Figure 26). The remaining 6% only showed expression in one or two cells of the pharynx, representing a significant reduction in activity compared to wild type. When these transgenes were outcrossed to wild-type animals, expression was restored in 98% of animals, verifying that *ceh-39* activates *M05B5.2* expression (Figure 26).

The HNF6/ONECUT factor CEH-39 is required for PIG1 activity in the pharynx, and the PIG1 site also contains a ONECUT-like binding site. Therefore, it is likely that CEH-39 acts directly on this element to activate pharyngeal expression. Also, the loss of pharyngeal expression of the truncated M05B5.2 reporters shows that CEH-39 is required for pharyngeal expression of M05B5.2.

The model of M05B5.2 regulation

This thorough deletion, mutation and transcription factor analysis has identified four critical regulatory sites PIG1-4 for *M05B5.2* expression. I have also found one likely

activating factor, CEH-39, which drives expression of *M05B5.2* through the PIG1 site. The other three sites are likely general elements required for expression in all of the tissues in which *M05B5.2* is expressed. The factor binding to PIG2 is possibly a bZIP factor, the PIG3 site resembles the Early-1 element and the PIG4 element is a previously undetermined binding site. The final model of regulation with the identified transcriptional regulators is shown in Figure 27.

Table 4. The timing of the lethality in *M05B5.2(RNAi)* and *let-527*

Strain	Percentage of death as embryos or L1 larvae of animals that fail to develop to adulthood	
	Embryo	Larvae
KR523 (*let-527*)	81.2% (65/80)	18.8% (15/80)
M05B5.2(RNAi) in GD73	86.3% (113/131)	13.7% (18/131)
M05B5.2(RNAi) in SM469	92.4% (85/92)	7.6% (7/92)

Table 5. A summary of the four *cis*-acting regulatory sequences

cis-regulatory element	Necessary	Sufficient	Tissue(s) in which the element is acting
PIG1	Yes	Yes	pharynx, hypodermis
PIG2	Yes	No	pharynx, intestine, hindgut and somatic gonad
PIG3	Yes	No	pharynx, intestine, hindgut and somatic gonad
PIG4	Yes	No	pharynx, intestine, hindgut and somatic gonad

Table 6. The bZIP genes of C. elegans.

Gene	Gene name	Tissue expression or homolgy	Origin of expression data*	Candidate
K08F8.2	*atf-2*	pharynx, gut and hypodermis	*in situ*	Yes
K02F3.4	*zip-2*	pharynx, gut and hypodermis	*in situ*	Yes
C27D6.4		pharynx, gut, hypodermis and neurons	*in situ* and reporter data	Yes
T24H10.7		Best jun homolog, jun binds to the AP1 site		Yes
ZC376.7		pharynx and gut	*in situ*	Yes
W02H5.7	*sknr-1*	pharynx, intestine, hypodermis and nervous system	reporter data	Yes
Y75B8A.35	*zip-1*	pharynx, hypodermal and rectum	*in situ*	Possibly
T19E7.2	*skn-1*	maternal, gut, hypodermis and neurons	*in situ* and reporter data	Possibly
C07G2.2	*atf-7*	germline and gut	*in situ*	Possibly
F23F12.9		intestine	reporter data	Possibly
C34D1.5	*zip-5*	not known		Possibly
D1005.3		not known		Possibly
R07H5.10		not known		Possibly
T23C6.5		not known		Possibly
W07G1.3	*zip-3*	not known		Possibly
W08E12.1		not known		Possibly
Y44E3B.1	*zip-4*	not known		Possibly
ZK909.4	*ces-2*	not known		Possibly
F59B1.7	*srx-41*	not known		Possibly
Y41C4A.4	*crh-1*	broad	*in situ*	Possibly
Y51H4A.4		broad	*in situ*	Possibly
F45E6.2	*atf-6*	none visible	*in situ*	Possibly
C48E7.11		none visible	*in situ*	Possibly
F57B10.1		hypodermis	*in situ*	No
T27F2.4		gut	*in situ*	No
Y75B8A.29		gonad	*in situ*	No
ZC8.4	*lfi-1*	neurons	*in situ*	No
F45H11.4	*mgl-2*	neurons	*in situ* and reporter data	No
T04C10.4	*atf-5*	gonad	*in situ*	No
R74.3	*xbp-1*	neurons	*in situ* and reporter data	No
F17A9.3		hypodermis	*in situ* and reporter data	No
F29G9.4	*fos-1*	gonad	*in situ* and reporter data	No

* The *in situ* expression data comes from the Nematode Expression Pattern Database (http://nematode.lab.nig.ac.jp/db2/index.php) and the reporter expression data is from the BC Gene Expression Consortium (http://elegans.bcgsc.ca/home/ge_consortium.html) except for *fos-1* expression found in Sherwood et al. (2005).

Table 7. The forkhead factors of *C. elegans*.

Gene	Name	Tissue Expression	Origin of expression data*	Candidate
C47G2.2	*unc-130*	body wall muscle, tail and intestine	*in situ* and reporter data	Yes
R13H8.1	*daf-16*	pharynx, neurons, somatic gonad, muscles and vulva	*in situ* and reporter data	Yes
F40H3.4	*fkh-8*	hypodermis, pharynx, intestine, neurons	*in situ,* SAGE and reporter data	Yes
F26D12.1	*fkh-7*	hypodermis, pharynx, intestine, neurons	*in situ,* SAGE and reporter data	Yes
F26A1.2	*fkh-5*	broad	reporter data	Yes
C29F7.5	*fkh-4*	broad	*in situ* and reporter data	Yes
C29F7.4	*fkh-3*	broad	*in situ* and reporter data	Yes
T14G12.4	*fkh-2*	embryonic D and MS cells, no adult	reporter data	No
T28H11.4	*pes-1*	embryonic D and MS cells, no adult	reporter data	No
K10G6.1	*lin-31*	ventral ectodermal cells, male tail blast cells, excretory duct cell, tail neurons, neurons	reporter data	No
F26B1.7	*let-381*	neuronal	*in situ* and reporter data	No
C34B4.2		none	*in situ*	No
F38A6.1	*pha-4*	pharynx, intestine, hindgut, somatic gonad	*in situ* and reporter data	No
C25A1.2	*fkh-10*	neurons, pharynx and gonad	*in situ* and reporter data	No
K03C7.2	*fkh-9*	neurons and digestive tract	*in situ* and reporter data	No
B0286.5	*fkh-6*	spermatheca, oviduct sheath cells	reporter data	No
T27A8.2		neurons	reporter data	No

* The *in situ* expression data comes from the Nematode Expression Pattern Database (NexDb; http://nematode.lab.nig.ac.jp/db2/index.php), the reporter expression data and SAGE data is from the BC *C. elegans* Gene Expression Consortium with some of the expression data from Hope et al. (2003).

Table 8. The method used to test the forkhead factors as inhibitory candidates.

Gene(s)	Name	Tested by RNAi or mutant strain*	Inhibitory factor(s)
C29F7.4	*fkh-3*	RNAi	No
C29F7.5	*fkh-4*	RNAi	No
F26A1.2	*fkh-5*	mutant strain	No
C29F7.4/C29F7.5/F26A1.2	*fkh-3/fkh-4/fkh-5*	RNAi/RNAi/mutant strain	No
F26D12.1	*fkh-7*	RNAi	No
F40H3.4	*fkh-8*	RNAi	No
C47G2.2	*unc-130*	mutant strain	No
R13H8.1	*daf-16*	mutant strain	No

*The mutant strains and their genotypes that are used in this analysis are listed in Table 1.

Table 9. The ONECUT candidate factors in *C. elegans* for inhibition and activation through the PIG1 site.

Gene	Name	Previously determined expression by *in situ*	Inhibitory Candidate	Activating Candidate
F17A9.6	*ceh-49*	broad	Yes	Yes
T26C11.7	*ceh-39*	broad	Yes	Yes
F22D3.1	*ceh-38*	pharynx, gut, hypodermis	Yes	Yes
T26C11.6	*ceh-21*	hypodermis	Yes	No
C17H12.9	*ceh-48*	not known	Possibly	Possibly
T26C11.5	*ceh-41*	broad in embryo, none in adult	No	No

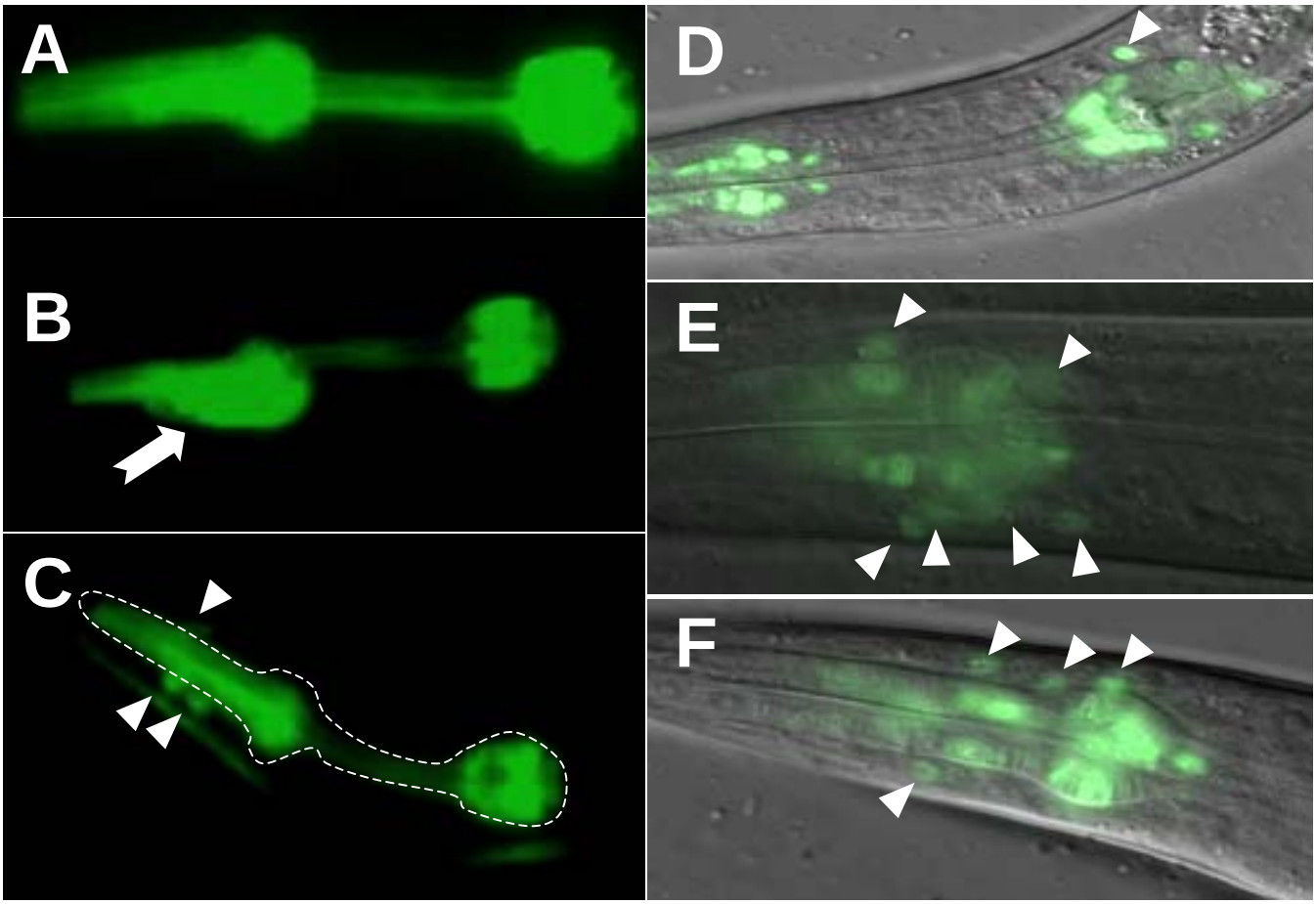
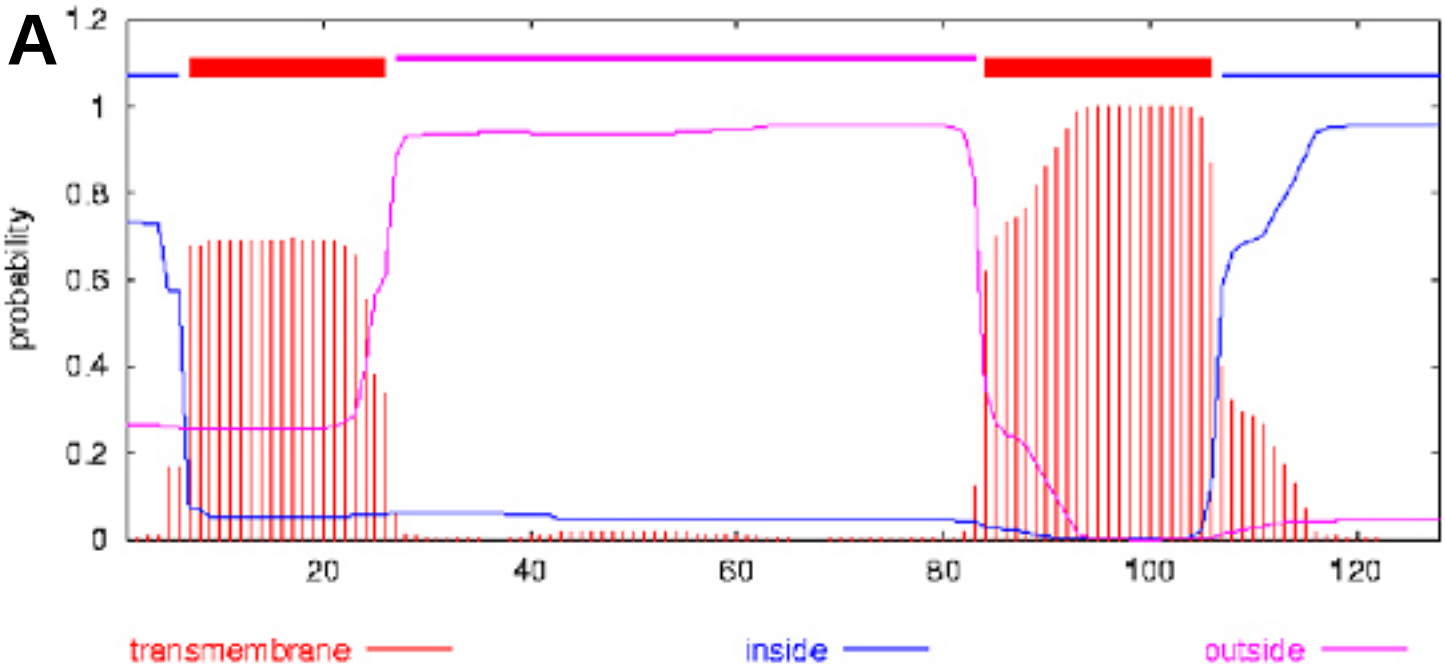


Figure 7. The pharyngeal phenotypes due to injection of double stranded RNA to M05B5.2. A) Wildtype pharynx expressing the *myo-2*::*GFP* reporter. B) *M05B5.2(RNAi)* induced pharyngeal morphology defects of the anterior bulb indicated by the arrow. C-F) The arrowheads indicate GFP expressing cells outside of the pharynx. C) Cells that express the *myo-2*::*GFP* reporter outside of the pharyngeal membrane after RNAi injection. D) Uninjected pharynx expressing the *pha-4*::*GFP* reporter showing that normally 1 to 2 cells express GFP outside of the pharynx. E-F) More than two *pha-4*::*GFP* expressing cells that are not bound by the pharyngeal membrane after RNAi injection.



B

Most likely signal peptide and its passenger peptide

1 MEIRLRHLAFA **11** AVLCTLCYYA **21** NG | QGLNDPIT **31** QYIEGRLGNR **41** RIFWCPSGYG **51** YLAFCPQPTD **61** WDNYNWCCTF **71** PYMGSWKPSC **81** CQFAIPTGAV **91** VAIILAAIVL **101** LLVLIAMSCW **111** CCWCCPLYKQ **121** LYDFEDE

Figure 8. The transmembrane and signal peptide predictions of M05B5.2. A) Transmembrane prediction from the TMHMM program [http://www.cbs.dtu.dk/services/TMHMM/] of the protein encoded by M05B5.2 showing the signal peptide as the first transmembrane site and the predicted transmembrane region as the second transmembrane site. B) The signal peptide prediction from SignalP [www.cbs.dtu.dk/services/SignalIP/] showing the entire protein encoded by M05B5.2 with the signal peptide in red and the rest of the protein in black.

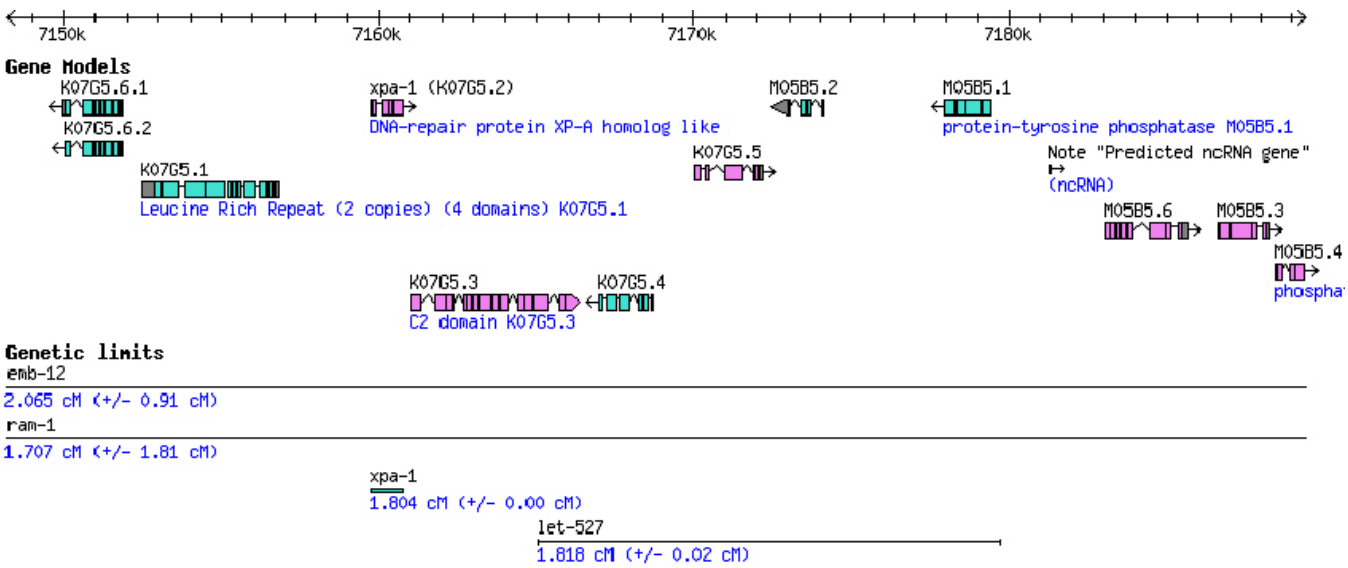


Figure 9. The chromosomal region to which *let*-527 is mapped. Figure from www.wormbase.org

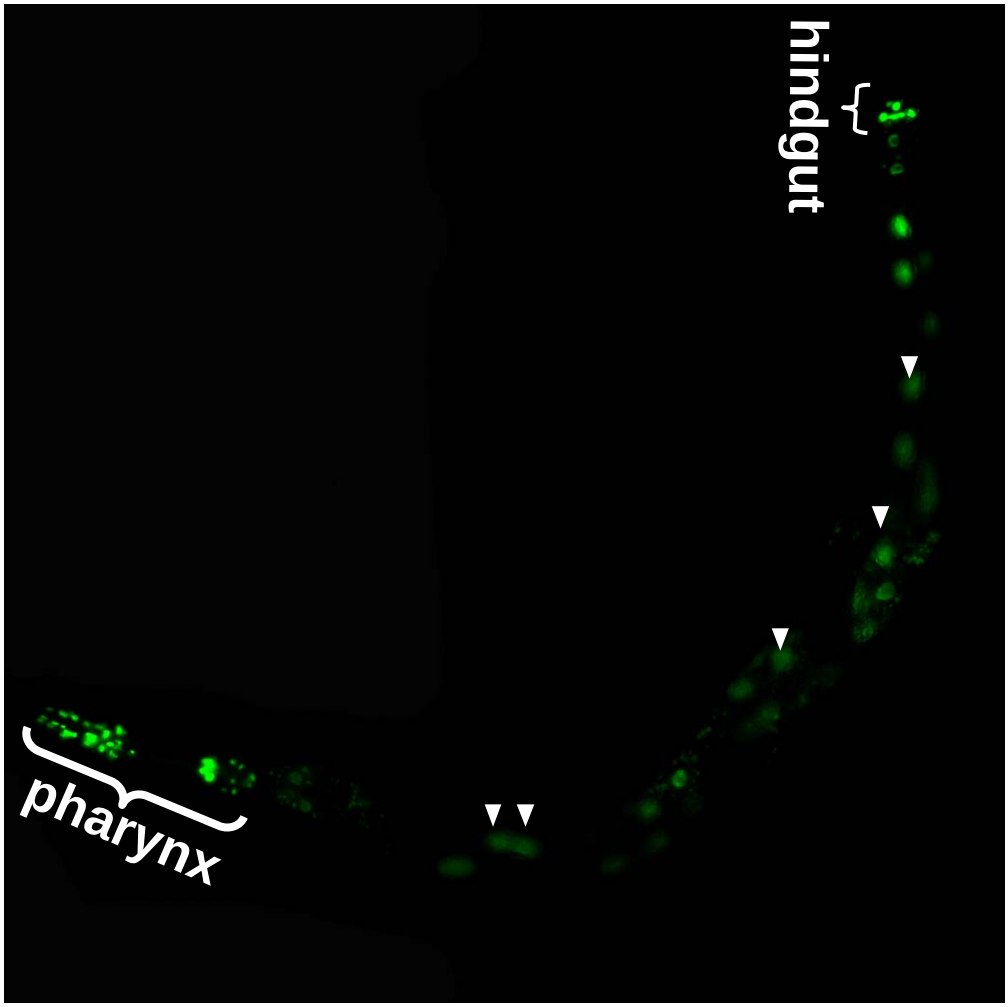


Figure 10. The expression of Cb-M05 in *C. briggsae*. Cb-M05 is expressed throughout the digestive tract in the pharynx, intestine and hindgut in *C. briggsae* similar to the expression of *M05B5.2*::*GFP*::*HIS2B* in *C. elegans.* The Cb-M05 reporter however does not express in the somatic gonad as does the *M05B5.2*::*GFP*::*HIS2B* reporter.

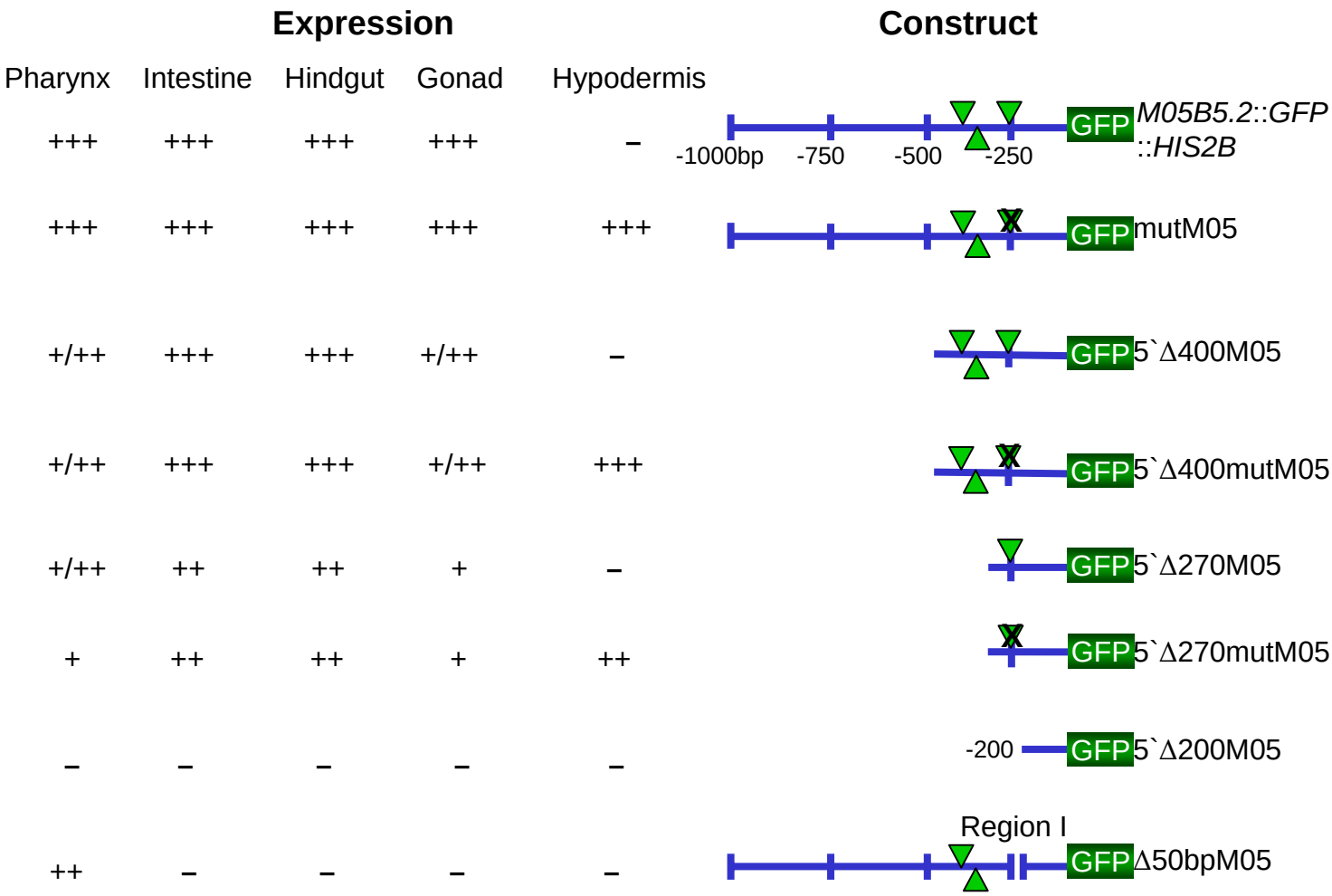


Figure 11. Promoter constructs and their expression patterns for the 5` deletion analysis of M05B5.2. The green triangles represent PHA-4 sites and the Xs represent the mutated PHA-4 site that turns on ectopic hypodermal expression. The expression pattern is represented by +++ for strong expression, ++ for moderate expression, + for weak expression and – for no expression.

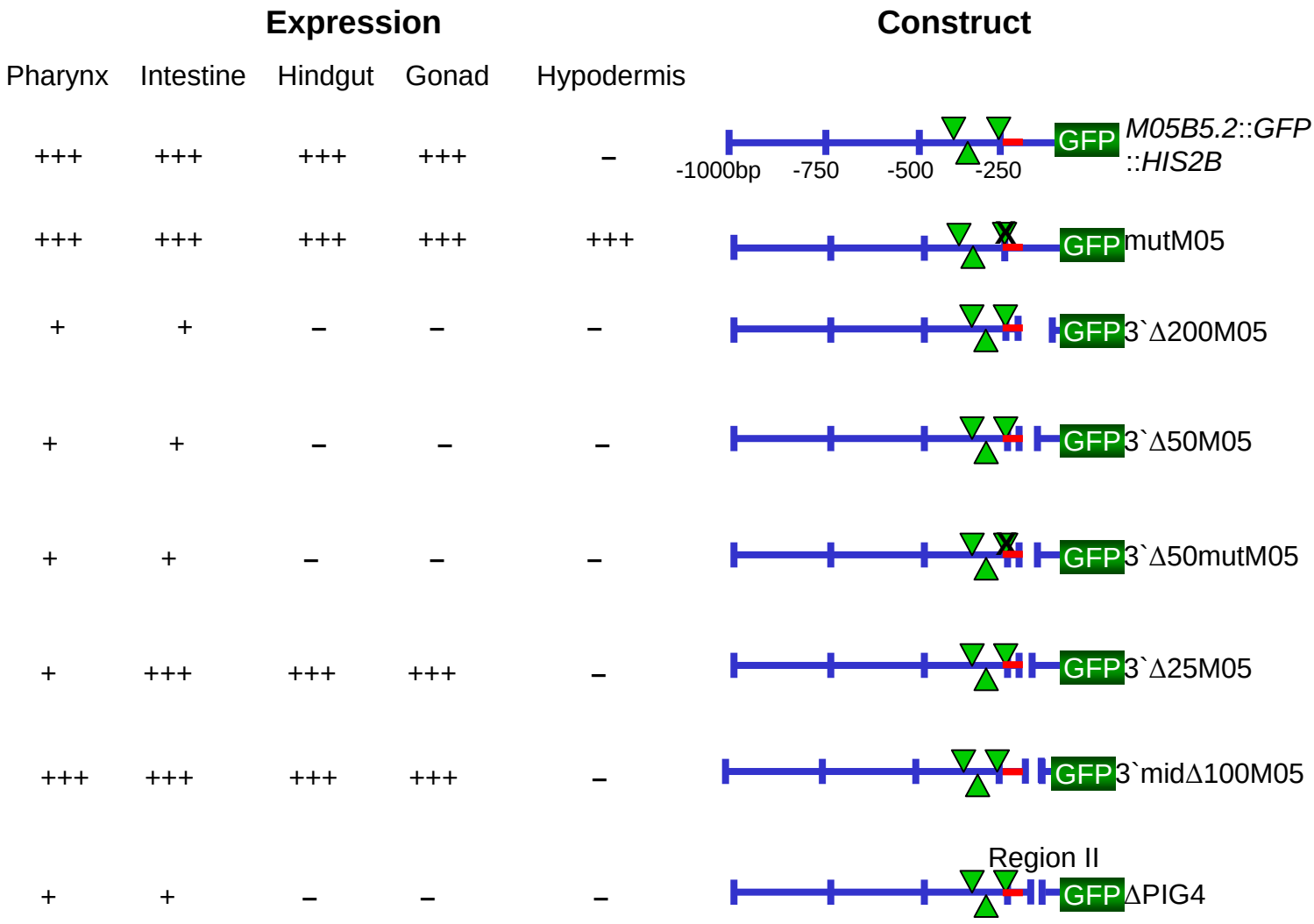


Figure 12. Promoter constructs and their expression patterns for the 3` deletion analysis of M05B5.2. The red bars represent Region I, the green triangles represent PHA-4 sites and the Xs represent the mutated PHA-4 site that turns on ectopic hypodermal expression. The expression pattern is represented by +++ for strong expression, ++ for moderate expression, + for weak expression and – for no expression.

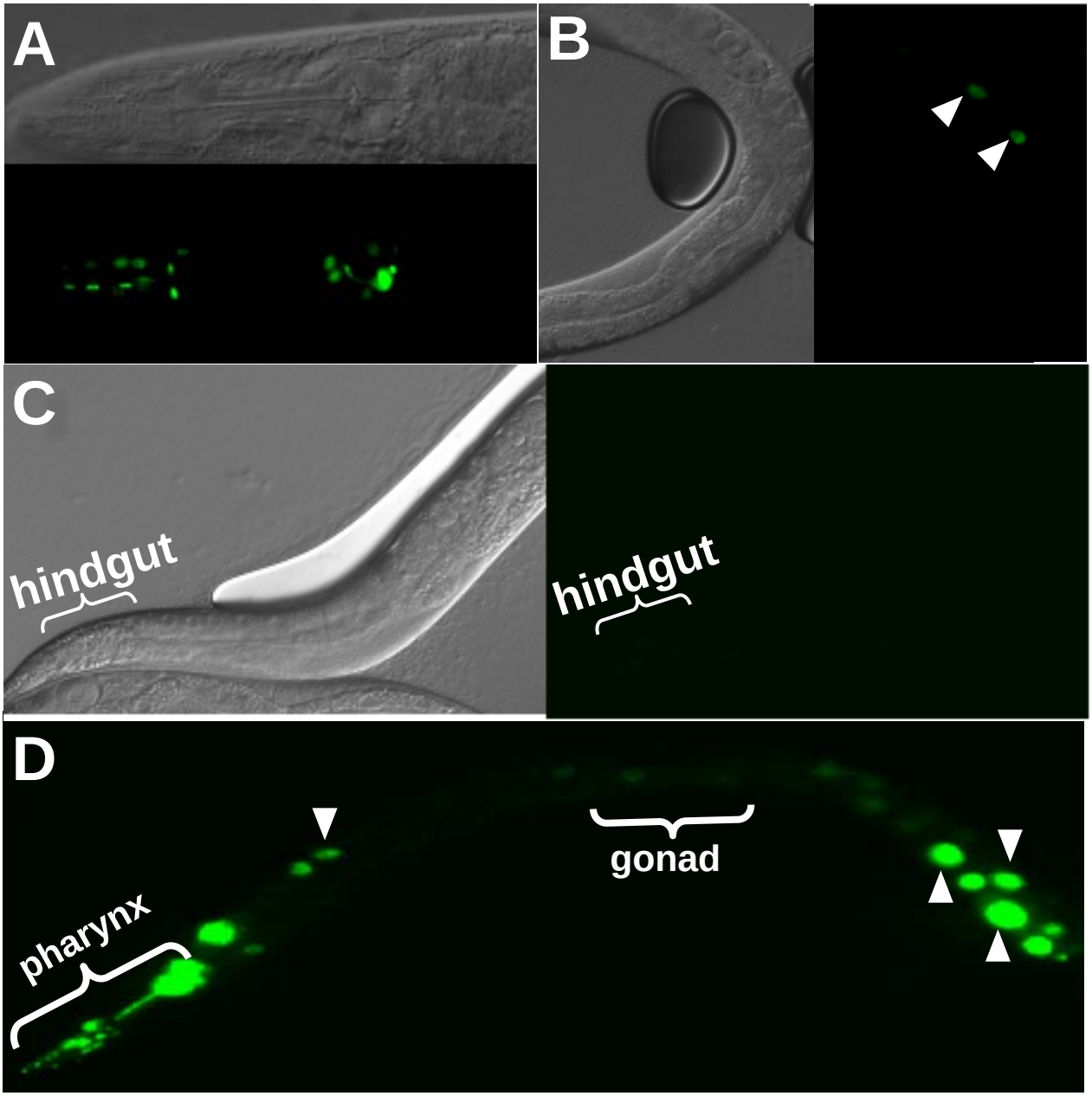


Figure 13. The expression of the full-length *M05B5.2* promoter fluorescent reporter with either Region I or PIG4 deleted. A-C) The deletion of Region I from the full-length promoter (Δ50bpM05) . A) Pharyngeal expression is weaker in reporters with Region I removed as compared to the full-length reporter. B) Intestinal expression is very weak or completely abolished when Region I is deleted. C) Hindgut expression is abolished when Region I is deleted. D) A reporter with PIG4 deleted (ΔPIG4) loses expression in the hindgut and somatic gonad. This reporter is still expressed in the pharynx and intestine, but weaker than the full-length reporter.

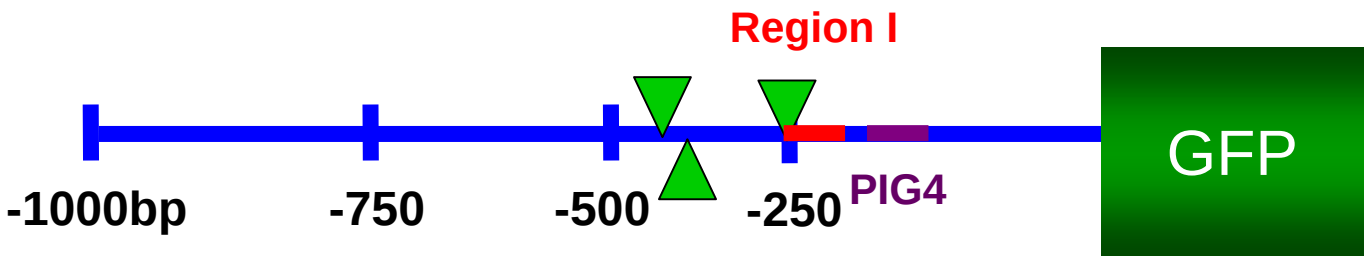


Figure 14. The *M05B5.2*::*GFP*::*HIS2B* reporter showing both Region I and PIG4. Region I is highlighted in red and PIG4 is highlighted in purple

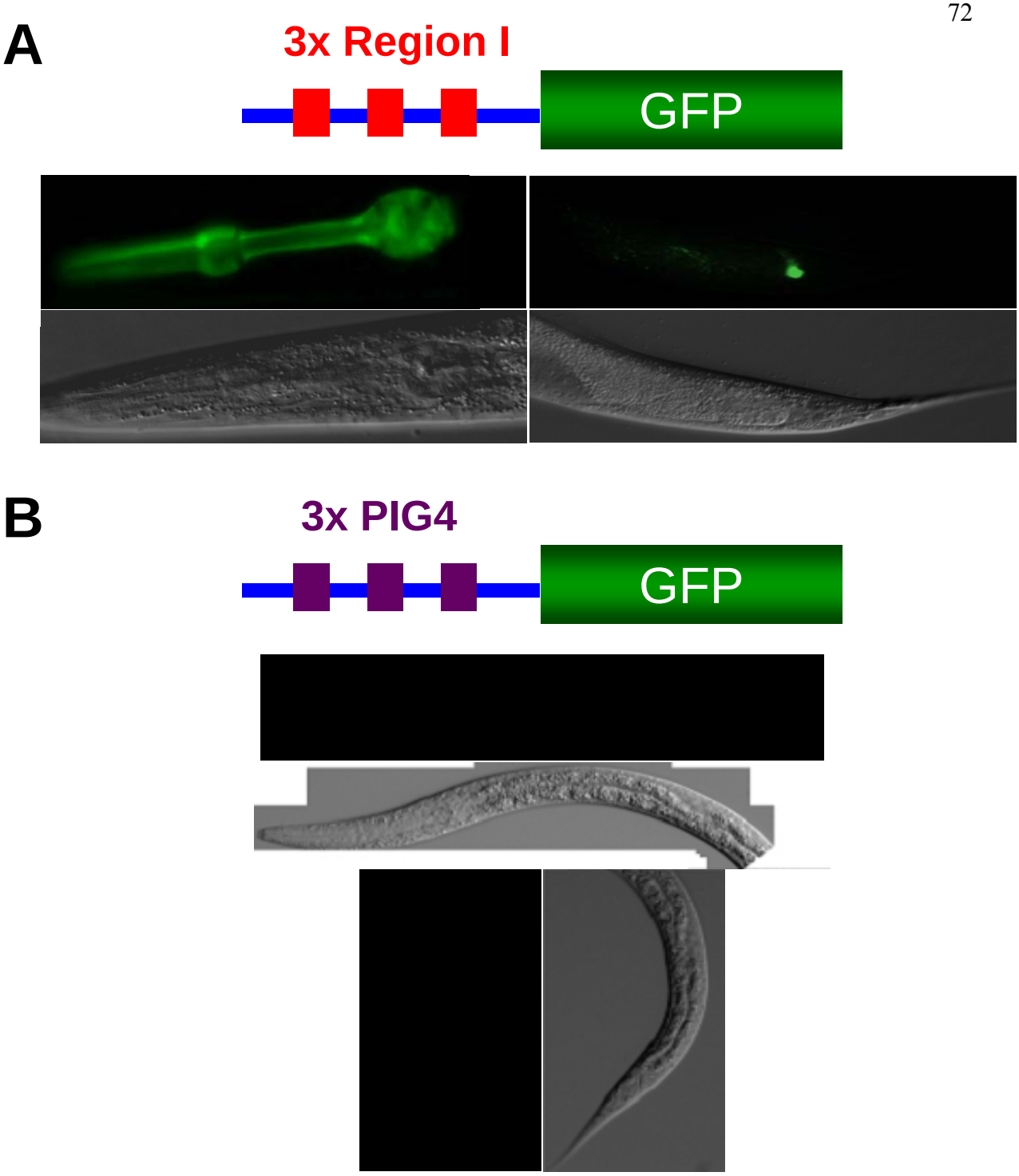


Figure 15. The expression of enhancer constructs containing multiple copies of the critical regions, Regions I and II from the M05B5.2 promoter. A) The 3xRegion I enhancer construct containing three copies of the critical 50bp region expresses in the pharynx, pharyngeal-intestinal valve and rectal gland cells. B) 3xPIG4 enhancer construct does not express in any tissues in *C. elegans*.

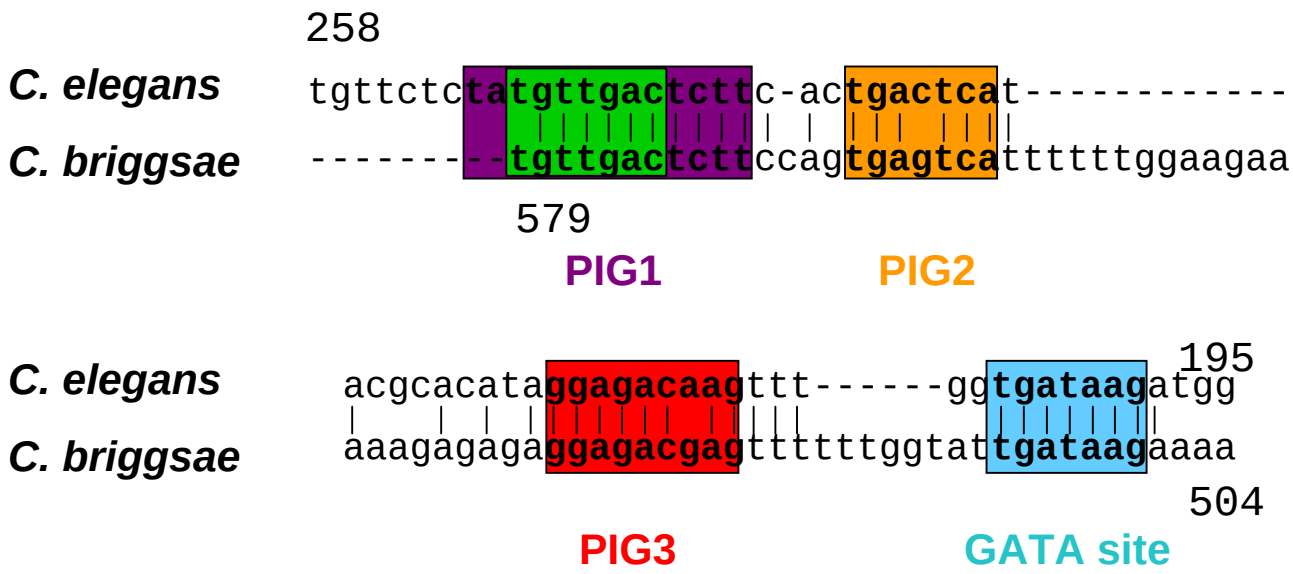


Figure 16. Comparison between *C. elegans M05B5.2* and its homologous gene in *C. Briggsae.* The *C. elegans* sequence is shown on top and the *C. briggsae* sequence is on the bottom with conserved bases depicted by a vertical lines. The PHA-4 site is highlighted by a green box, PIG1 is highlighted by a purple box, PIG2 is highlighted by an orange box and PIG3 is highlighted by a red box. The numbers 258 and 195 are the number of DNA bases upstream of the predicted translational start site of this gene at which each end of this region occurs in *C. elegans*. The numbers 579 and 504 are the number of DNA bases upstream of the predicted translational start site of this gene at which each end of this region occurs in *C. briggsae*.

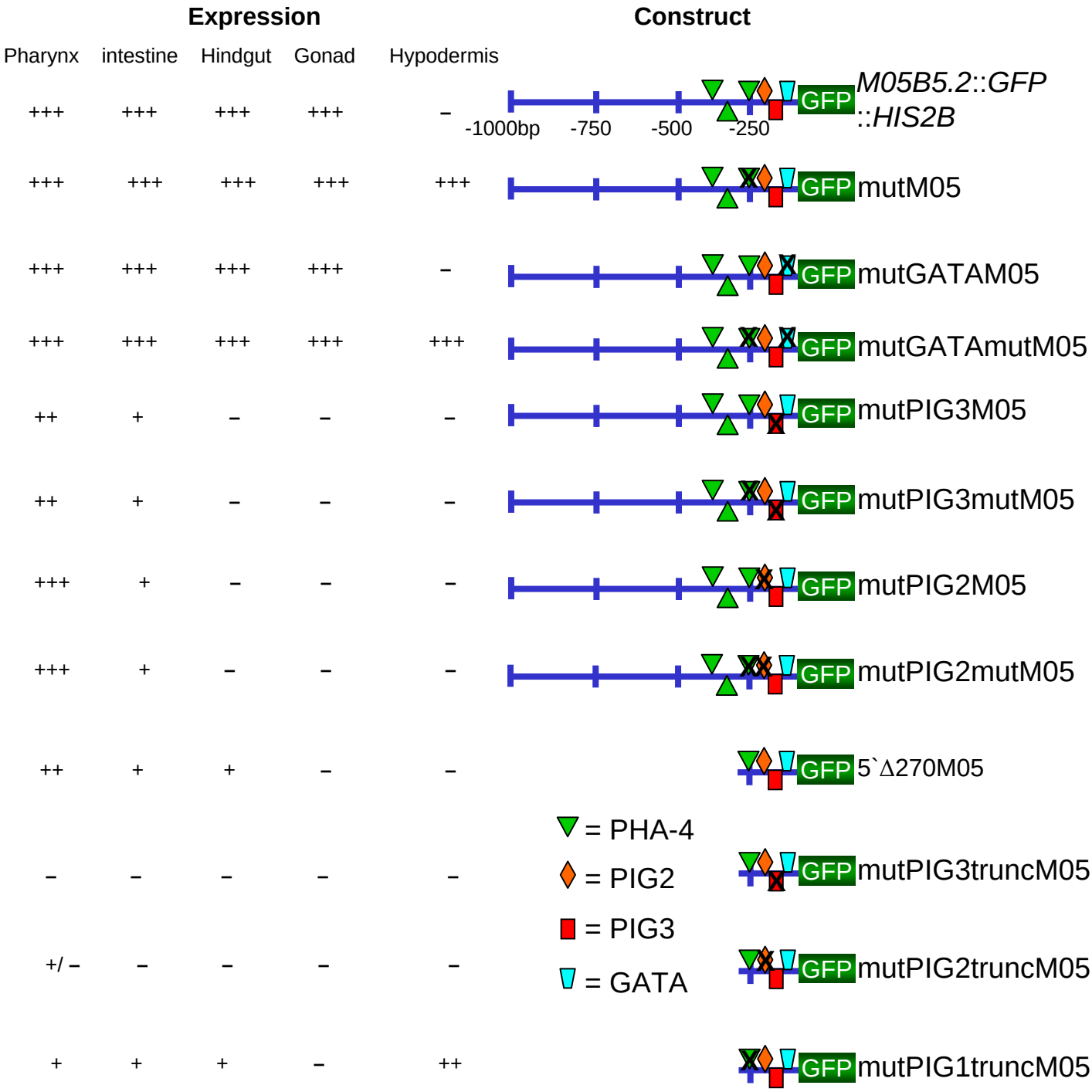


Figure 17. Reporter constructs and their expression patterns for mutation analysis of the conserved sites in the promoter of *M05B5.2*. The green triangles represent PHA-4 sites, the light blue trapezoid represent the GATA site, the orange diamonds represent the PIG2 site and red squares represent the PIG3 site. The expression pattern is represented by +++ for strong expression, ++ for moderate expression, + for weak expression and – for no expression. The X's represent the mutations of the different conserved sites.

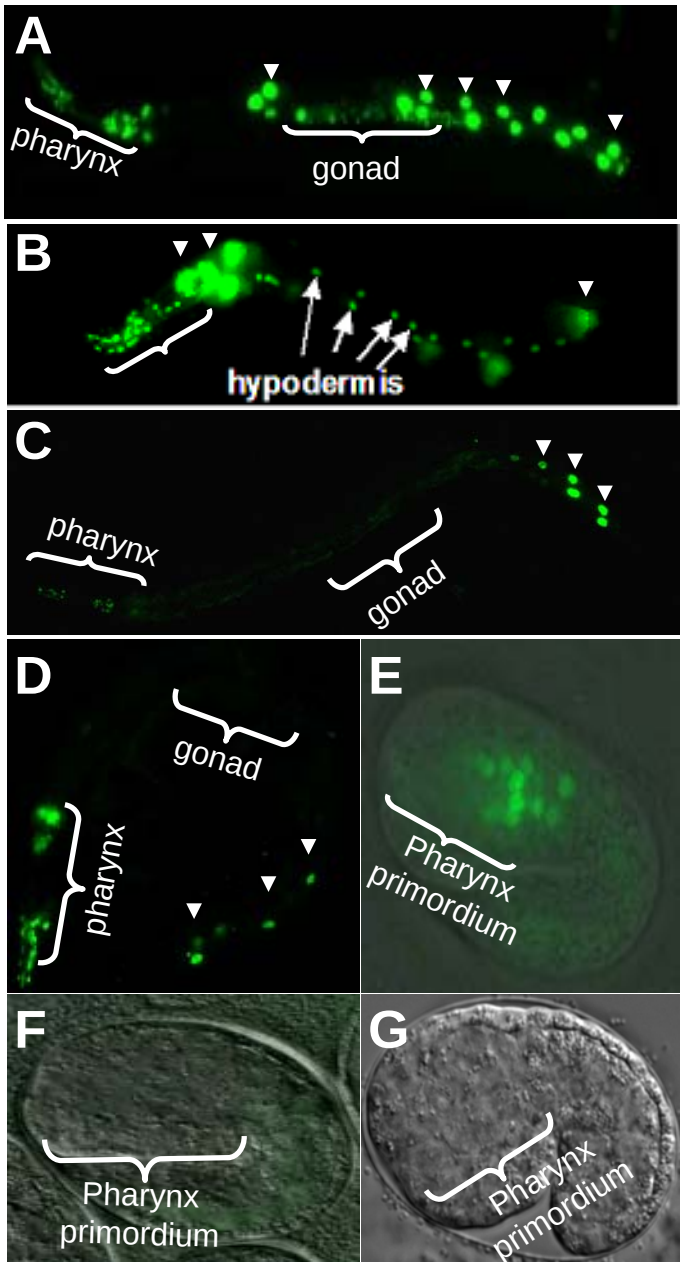


Figure 18. The expression patterns of the mutated PIG1-3 sites in the *M05B5.2* full-length and truncated promoters. A) Wildtype expression of the full-length *M05B5.2*::*GFP*::*HIS2B* reporter in the pharynx, intestine, hindgut and somatic gonad. B) Expression of mutM05, the full-length promoter with a mutation of the PIG-1 site, in the pharynx, intestine, hindgut, somatic gonad and hypodermis. C) The expression of mutPIG2, a reporter with a mutation of the PIG2 site in the full-length *M05B5.2* promoter, in the pharynx and weakly in the intestine. D) The expression of mut PIG3, a reporter with a mutation of the PIG3 site in the full-length *M05B5.2* promoter, in the pharynx and weakly in the intestine. E) Expression of 5`Δ270M05, the truncated *M05B5.2* reporter in embryos with expression in the pharynx and intestine. F) mutPIG2truncM05 expresses very weakly in the intestine. G) mutPIG3truncM05 does not express GFP in any tissues.

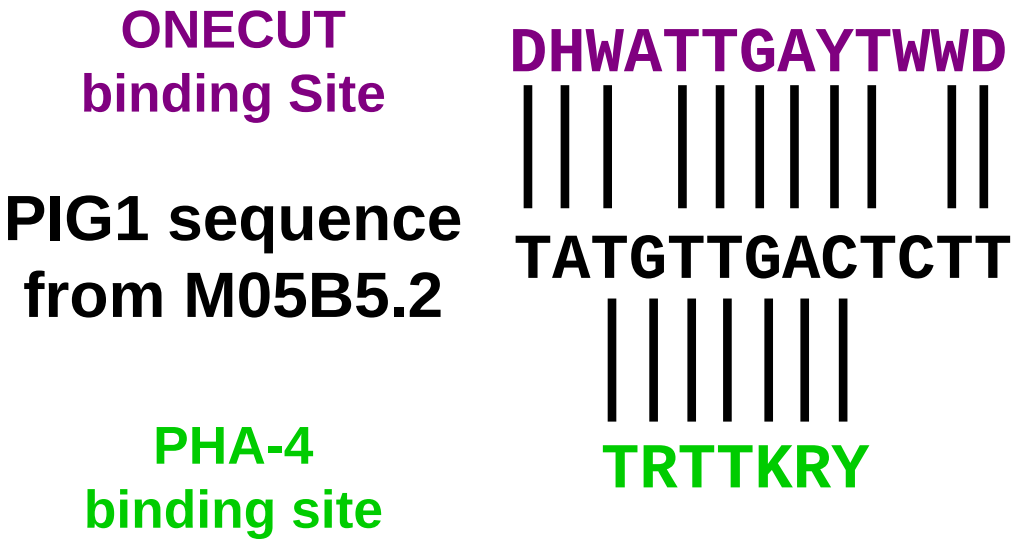


Figure 19. A comparison between the PIG1, PHA-4 site and HNF6/ONECUT binding sites. The PIG1 site contains a perfect match for a PHA-4 binding site and a very close match to a ONECUT binding site.

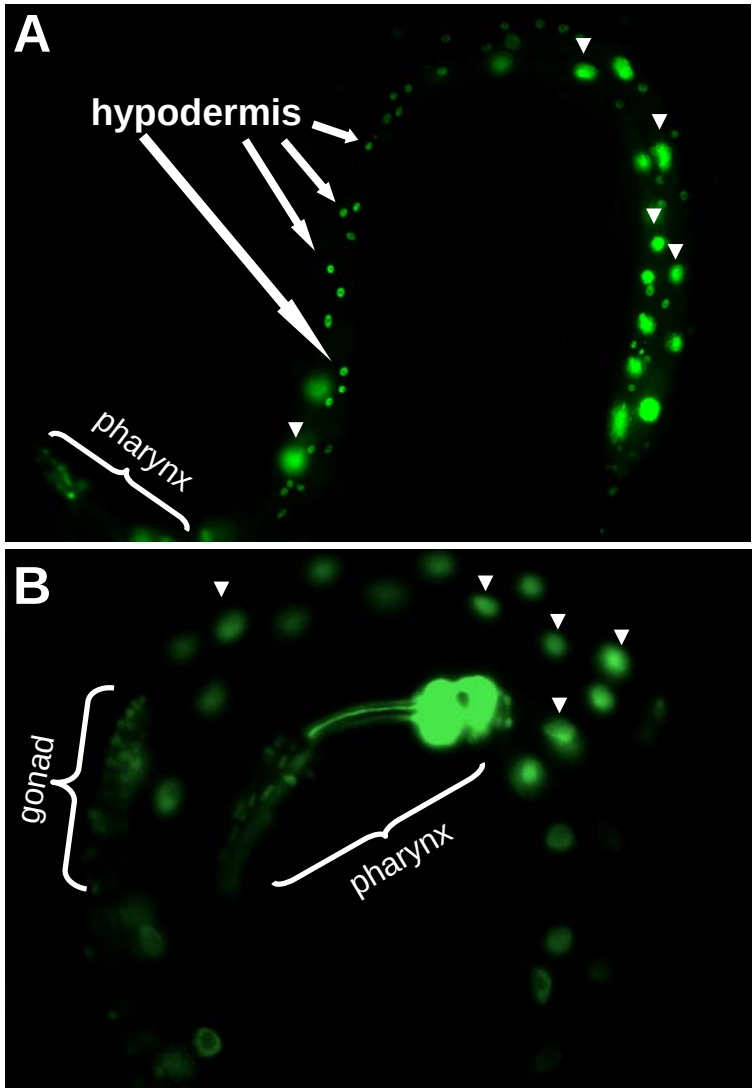


Figure 20. The expression of the *M05B5.2*::*GFP*::*HIS2B* reporter with a mutation in the PIG1-PHA-4 and PIG1-ONECUT-like sites. The arrowheads point to intestinal nuclei expressing GFP and the arrows point to hypodermal cells that are ectopically expressing GFP. A) The mutation of the ONECUT-like site that leaves the PHA-4 site intact (PIG1-ONECUT(-)) leads to ectopic hypodermal expression. B) The mutation of the PHA-4 site that leaves the ONECUT-like site intact (PIG1-PHA-4(-)) does not affect expression.

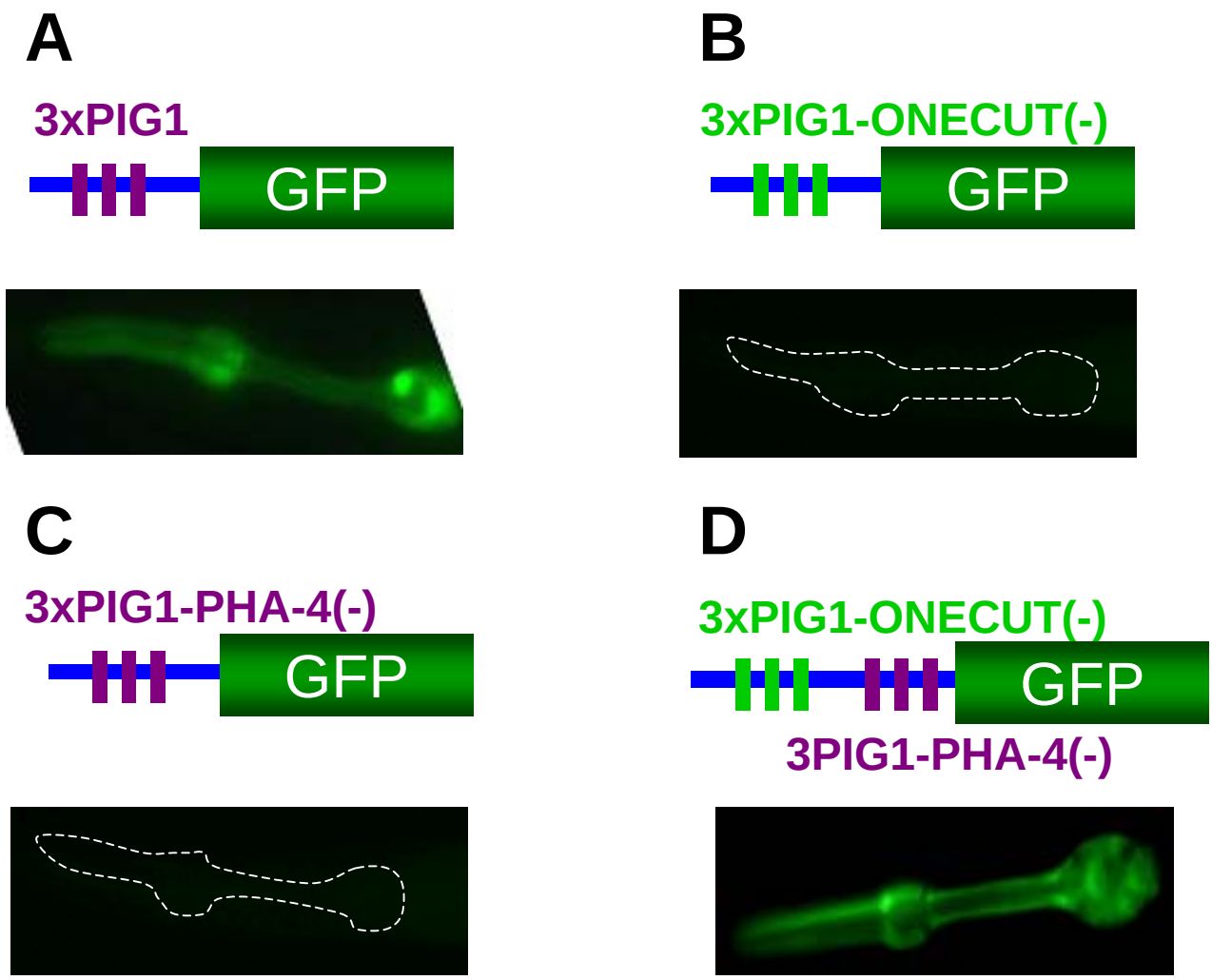


Figure 21. The expression of the PIG1 enhancer constructs. The cartoons depict the different enhancer constructs, with the GFP images showing the expression of the construct. A) Expression of the 3xPIG1 construct throughout the pharynx. B) The 3xPIG1-ONECUT(-) enhancer construct does not express in any tissues in *C. elegans.* C) The 3xPIG1-PHA-4(-) enhancer construct does not express in any tissues in *C. elegans.* D) The enhancer construct with a combination of PHA-4 and ONECUT-like sites (3xPIG1-ONECUT(-)::3xPIG1-PHA-4(-) is expressed throughout the pharynx.

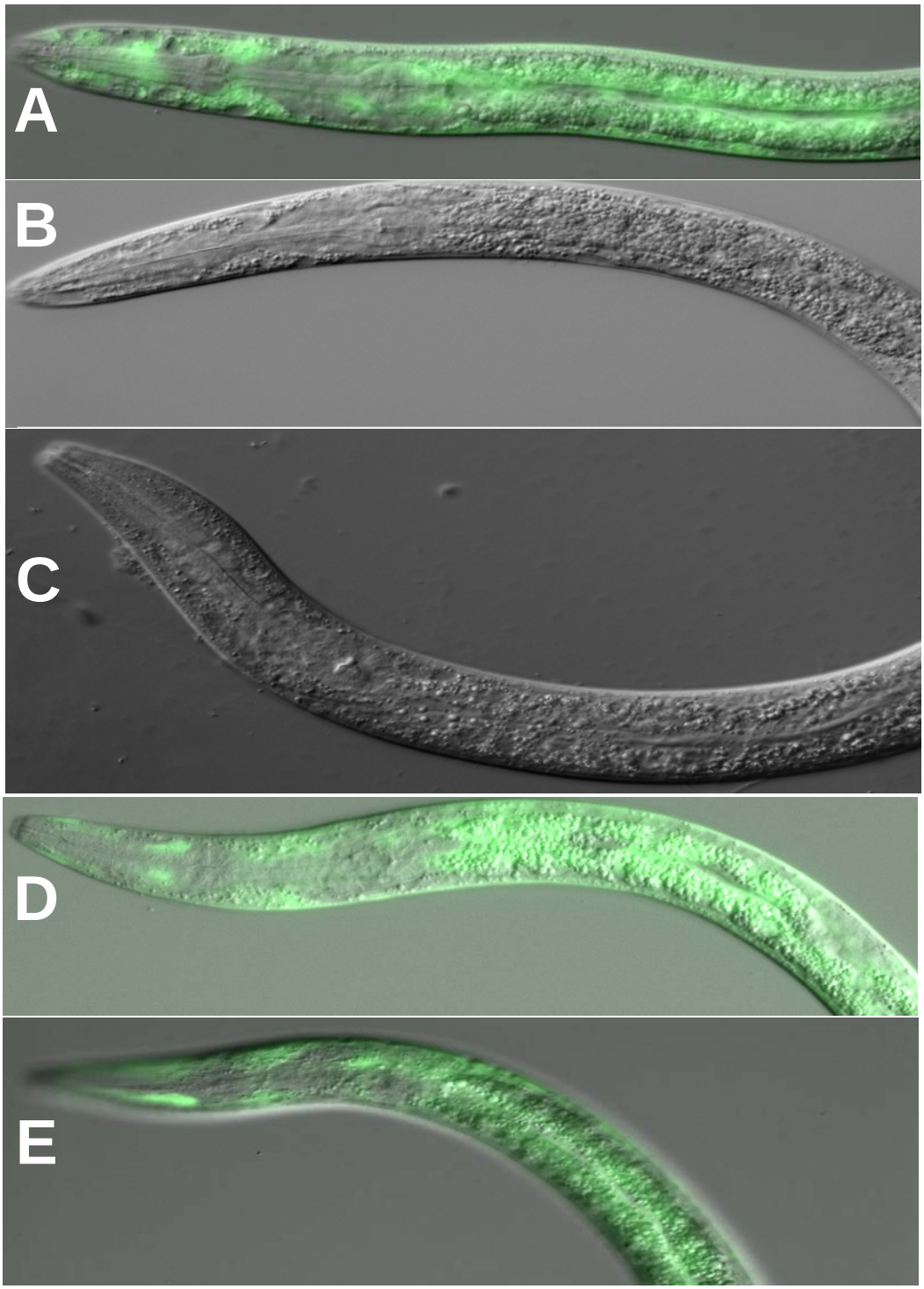


Figure 22. The expression of the *dpy-7* repressor constructs. A-D) These images are merged Nomarski Differential Interference Contrast and GFP images. A) Expression of the minimal *dpy-7*::GFP reporter; B) The PIG1::*dpy*-7::*GFP* repressor lacks hypodermal expression. C) The PIG-1-ONECUT(-)::PIG1-PHA-4(-)::*dpy*-7::*GFP* repressor lacks hypodermal expression. D) The PIG1-ONECUT(-)::*dpy-7*::*GFP* construct expresses in the hypodermis. E) The PIG1-PHA-4(-)::*dpy-7*::*GFP* construct expresses in the hypodermis.

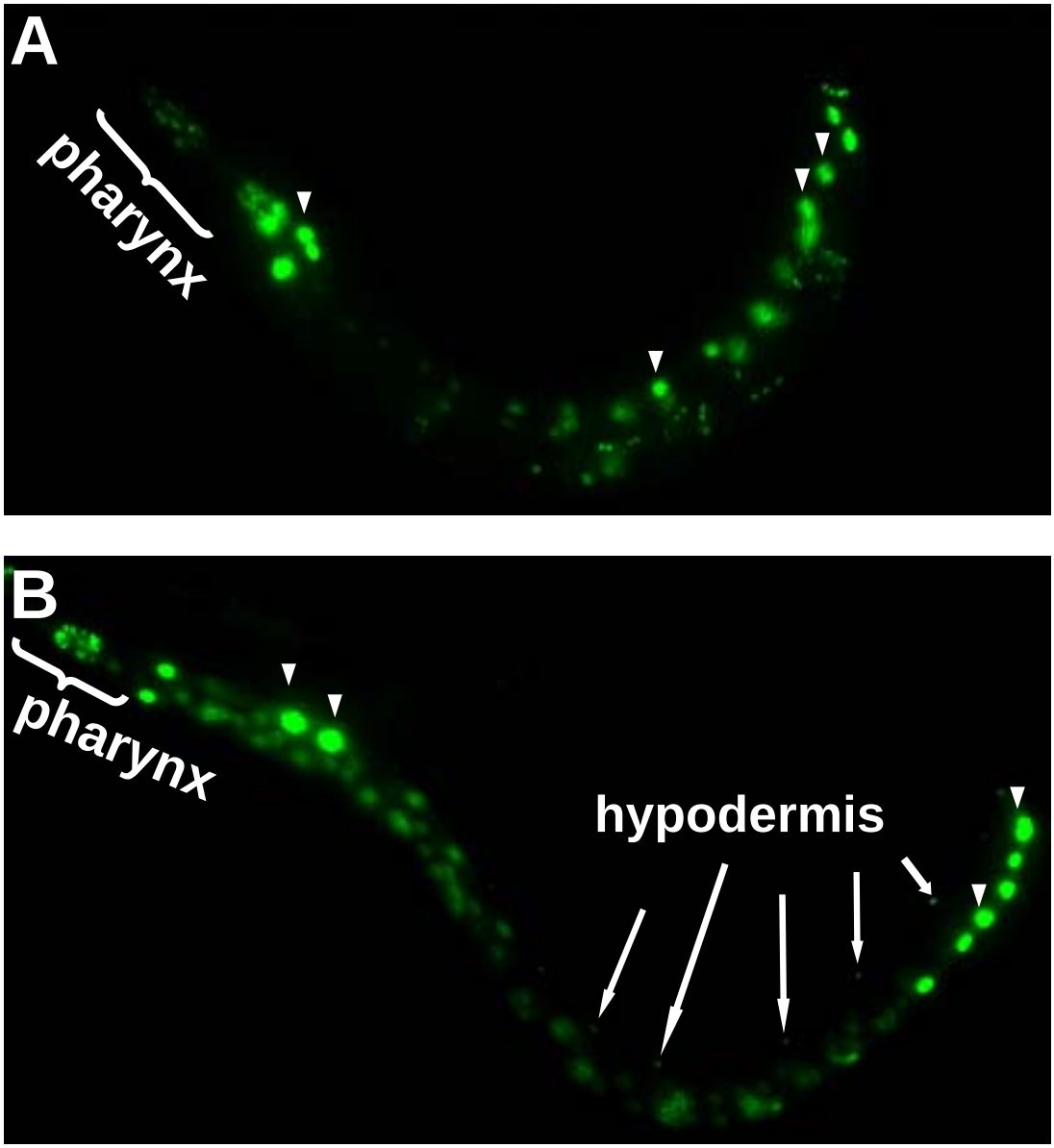


Figure 23. The expression of *M05B5.2*::*GFP*::HIS2B and mutM05 is not affected in *daf-4* mutant animals. A-B) The arrowheads highlight some of the expressing intestinal nuclei and the arrows highlight some of the expressing hypodermal nuclei. A) Expression of the *M05B5.2*::*GFP*::*HIS2B* reporter in *daf-4* mutant animals in the pharynx, intestine, hindgut and somatic gonad. B) Expression of mutM05 reporter in *daf-4* mutant animals in the pharynx, intestine, hindgut, somatic gonad and hypodermis.

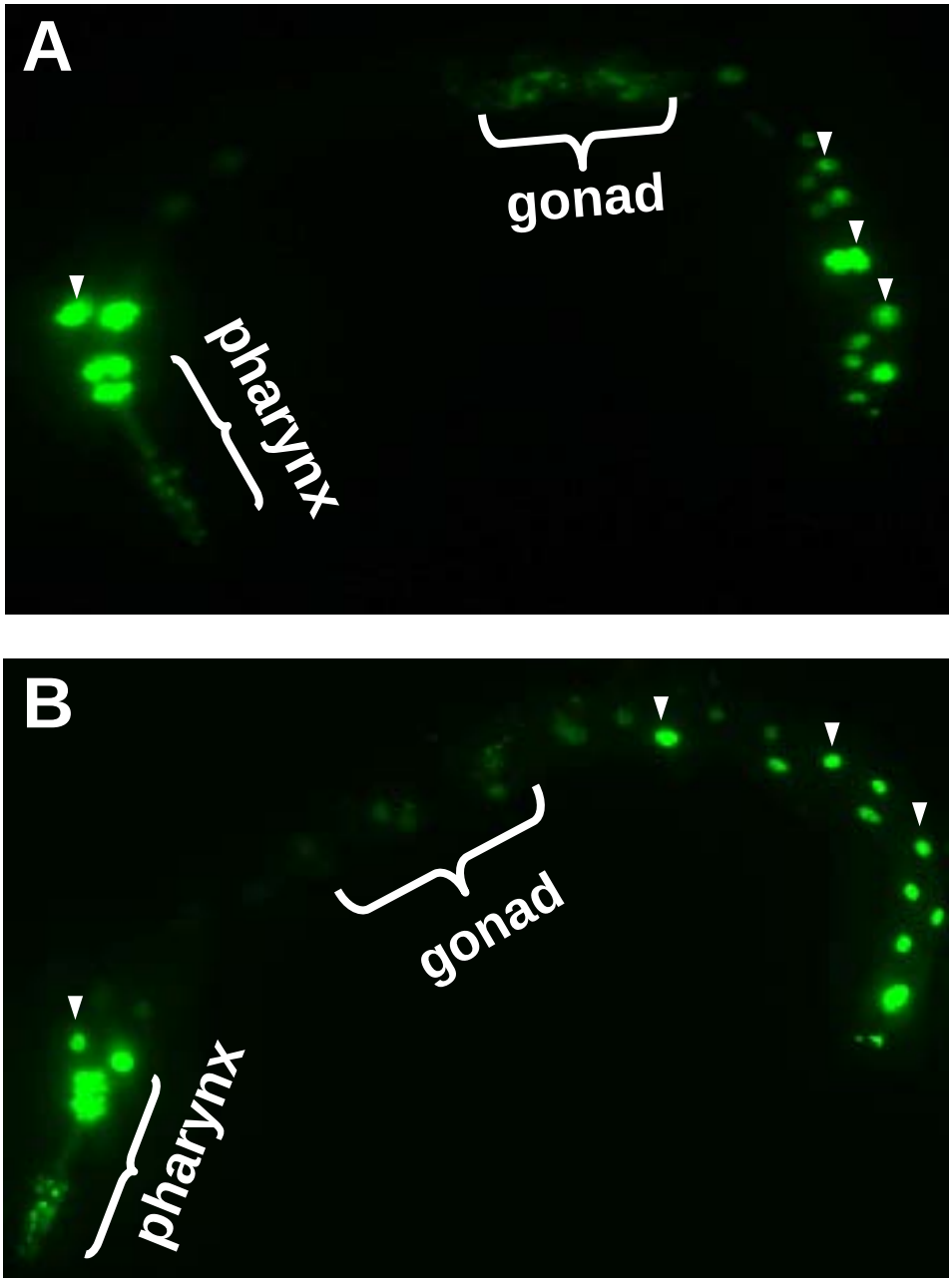


Figure 24. The expression of *M05B5.2*::*GFP*::*HIS2B* is not affected in *atf-2* mutant animals, or in ones injected with *atf-2* dsRNA. A-B) The arrowheads point to some of the expressing intestinal nuclei. A) Expression of the *M05B5.2*::*GFP*::*HIS2B* reporter in *atf-2* mutant animals in the pharynx, intestine hindgut and somatic gonad. B) Expression of *M05B5.2*::*GFP*::*HIS2B* reporter in animals injected with dsRNA from *atf-2* in the pharynx, intestine, hindgut and somatic gonad

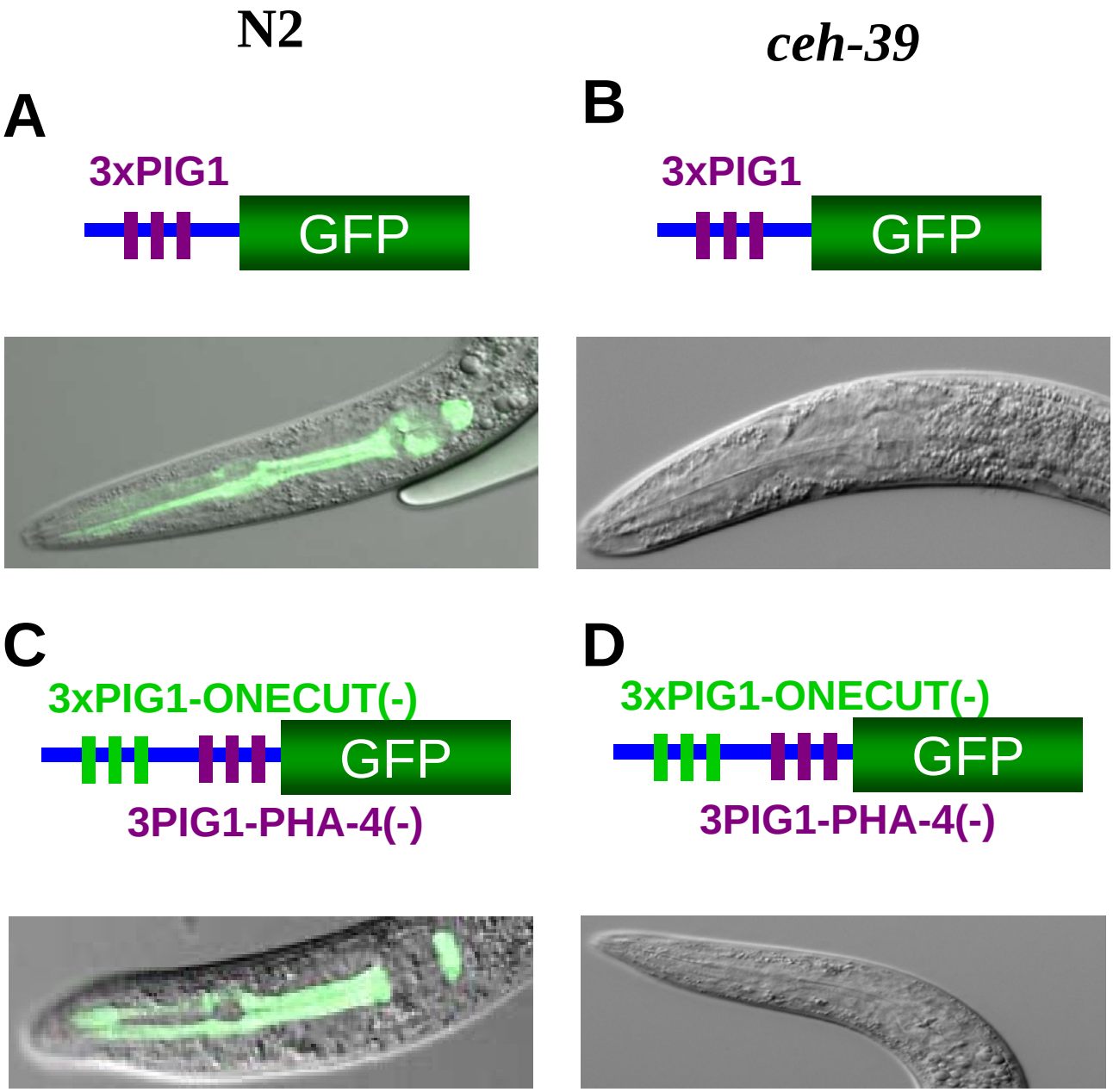


Figure 25. The expression of the 3xPIG1 and 3xPIG1-ONECUT(-)::3xPIG1-PHA-4(-) enhancer constructs in N2 and *ceh-39* animals. A) The pharyngeal expression of the 3xPIG1 construct in N2 animals. B) The 3xPIG1 enhancer construct does not express in the pharynx of *ceh-39* animals. C) The pharyngeal expression of the 3xPIG1-ONECUT(-)::3xPIG1-PHA-4(-) enhancer construct in N2 animals. D) 3xPIG1-ONECUT(-)::3xPIG1-PHA-4(-) enhancer construct does not express in the pharynx of *ceh-39* animals.

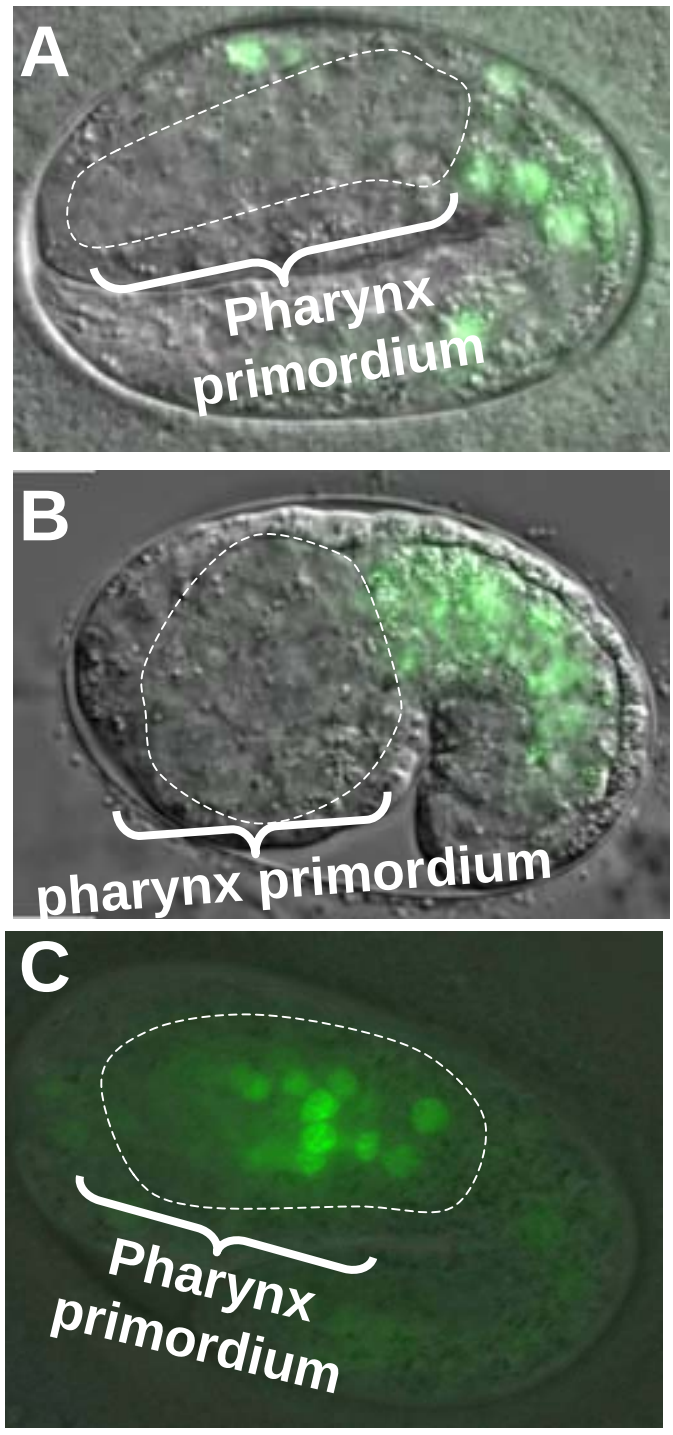


Figure 26. The expression of the 5`Δ270M05 reporter in *ceh-39* mutants and outcrossed animals. The pharyngeal membrane is highlighted by the dashed white lines. A-B) The 5`Δ270M05 reporter does not express in the primordial pharynx of *ceh-39(gk296)* mutant animals C) The 5`Δ270M05 reporter is expressed in the pharyngeal primordium of *ceh-39(gk296)* animals that were outcrossed with N2 to get wild type animals.

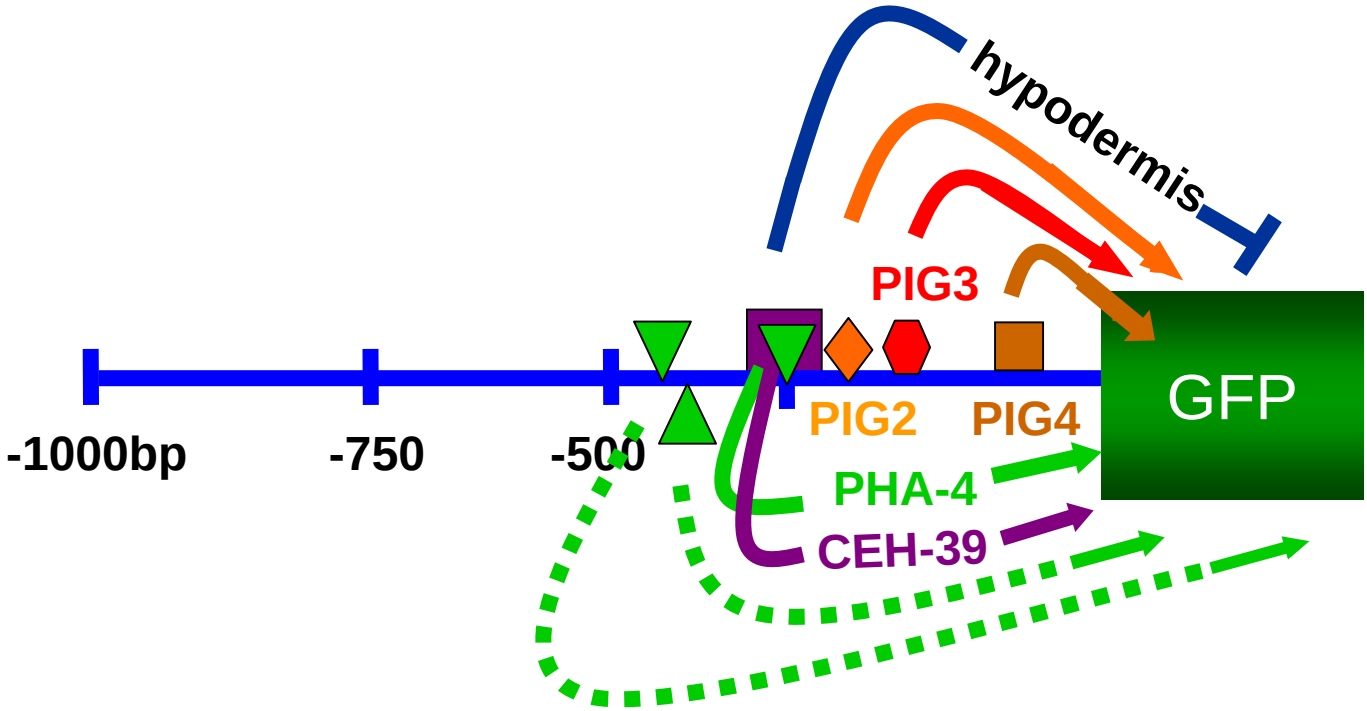


Figure 27. Transcriptional regulation of M05B5.2. The green triangles represent PHA-4 binding sites. It is not currently known whether PHA-4 and CEH-39 are acting directly or indirectly through these binding sites, however based on the current data it is likely that these factors are directly regulating the expression of M05B5.2.

Discussion

This study was undertaken to identify transcriptional regulators of pharyngeal development. The promoter of the pharyngeally-expressed gene *M05B5.2* was analyzed to identify pharyngeal *cis*-regulatory elements. Previous research by Gaudet and Mango (2002) had determined that the expression of *M05B5.2* was unusual in that it is not wholly dependent on PHA-4 and thus analysis of the regulation of this gene should identify other pharyngeal regulators.

Regulation of *M05B5.2* requires a combination of positive and negative signals that combine to drive expression in the pharynx, intestine, hindgut and somatic gonad while inhibiting hypodermal expression. The complex PIG1 *cis*-regulatory element contains at least two overlapping binding sites that are required for both activation and inhibition of *M05B5.2*. Thus, I have identified a novel pharyngeal regulator that activates pharyngeal expression of *M05B5.2*.

The Function of M05B5.2

Examination of the protein encoded by M05B5.2 using BLAST and PFAM searches was not able to identify homologous genes to *M05B5.2* outside of the *Caenorhabditis* genus. The other two *Caenorhabditis* species whose genomes have been sequenced and annotated, *C. briggsae and C. remanei,* both have orthologs of *M05B5.2*, suggesting that this gene is at least present within *Caenorhabditis* and may be present more broadly in other nematodes. Interestingly, the Cb-M05 reporter with 1 kb of promoter from the homologous *C. briggsae* gene CBG21925 also expresses in the digestive tract of *C. briggsae*, suggesting that the digestive tract expression of this gene is

likely also conserved in other species. Further analysis of the role of M05B5.2 in the digestive tract of *C. elegans* will also determine the role of M05B5.2 in the digestive tract of other nematodes.

The predicted product of M05B5.2 is probably a single-pass transmembrane protein, which suggests some possible mechanisms for its function. For example, this protein may act in cell signalling as a membrane-bound receptor or in cell-cell interactions possibly at cell junctions. Further analysis of M05B5.2 and its predicted protein domains will clarify its role in *C. elegans.* An *M05B5.2* mutant would allow me to test versions of the M05B5.2 protein that contain different predicted protein domains to determine which of these domains are critical. For example, to determine whether this protein is acting as a membrane receptor for cell signalling the small intracellular domain could be removed from a rescuing construct. If the intracellular portion of M05B5.2 is critical for rescue then this part of the protein may be required to induce a second messenger signal inside the cell, suggesting that the intracellular portion is required for activity. In contrast, if the intracellular portion is not required to rescue a mutation of *M05B5.2*, it is probable that this protein is acting only at the cell surface and could be involved in cell-cell interactions.

Identification of a mutant allele of *M05B5.2* would facilitate further analysis of the function of this protein. The *let-527* bearing strain was investigated as a probable mutant allele of *M05B5.2*, however the mutant gene in this strain was not conclusively identified as *M05B5.2* due to the contradictory rescue and sequencing results. It is possible though that the mutation in *let-527* could be in a region of *M05B5.2* that was not sequenced. The successful rescue of *let-527* with M05B5.2(+) suggests that this strain

contains a mutation in *M05B5.2*, however, without an identified molecular lesion this can not be conclusively confirmed. Thus, to further investigate a confirmed allele of *M05B5.2* with a known molecular lesion, a mutant allele of *M05B5.2* has been requested from the *C. elegans* gene knockout consortium.

Further investigation of the origin of the cells that express the pharyngeal *myo-2*::*GFP* and *pha-4*::*GFP* markers outside of the pharyngeal membrane after *M05B5.2(RNAi)* injection will also lead to a better understanding of the role of M05B5.2. The cells that inappropriately express these pharyngeal markers could be non-pharyngeal cells that are mis-fated or pharyngeal cells that are now in the wrong place. A lineage analysis of the cells that express the pharyngeal markers outside of the pharynx would determine whether these cells are mis-fated or now in the wrong place. As well, determination of the number of PHA-4 positive cells present in *M05B5.2(RNAi)* affected animals would determine whether cells outside the pharynx are mis-specified cells or pharyngeal cells that have not been incorporated into the organ.

Regulation of expression of *M05B5.2*

The regulation of expression of *M05B5.2* is surprisingly complex (Figure 27). It was originally hypothesized that *M05B5.2* would be solely regulated by PHA-4 as these two genes have identical expression patterns. However, the removal of all of the PHA-4 sites from the promoter of *M05B5.2* only decreases but does not abolish pharyngeal expression. PHA-4 is required for full expression of *M05B5.2*; however, there are also four other sites in the promoter of *M05B5.2* that are critical for normal expression, PIG1-4. The PIG1 site contains the elements required for activation in the pharynx and hindgut

and inhibition in the hypodermis. The other three sites, PIG2-4, are required for expression in all tissues in which *M05B5.2* is expressed, the pharynx, intestine, hindgut and somatic gonad. The PIG2 site is possibly a bZIP binding site, based on the identical match of this site to the AP-1 bZIP binding site (Patel, 1999), the PIG3 site is a close match to a general activating site "Early-1" (Gaudet et al., 2004), while the PIG4 site does not show homology to any known transcription factor binding sites.

PIG1 is a complex regulatory element that contains overlapping binding sites

PIG1 is only 13 base pairs, yet contains sites for both positive and negative regulation of *M05B5.2*. The PIG1 site contains a predicted PHA-4 binding site that has been previously shown to bind PHA-4 *in vitro* and is required to activate pharyngeal expression in enhancer assays (Gaudet and Mango, 2002). The PIG1 site also contains an overlapping ONECUT-like binding site that is required for pharyngeal activation. The HNF6/ONECUT transcription factor CEH-39 is required for the activation of the PIG1 enhancers and *M05B5.2* reporters; CEH-39 is likely acting directly through this site to activate pharyngeal expression.

PIG1 is also required for hypodermal inhibition of *M05B5.2*. The *dpy-7* repressor assay suggests that two transcription factors, an HNF6/ONECUT and a forkhead factor are required for inhibition by this site. Both the PHA-4 and ONECUT-like binding sites were required to inhibit hypodermal expression of *dpy-7*. In contrast, the mutation of only the ONECUT-like site and not the PHA-4 site turned on ectopic expression in the hypodermis. It is likely that the factor binding to the PHA-4 site in the hypodermis is not PHA-4, as this factor is not expressed in the hypodermis; a different forkhead factor may

act on this site in the hypodermis to inhibit hypodermal expression. It is possible that the specific mutation of the PHA-4 site did not significantly affect the binding of the relevant hypodermal factor, and was thus still able to repress the hypodermal expression even with the mutation of the PHA-4 site.

Identification of the hypodermal inhibitory factors has been unsuccessful. Investigation of the forkhead and ONECUT transcription factors that are expressed in the hypodermis of *C. elegans* has not identified inhibitory factors. It is possible that forkhead or ONECUT factors that we excluded as candidates may in fact have some hypodermal activity, thus it is still possible that forkhead and ONECUT transcription factors are required for hypodermal inhibition. It is also possible that two different factors (neither forkhead nor ONECUT) that bind to similar sequences may be required for inhibition in the hypodermis. To determine the identity of the factors that inhibit hypodermal expression, RNAi or mutant strain analysis of the rest of the forkhead and ONECUT transcription factors could be done. Another method for identifying the inhibitory factors is to perform a mutagenesis screen on a strain of worms containing the integrated *M05B5.2*::*GFP*::*HIS2B* reporter. The investigation of animals that ectopically express the integrated M05B5.2 reporter in the hypodermis after mutagenesis would identify the factor(s) involved in normal hypodermal inhibition. As well, an RNAi screen of transcription factors could be performed to look for genes that when knocked out lead to ectopic hypodermal expression of the integrated *M05B5.2*::*GFP*::*HIS2B* reporter.

Other examples of overlapping binding sites

The PIG1 regulatory element in the promoter of *M05B5.2* is not the only example of a single regulatory element at which multiple factors bind to modulate expression. Interestingly, other examples of overlapping binding sites in the promoters of other *C. elegans* genes also occur at PHA-4 binding sites. For example, one of the PHA-4 sites in the promoter of the pharyngeal gene *myo-2* also contains an overlapping PEB-1 binding site (Thatcher et al., 2001). The PEB-1 pharyngeal transcription factor was originally identified in analysis of the *myo-2* promoter. The binding site for this factor was found to overlap the binding site for PHA-4, similar to the overlap of sites in PIG1 (Kalb et al., 1998; Thatcher et al., 2001). PEB-1 has been found to inhibit *in vitro* binding of PHA-4, suggesting that the overlap of these sites may be necessary for modulation of PHA-4 activity by PEB-1 inhibition (Thatcher et al., 2001). However, the interaction between PEB-1 and PHA-4 *in vivo* remains unclear as does the pharyngeal function of PEB-1.

As well, the Gaudet lab has identified at least two other PHA-4 sites whose expression is likely modified by a second factor whose binding site overlaps that of PHA-4 (personal communication). Different PHA-4 enhancer constructs show different *in vivo* expression patterns depending on the sequence of the enhancer suggesting that there is likely more than simple PHA-4 regulation of expression. Some of the enhancer constructs with multiple copies of PHA-4 sites express at very low levels in the pharynx, whereas others are expressed strongly and often in a specific subset of pharyngeal cells. It is possible that these PHA-4 enhancers may have different affinities for PHA-4 binding, which leads to the varied expression of PHA-4 enhancers. However, this is unlikely since the different enhancers do not simply express at different levels, but also express in

different cell types. Another explanation for the different activity of different PHA-4 enhancers (including PIG1) is that they may not be PHA-4 sites at all, but each may be a binding site for a different transcription factor, which drives the cell-type specific expression pattern. Finally a third possibility is that these sites are bound by PHA-4 and also by a second factor that modulates expression to drive the tissue specific expression pattern. My data suggests, that at least for PIG1, two factors that bind the same regulatory site are required for pharyngeal expression.

The PIG1 site can also be broken down into two different binding sites, a PHA-4 site and an ONECUT-like site. Neither of these sites drive expression in the pharynx alone; however, placing these sites in tandem in the same enhancer recapitulates the expression of the whole PIG1 site. The varied expression of the different PHA-4 sites and the ability to break down the PIG1 site into two different binding sites suggest that the model of regulation most likely involves two factors binding to the PHA-4 site. It is likely that PHA-4 binds to activate expression in the pharynx while the second factor binds to modulate expression.

Previous work identifying the PHA-4 homolog HNF3 as a “Pioneer” factor may explain the two-factor regulation through these PHA-4 binding sites. It has been proposed that HNF3 acts as a “pioneer” factor by de-compacting local chromatin making the DNA accessible to other proteins (Cirillo et al., 2002; Holmqvist et al., 2005). It is possible that PHA-4 acts similarly to HNF3 by binding to the promoter of different genes and opening the chromatin to give access to additional factors like CEH-39 instead of acting as a general pharyngeal activator. As well, the number of PHA-4 enhancers that have different

expression patterns suggest that this modulation of expression by a second factor is likely a common phenomenon in pharyngeal regulation.

ceh-39 encodes a ONECUT factor that may directly regulate PIG1

In a *ceh-39* mutant background, pharyngeal expression of both the 3xPIG1 enhancer constructs and a truncated *M05B5.2* reporter, 5`Δ270M05, are significantly reduced. However, the two different mutant alleles of *ceh-39* have different effects on the expression of these reporters. One allele, *ceh-39(gk296)*, completely abolishes the expression at all developmental stages, while the second, *ceh-39(gk329)*, abolishes only late larval and adult expression leaving weak embryonic and young larval expression in the pharynx of a small percentage of animals. One possible explanation for this difference is that *ceh-39(gk329)* may be a weaker allele than *ceh-39(gk296)*. This possibility seems unlikely given that both alleles are deletions within *ceh-39* that remove the predicted start codon. However, it remains possible that *ceh-39(gk329)* retains some activity by use of an alternate ATG; such a product would include the predicted homeobox of CEH-39, which is unaffected by the *gk329* deletion. Further examination of these two strains using reverse transcriptase PCR will determine whether either of these two alleles produce messages that may be functional. As well, injection of *ceh-39* dsRNA into the strain carrying the *ceh-39(gk329)* allele and the 3xPIG1 enhancer construct would determine whether further knockdown of *ceh-39* abolishes the early expression of this reporter. A loss of the early pharyngeal expression due to the injection of *ceh-39* dsRNA into the *ceh-39(gk329)* allele would suggest that this allele is making a functional protein product that retains some activity.

ceh-39 is required for expression of both artificial enhancers containing multiple copies of PIG1 and a truncated M05B5.2 reporter construct, 5`Δ270M05, that contains the PIG1 element. The PIG1 regulatory element contains both ONECUT-like and PHA-4 binding sites. *ceh-39* is predicted to encode a HNF6/ONECUT transcription factor and since the PIG1 site contains a ONECUT-like binding site the regulation of M05B5.2 by CEH-39 is most likely direct. An *in vitro* or *in vivo* binding assay looking at whether CEH-39 is able to bind the PIG1 site would determine whether this regulation is direct or indirect. Further binding experiments with both PHA-4 and CEH-39 would determine whether these two proteins bind the PIG1 site independently, cooperatively or competitively. In this binding assay if the presence of PHA-4 does not affect the binding of CEH-39 or vice versa then these two factors are binding independently. Alternatively, the binding of CEH-39 may enhance the binding of PHA-4 or vice versa suggesting that these two factors are binding cooperatively. A third possibility is that the binding of CEH-39 inhibits the binding of PHA-4 or vice versa, which would suggest that these two factors are binding competitively.

As well, further examination of *ceh-39* will help identify the role of this transcription factor in pharyngeal development. The two deletion alleles of *ceh-39* are both viable suggesting that this factor is not critical for pharyngeal development. However, the loss of pharyngeal expression of the enhancer and promoter constructs indicates that CEH-39 is at least capable of activating pharyngeal expression. CEH-39 may play a similar role as PHA-4 or other pharyngeal regulators in the activation of pharyngeal genes. The loss of both PHA-4 and CEH-39 may have a more severe effect than the loss of either factor alone. The loss of PHA-4 leads to lethality in *C. elegans*

making it difficult to look for functional interactions with PHA-4 and thus, an assay was previously designed to look for these functional interactions by knocking out a second gene (by *tam-1(RNAi)*) in a background containing a weak allele of PHA-4 (Kaltenbach, Updike and Mango, 2005; Keifer, Smith and Mango, 2006). Therefore, testing the loss of CEH-39 in a strain of *C. elegans* that contains a weak allele of PHA-4 will determine if these two factors functionally interact. If the loss of CEH-39 in an already compromised weak PHA-4 background causes further pharyngeal defects or lethality then it would support a role for CEH-39 in pharyngeal development or function.

Sufficiency of the PIG elements only accounts for pharyngeal expression

Of the four regulatory sites within the M05B5.2 promoter, only PIG1 is sufficient to drive expression. Interestingly, the PIG1 site is only sufficient to drive expression in the pharynx and hindgut, but not the intestine or somatic gonad. The individual or combinatorial enhancer constructs of the other regulatory elements, PIG2-4 have not identified the elements that are sufficient to drive expression of M05B5.2 in tissues other than the pharynx and hindgut. It remains unclear as to what is driving the expression of M05B5.2 in the intestine and somatic gonad. PHA-4 has been shown to be present in the intestine and somatic gonad of *C. elegans*, however none of the PHA-4 enhancer constructs express in these tissues. It is possible that an enhancer containing the PIG-1 site one or more of the PIG2-4 sites could activate expression in the intestine and somatic gonad. Thus, further experiments with enhancers containing PIG1 and one or more of the other three regulatory sites will determine whether these sites can be sufficient to activate intestinal or gonadal expression.

Further investigation of the promoter may also identify the sites that are sufficient to activate expression in the intestine or somatic gonad. It has been previously shown that the intestinal transcription factor *elt-2* is necessary for intestinal gene expression and probably accounts for the regulation of the majority of intestinal genes (Fukushige et al., 2005; McGhee et al., 2007). As well, this factor has been shown to bind through a specific WGATAR site (GATA site) in the promoter of intestinal genes (Fukushige et al., 2005). I have shown that of one of the GATA sites in the M05B5.2 promoter is not necessary for expression, however there are other predicted GATA sites present in the promoter of M05B5.2 that may act to drive intestinal expression of this gene. Tests of the GATA sites may determine the sufficient element that drives M05B5.2 expression in the intestine.

Conservation of regulation of foregut genes

The regulatory factors and candidate factors identified in this study suggest that components of the transcriptional network in the foregut of *C. elegans* is conserved in other organisms. It is likely that a bZIP factor is binding through the PIG2 regulatory site, as it is identical to the bZIP binding site for AP-1 (Patel, 1999). This suggests conservation with Drosophila and mammalian foregut development as both the Drosophila *cap'n'collar (cnc)* and mammalian C/EBPα bZIP factors are required for normal foregut development and gene expression (Mohler et al., 1995; Suzuki et al., 2003).

The development of mammalian and zebrafish foregut organs require both the PHA-4 homolog HNF3 (FoxA2) and the CEH-39 homolog HNF6 (Cheng et al., 2006;

Lemaigre and Zaret, 2004; Schrem, Klempnauer and Borlak, 2002). The loss of either of these factors results in developmental defects in the liver and pancreas (Gannon et al., 2003, Schrem, Klempnauer and Borlak, 2002). The loss of HNF6 in the liver and pancreas is different from the loss of CEH-39 in *C.elegans. ceh-39* mutants do not have foregut defects, possibly because CEH-39 is redundant with PHA-4 or other transcription factors so that the loss of this protein on its own is not detrimental to *C. elegans* . Gene expression in mammalian and zebrafish foregut organs requires both HNF3 (FoxA) and HNF6 activation (Cereghini, 1996; Cheng et al., 2006; Odom et al., 2004) and HNF3 and HNF6 often bind to the same promoters of foregut genes (Cheng et al., 2006; Rausa, Tan and Costa, 2003). Interestingly, HNF3 and HNF6 have been found to physically interact to regulate mammalian foregut gene expression (Rausa, Tan and Costa, 2003; Rubins et al., 2005), suggesting that there may be a similar physical interaction between PHA-4 and CEH-39. The identification of the novel pharyngeal regulator, CEH-39, will lead to a better understanding of foregut development in *C. elegans* as well as other organisms.

References

Albertson, D. G., and Thomson, J. N. (1976). The pharynx of *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society of London 275,* 299-325.

Altun, Z.F. and Hall, D.H. (2005). The Pharynx. In Worm Atlas. http://www.wormatlas.org/handbook/alimentary/alimentary1.htm

Avery, L. (1993). The genetics of feeding in *Caenorhabditis elegans. Genetics, 133,* 897-917.

Azzaria, M., Goszczynski, B., Chung, M.A., Kalb, J.M., McGhee, J.D. (1996). A *fork head*/HNF-3 homolog expressed in the pharynx and intestine of the *Caenorhabditis elegans* embryo. *Developmental Biology, 178,* 289-303.

Bendtsen, J.D., Nielsen, H., von Heijne, G. and Brunak, S. (2004). Improved prediction of signal peptides: SignalIP 3.0. *Journal of Molecular Biology, 340,* 783-795.

Blackwood, E.M., Kadonga, J. T. (1988). Going the distance: Current view of enhancer action. *Science, 281,* 61-63.

Breathnac, R. and Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. *Annual Review of Biochemistry, 50,* 349-383.

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

Cassata, R., Kagoshima, H., Pretot, R.F., Aspock, G., Niklaus, G. and Burglin, T.R. (1998). Rapid expression screening of *Caenorhabditis elegans* homeobox open reading frames using a two-step polymerase chain reaction promoter-gfp reporter construction technique. *Gene, 212,* 127-135.

Cereghini, S. (1996). Liver-enriched transcription factors and hepatocyte differentiation. *The Journal of the Federation of American Societies for Experimental Biology, 10,* 267-282.

Chandler, V.L., Maler, B.A. and Yamamoto, K.R. (1983). DNA sequences bound specifically by glucocorticoid receptor in vitro render a heterologous promoter hormone responsive in vivo. *Cell, 33,* 489-499.

Cheng, W., Guo, L., Zhang, Z., Soo, H.M., Wen, C. et al. (2006). HNF factors form a network to regulate liver-enriched genes in zebrafish. *Develpomental Biology, 294,* 482-496.

Chowdhuri, R.S., Crum, T., Woollard, A., Aslam, S. and Okkema, P.G. (2006). The T-box factor TBX-2 and the SUMO conjugating enzyme UBC-9 are required for ABa-derived pharyngeal muscle in *C. elegans. Developmental Biology, 295,* 664-677.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M. and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell, 9,* 279-289.

Clark, K.L., Halay, E.D., Lai, E. and Burley, S.K. (1993). Cocrystal structure of the HNF-3/*fork head* DNA-recognition motif resembles histone H5. *Nature, 364,* 412-420.

Davis-Silberman, N. and Ashery-Padan, R. (2007). Iris development in vertebrates; genetic and molecular considerations. *Brain Research, Epub.*

Fernandez, A.P., Gibbons, J. and Okkema, P.G. (2004). *C. elegans peb-1* mutants exhibit pleiotropic defects in molting, feeding and morphology. *Developmental Biology, 276,* 352-366.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.A. and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans. Nature, 391,* 806-811.

Fraser, A.G., Kamath R.S., Zipperlen, P., Martinez-Campos, M., Sohrmmann, M. and Ahringer, J.A. (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA intereference. *Nature, 408,* 325-330.

Fukushige, T., Goszczynski, B., Yan, J., and McGhee, J.D. (2005). Transcriptional control and patterning of the pho-1 gene, an essential acid phosphatase expressed in the *C. elegans* intestine. *Developmental Biology, 279,* 446-461.

Fukushige, T., Hendzel, M.J., Bazett-Jones, D.P. and McGhee, J.D. (1999). Direct visualization of the elt-2 gut-specific GATA factor binding to a target promoter inside the living Caenorhabditis elegans embryo. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 11883-11888.

Gannon, M., Ray, M.K., Van Zee, K., Rausa, F., Costa, R.H. and Wright, C. V. (2000). Persistent expression of HNF6 in islet endocrine cells causes disrupted islet architecture and loss of beta cell function. *Development, 127,* 2883-2895.

Garcia, J.V., Bich-Thuy, L.T., Staford, J. and Queen, C. (1986). Synergism between immunoglobulin enhancers and promoters. *Nature, 322,* 383-385.

Gaudet, J. and Mango, S.E. (2002). Regulation of organogenesis by the *Caenorhabditis elegan*s FoxA protein PHA-4. *Science, 295*, 821-825.

Gaudet, J., Muttumu, S., Horner, M. and Mango, S.E. (2004). Whole-genome analysis of temporal gene expression during foregut development. *Public Library of Science, 2,* 1828-1842.

Gilleard, J.S., Barry, J.D. and Johnstone, I.L. (1997) *cis* regulatory requirements for hypodermal cell-specific expression of the *Caenorhabditis elegans* cuticle collagen gene *dpy-7. Molecular and Cellular Biology, 17,* 2301-2311.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C. and Gelbart, W.M. (1999). An introduction to genetic analysis. New York: W. H. Freeman, 350-356.

GuhaThakurta, D., Palomar, L., Stormo, G.D., Tedesco, P., Johnson, T.E., Walker, D.W., Lithgow, G. and Kim, S. (2002). Identification of a novel cis-regulatory element involved in the heat shock response in *Caenorhabditis elegans* using microarray gene expression and computational methods. *Genome Research, 12,* 701-712.

Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G. and Gibson T.J.(1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Re*search, *22,* 4673-4680.

Ho, S.N., Hunt, H.D., Horton, R.M., Pallen, J.K. and Pease, L.R. (1989). Ste-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene, 77,* 51-59.

Holmqvist, P.H., Belikov, S., Zaret, K.S. and Wrange, O. (2005). FoxA1 binding to the MMTV LTR modulates chromatin structure and transcription. *Experimental Cell Research, 304,* 593-603.

Hope, I.A., Mounsey, A., Bauer, P. and Aslam, S. (2003). The forkhead gene family of *Caenorhabditis elegans*. *Gene, 304,* 43-55.

Horner, M.A., Quintin, S., Domeier, M.E., Kimble, J., Labouesse, M., et al. (1998). *pha-4* an HNF-3 homologue, specifies organ identity in *Caenorhabditis elegans. Genes and Development, 12,* 1947-1952.

Horvitz, H.R., Brenner, S., Hogkin, J. and Herman, R.K. (1979). A uniform genetic nomenclature for the nematode *Caenorhabditis elegans. Molecular and General Genetics, 175,* 129-133.

Jin, Y. (1999). Transformation, In *C. elegans* a practical approach, I. A. Hope, eds. New York: Oxford University Press, pp. 69-88.

Kalb, J.M., Beaster-Jones, L., Fernandez, A.P., Okkema, P.G., Goszczynski, B and McGhee, J.D. (2002). Interference between the PHA-4 and PEB-1 transcription factors in formation of the *Caenorhabditis elegans* pharynx. *Journal of Molecular Biology, 320,* 697-704.

Kaltenbach, L.S., Updike, D.L. and Mango, S.E. (2005). Contribution of the amino and carboxyl termini for PHA-4/FoxA function in *Caenorhabditis elegans*. *Developmental Dynamics, 234,* 346-354.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature 421,* 231–237.

Keifer, J.C., Smith, P.A. and Mango S.E. (2006). PHA-4/FoxA cooperates with TAM-1/TRIM to regulate cell fate restriction in the *C. elegans* foregut. *Developmental Biology, 302,* 611-624.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology, 305,* 567-580.

Kume, S. (2005). The molecular basis and prospects in pancreatic development. *Development, Growth and Differentiation, 47,* 367-374.

Kwon, J.Y., Hong, M., Choi, M.S., Kang, S., Duke, K., Kim, S., Lee, S. and Lee, J. (2004). Ethanol-response genes and their regulation analyzed by a microarray and comparative genomic approach in the nematode *Caenorhabditis elegans*. *Genomics, 83,* 600-614.

Lavon, N. and Benvenisty, N. (2005). Study of hepatocyte differentiation using embryonic stem cells. *Journal of Cellular Biochemistry, 15,* 1193-1202.

Lemaigre, F. and Zaret, S. (2004). Liver development update: new embryonic models, cell lineage control and morphogenesis. *Current Opinion in Genetics & Development, 14,* 582-590.

Mango, S.E., Lambie, E.J., Kimble, J. (1994). The *pha-4* gene is required to generate the pharyngeal primordium of *Caenorhabditis elegans. Development, 120,* 3019-3031.

Massague, J. and Chen, Y.G. (2000). Controlling TGF-β signaling. *Genes & Development, 14,* 627-644.

McDowall, J.S. and Rose, A.M. (1997). Genetic analysis of sterile mutants in the dpy-5 unc-13 (I) genomic region of *Caenorhabditis elegans. Molecular and General Genetics, 255,* 60-77.

McGhee, J.D., Sleumer, M.C., Bilenky, M.C., Wong, K., McKay, S.J., et al (2007). The ELT-2 GATA-factor and the global regulation of transcription in the *C. elegans* intestine. *Developmental Biology, 302,* 627-645.

McKim, K.S., Starr, T.V.B. and Rose, A.M. (1992). Genetic and molecular analysis of the *dpy-14* region in*Caenorhabditis elegans. Molecular and General Genetics, 233,* 241-251.

Mello, C.C., Kramer, J.M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequence. *The European Molecular Biology Organization Journal, 10,* 3959-3970.

Mohler, J., Mahhaffey, J.W., Deutsch, E. and Vani, K. (1995). Control of Drosophila head segment identity by the bZIP homeotic gene *cnc. Development, 121,* 237-242.

Morck, C., Rauthan, M., Wagberg, F. and Pilon, M. (2004). *pha-2* encodes the *C. elegans* ortholog of the homeodomain protein HEX and is required for the formation of the pharyngeal isthmus. *Developmental Biology, 272,* 403-418.

Muller, M., Gerster, T. and Schaffner, W. (1988). Enhancer sequences and the regulation of gene transcription. *European Journal of Biochemistry, 176,* 485-495.

Murtaugh, L.C. and Melton, D.A. (2003). Genes, signals, and lineages in pancreas development. *Annual Review of Cell and Developmental Biology, 19,* 71-89.

Pelham, H.R. (1982). A regulatory upstream promoter element in *Drosophila* hsp 70 heat-shock gene. *Cell, 30*, 517-528.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors, *Science, 303,* 1378-1381.

Okkema, P.G. and Fire, A. (1994). The Caenorhabditis elegans NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle. *Development, 120*, 2175-2186.

Okkema, P.G., Ha, E., Haun, C., Chen, W. and Fire, A. (1997). The *Caenorhabditis elegans* NK-2 homeobox gene *ceh-22* activates pharyngeal muscle gene expression in combination with *pha-1* and is required for normal pharyngeal development. *Development, 124,* 3965-73.

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

Overdier, D.G., Porcella, A and Costa, R.H. (1994). The DNA-binding specificity of the hepatocyte nuclear factor 3/*forkhead* domain is influenced by amino acid residues adjacent to the recognition helix. *Molecular and Cellular Biology, 14,* 2755-2766.

Panowski, S.H., Wolff, S., Aguilaniu, H., Durieux, J. and Dillin, A. (2007). PHA-4/FoxA mediates diet-restriction-induced longevity of *C. elegans. Nature, 447,* 536-537.

Patel, L., Abate, C. and Curran, T. (1999). Altered protein conformation on DNA binding by Fos and Jun. *Nature, 347,* 572-575.

Portereiko, M.F. and Mango, S.E. (2001). Early morphogenesis of the *Caenorhabditis elegans* pharynx. *Developmental Biology, 233,* 482-494.

Raharjo, I. and Gaudet, J. (2007). Gland-specific expression of *C. elegans hlh-6* requires the combinatorial action of three distinct promoter elements. *Developmental Biology, 302,* 295-308.

Rausa, F.M., Tan, Y. and Costa, R.H. (2003). Association between hepatocyte nuclear factor 6 (HNF6) and FoxA2 DNA binding domains stimulates FoxA2 transcriptional activity but inhibits HNF-6 DNA binding. *Molecular Cell Biology, 23,* 437-449.

Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A. and Walhout, A.J.M. (2005). A compendium of *Caenorhabditis elegans* transcription factors: a resource for mapping transcription networks. *Genome Biology, 6,* R110.

Robinett, C.C., Straight, A., Li, G., Willhelm, C., Sudlow, G. et al. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using Lac operator/repressor recognition. *Journal of Cell Biology, 135,*1685-1700.

Rual J.F., Ceron, J., Koreth, J., Hao, T., Nicot A. Z. et al (2004). Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome based RNAi library. *Genome Resources, 14,* 2162-2168.

Rubins, N.E., Friedman, J.R., Le, P.P., Zhang, L., Brestelli, J. and Kaestner, K.H. (2005). Transcriptional networks in the liver: hepatocyte nuclear factor 6 function is largely independent of Foxa2. *Molecular and Cellular Biology, 25,* 7069-7077.

Ruvinsky, I., Ohler, U., Burge, C.B. and Ruvkun, G. (2007). Detection of broadly expressed neuronal genes in *C. elegans*. *Developmental Biology, 302,* 617-626.

Samadani, U. and Casta, R.H. (1996). The transcriptional activation of hepatocyte nuclear factor 6 regulates liver gene expression. *Cell Biology, 16,* 6273-6284.

Sambrook, J. and Russel, D.W. (2001). Molecular cloning a laboratory manual third edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Savage-Dunn, C. (2001). Targets of TGF-β signaling in *Caenorhabditis elegans. Cytokine and Growth Factor Reviews, 12,* 305-312.

Schrem, H., Klempnauer, J. and Borlak, J. (2002). Lver-enriched transcription factors in liver function and development. part I: the hepatocyte nuclear factor network and liver-specific gene expression. *Pharmacological Reviews, 54,* 129-158.

Sharp, P.A. (2001). RNA interference – 2001. *Genes and Development, 15,* 485-490.

Sherwood, D.R., Butler, J.A., Kramer, J.M. and Sternberg, P.W. (2005). FOS-1 promotes basement membrane removal during anchor-cell invasion in *C. elegans*. *Cell, 121,* 951-962.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S.P., Nonet, M.L., Fire, A., Ahringer, J., Plasterak, R.H. (2002). Loss of the putative RNA-directed RNA polymerase RRF3 makes *C. elegans* hypersensitive to RNAi. *Current Biology, 12,* 1317-1319.

Smith, P.A. and Mango, S.E. (2007). Role of T-box gene *tbx-2* for anterior foregut muscle development in *C. elegans. Developmental Biology, 302,* 25-39.

Sonnichsen, B., Koski, L.B., Walsh, A., Marschall, P. Neumann, B., Brehm, M. et al. (2005). Full genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans. Nature, 434,* 462-469.

Srivastava, D. (2006). Making or breaking the heart: from lineage determination to morphogenesis. *Cell, 126,* 1037-1048.

Sulston, J.E., Schierenberg, E., White, J.G. and Thomson, J.N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans. Developmental Biology, 100,* 64-119.

Suzuki, A., Iwama, A., Miyashita, H., Nakauchi, H. and Taniguchi, H. (2003). Role for growth factors and extracellular matrix in controlling differentiation of prospectively isolated hepatic stem cells. *Development, 130,* 2513-2524.

Thatcher, J.D., Fernandez, A.P., Beaster-Jones, L., Haun, C. and Okkema, P.G. (2001). The *Caenorhabditis elegans peb-1* gene encodes a novel DNA-binding protein involved in morphogenesis of the pharynx, vulva and hindgut. *Developmental Biology, 229,* 480-493.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *M0lecular Cell, 10,* 1453-1465.

Vilimas, T., Abraham, A. and Okkema, P. (2004). An early pharyngeal muscle enhancer from the *Caenorhabditis elegans ceh-22* gene is targeted by the forkhead factor PHA-4. *Developmental Biology, 266,* 388-398.

Wenick, A.S. and Hobert, O. (2004). Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans*. *Developmental Cell, 6*, 757-770.

Williams, B.D., Schrank, B., Huynh, C., Shownkeent, R. and Waterston, R.H. (1992). A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. *Genetics, 131,* 609-624.

Wood, W.B. (1988a) Determination of pattern and fate in early embryos of C*aenorhabditis elegans. Developmental Biology, 5*, 57-78.

Wood, W.B. (1988b). Introduction to *C. elegans* biology, In The nematode *Caenorhabditis elegans*, W. B. Wood, eds. (Cold Spring Harbor Laboratory Press), 1-14.