University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2015-06-01

Cell Signaling and Transcriptional Regulation in the Nematode C. elegans

Dineen, Aidan

Dineen, A. (2015). Cell Signaling and Transcriptional Regulation in the Nematode C. elegans (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. doi:10.11575/PRISM/25234

http://hdl.handle.net/11023/2285

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Cell Signaling and Transcriptional Regulation in the Nematode *C. elegans*

by

Aidan Dineen

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MOLECULAR AND MEDICAL GENETICS

DEPARTMENT OF MEDICAL SCIENCE

FACULTY OF MEDICINE

CALGARY, ALBERTA

APRIL, 2015

© Aidan Dineen 2015

Abstract

This thesis describes three investigations of basic principles in biology using the nematode *C. elegans*. It is centered on two general features of biology, cell to cell communication and the activation or inactivation of genes within an organ. In the first study, I examined the role of a conserved TGF-β signaling pathway in body size regulation. This signaling pathway is active in three organs of *C. elegans*; the pharynx, hypodermis and intestine. Previous research had concluded that signaling in the hypodermis was necessary and sufficient to regulate body size of *C. elegans.* My results demonstrated two key findings that modify the model of body size regulation. One, that this signaling pathway regulates size of the pharyngeal organ and two, that pharyngeal signaling can contribute in a minor way to overall body size regulation. These results suggest that TGF-β signaling in *C. elegans* could be coordinating growth of many cells. The second and third studies dealt with transcriptional regulation of terminal gene expression within the intestine. A hierarchy of GATA transcription factors work to specify the intestinal precursor during embryogenesis and mediate intestinal differentiation. Once embryogenesis is complete, two GATA factors, ELT-2 and ELT-7 function to regulate expression of terminal intestinal genes. These two factors are present at apparently equal levels in all intestinal cells, yet some intestinal genes are only expressed in a subset of these cells. The second project explored what other transcription factors are restricting expression of the *pho-1* gene to the posterior intestine. The results from repeating work of a previous student revealed the novel finding that *pho-1* expression is initially found in the anterior intestinal cells, then becomes restricted by the end of the first larval stage. It was confirmed that the heterochronic gene *lin-14* is repressing anterior *pho-1* expression, but not by a direct mechanism. The final project was an exploration of the intestinal targets of ELT-2 and ELT-7. RNAseq results indicate ELT-2 is the primary regulatory of gene expression in the intestine, with ELT-7 functioning partially redundantly. Furthermore, genes regulated by both factors were found to have more *cis*-regulatory elements in their promoters. In summary, these projects have furthered understanding of two conserved features of eukaryote biology, TGF-β signaling in growth regulation and GATA transcriptional regulation in the endoderm.

Acknowledgements

I would like to thank my supervisors, Dr. Jeb Gaudet, Dr. Paul Mains and Dr. Jim M^{c}Ghee for the guidance and positive words of encouragement they have provided to me over the last five years. Jim has especially been a great mentor and role model helping me to become a thorough biologist. My supervisory committee members Dr. Sarah Childs and Dr. Savraj Grewal are also deserving of my appreciation for helping guide me through the wandering path that this thesis took. I would also like to thank the dozens of lab mates and peers in and around the genes and development labs for all their backing and tireless help. Finally, I could not have finished this program without the constant love and support from my lovely wife Katrina.

Table of Contents

Abstract ii
Acknowledgements iii
Table of Contents iv
List of Tables vii
List of Figures ix
List of Symbols and Abbreviations xi
Chapter 1: *C. elegans* is a Versatile Model Organism 1
Chapter 2: TGF-β Signaling Acts From Multiple Tissues to Regulate *C. elegans* Body Size 3
2.1 Abstract 3
2.2 Background 3
2.3 Materials and Methods 5
2.3.1 *C. elegans* Strains 5
2.3.2 Plasmid Construction 5
2.3.3 Generating Transgenics by Microinjection 7
2.3.4 Length Measurements 7
2.3.5 Egg to Egg Timing 8
2.4 Results 8
2.4.1 The TGF-β Sma/Mab Pathway Regulates Pharynx Length 8
2.4.2 SMA-3 Can Act in the Pharynx to Regulate Body Length 9
2.4.3 Sma/Mab Signaling in Multiple Tissues Controls Overall Body Length 21
2.5 Discussion 28
2.5.1 The TGF-β Sma/Mab Pathway Regulates Pharynx Length 28
2.5.2 Activity of Sma/Mab Pathway in Multiple Tissues Controls Body Length 29
2.6 Conclusions 31
Chapter 3: *pho-1* Spatial Patterning is Regulated Indirectly by LIN-14 35
3.1 Abstract 35
3.2 Background 36
3.2.1 Overview 36
3.2.2 Anatomy and Function of the *C. elegans* Intestine 40
3.2.3 Morphogenesis of the *C. elegans* Intestine 44
3.2.4 Specification of the *C. elegans* Intestine 47

3.2.5 Spatial Patterning of Terminal Gene Expression in the *C. elegans* Intestine....52
3.2.6 LIN-14 and the Heterochronic Pathway of *C. elegans*....56
3.3 Materials and Methods....58
3.3.1 *C. elegans* Strains....58
3.3.2 Promoter Sequence Alignment....58
3.3.3 Plasmid Sequence Construction....58
3.3.4 Generating Transgenics by Microinjection....58
3.3.5 Image Acquisition and GFP Intensity Quantitation....58
3.4 Results....59
3.4.1 The *pho-1* Spatial Expression Pattern Forms During the L1 Larval Stage....59
3.4.2 Analysis of Regulatory Sites in the *pho-1* Promoter....71
3.4.3 LIN-14 is Required for Repression of *pho-1* Expression in the Intestine....74
3.4.4 LIN-14 Does Not Act Directly on the *pho-1* Promoter....80
3.4.5 Intestinal Development is not as Invariant as Previously Thought....80
3.5 Discussion....87
3.5.1 *pho-1* Expression is Patterned During the First Larval Stage....87
3.5.2 LIN-14 Regulates *pho-1* Spatial Patterning Indirectly....88
Chapter 4: Deciphering the Contributions to Intestinal Gene Expression by the GATA Transcription Factors ELT and ELT-7....90
4.1 Abstract....90
4.2 Background....91
4.2.1 Three GATA Transcription Factors are Expressed in the Intestine After Hatching....91
4.2.2 Regulation of Intestinal Gene Expression....92
4.3 Materials and Methods....94
4.3.1 *C. elegans* Strains....94
4.3.2 COPAS Biosorting and RNA Extraction....94
4.3.3 Library Preparation and RNAseq....95
4.3.4 Bioinformatic Analysis of RNAseq Results....96
4.3.5 Western Blot for ELT-2....96
4.4 Results....97
4.4.1 Isolation of Genetically Distinct *C. elegans* Populations....97
4.4.2 General Findings of the RNAseq....105

4.4.3 ELT-2 is Necessary to Activate Expression of Some Intestinal Genes.....118
4.4.4 ELT-2 is a Major Activator of Intestinal Genes While ELT-7 is a Minor Activator.....125
4.4.5 Joint Targets Have More *cis*-Regulatory Elements in their Promoters.....134
4.4.6 ELT-2 and ELT-7 are Required for Repression of Some Intestinal Genes.....141
4.4.7 Known Targets of ELT-2 and ELT-7.....149
4.4.8 Intestinal Transcription Regulation More Complex than ELT-2 and ELT-7..153
4.5 Discussion.....157
4.5.1 ELT-2 is the Major Effector of Intestinal Transcriptional Regulation.....157
4.5.2 Role of ELT-7 in the Intestinal Transcriptional Hierarchy.....160
4.5.3 More TGATAA Sites in Promoters of Genes Regulated by Both ELT-2 and ELT-7.....161
4.5.4 ELT-2 and ELT-7 Also Function as Repressors in the Intestine.....165
4.5.5 Regulation of Intestinal Genes Involves Many Transcription Factors.....167
4.5.6 Negative Feedback on ELT-2 and ELT-7 Expression.....170
4.6 Conclusions.....172
Chapter 5: Overall Thesis Conclusions.....**173**
5.1 Three Independent Studies with Connecting Threads.....173
Bibliography.....175
Appendices A: Chapter 2 Tables.....193
Appendices B: Chapter 2 Supplemental Figures.....196
Appendices C: Chapter 4 Tables.....206

List of Tables

Table 2.1: Frequency of Dpy phenotype rescue 193

Table 2.2: Percentage of Wt and *sma-3(e491)* animals (+ or - pharyngeal *sma-3* minigene constructs) that had laid an egg by the time point indicated (hrs after initial egg laid) 193

Table 4.1: Strains used to generate genetically distinct populations for RNAseq.....206

Table 4.2: Mean body length measurements of sorted L1 populations 206

Table 4.3: Mean total reads obtained from RNAseq per population and mean percentage of those reads that mapped to the *C. elegans* genome 206

Table 4.4: Differentially expressed gene counts in various comparisons arranged by log2 fold change and expression in the intestine 207

Table 4.5: Log2 fold changes and associated adjusted p values for the wild-type *vs elt-2 (+++)* comparison and related information for intestinal genes that positively respond to ELT-2 207

Table 4.6: Log2 fold changes and associated adjusted p values for the wild-type *vs elt-2(ca15)* comparison and related information for intestinal genes that are exclusively activated by ELT-2 209

Table 4.7: Log2 fold changes and associated adjusted p values for the wild-type vs *elt-7(tm840); elt-2(ca15)* comparison and related information for intestinal genes that are redundantly activated by ELT-2 or ELT-7 210

Table 4.8: Log2 fold changes and associated adjusted p values for the wild-type vs *elt-7(tm840)*, wild-type *vs elt-2(ca15)* and wild-type vs *elt-7(tm840); elt-2(ca15)* comparisons and related information for intestinal genes that are additively activated by ELT-2 and ELT-7 211

Table 4.9: Log2 fold changes and associated adjusted p values for the wild-type vs *elt-7(tm840)*, wild-type vs *elt-2(ca15)* and *elt-2(ca15)* vs *elt-7(tm840); elt-2(ca15)* comparisons and related information for intestinal genes that are not synergistically activated by ELT-2 and ELT-7.....212

Table 4.10: Position frequency matrices from the different classes of activated intestinal genes and the originally identified intestinal gene promoter PFM from McGhee et al. (2009)* 213

Table 4.11: Log2 fold changes and associated adjusted p values for the wild-type vs *elt-2(ca15)* comparison and related information for intestinal genes that are exclusively repressed by ELT-2.....213

Table 4.12: Log2 fold changes and associated adjusted p values for the wild-type vs *elt 7(tm840); elt-2(ca15)* comparison and related information for intestinal genes that are redundantly repressed by ELT-2 or ELT-7.....215

Table 4.13: Log2 fold changes and associated adjusted p values for the wild-type *vs elt 7(tm840)*, wild-type vs *elt-2(ca15)* and wild-type vs *elt-7(tm840); elt-2(ca15)*

comparisons and related information for intestinal genes that are additively repressed by ELT-2 and ELT-7 215

Table 4.14: Log2 fold changes and adjusted p values for the wild-type vs *elt-7(tm840)*, wild type vs *elt-2(ca15)* and wild-type vs *elt-7(tm840); elt-2(ca15)* comparisons for intestinal genes identified as repressed by both ELT-2 and ELT-7 by Sommermann et al. (2010) 216

Table 4.15: Differentially expressed intestinal transcription factors 216

Supplemental Table 2.1: Pharynx and Body Lengths of Various Strains.....193

Supplemental Table 2.2: Tissue Specificity of Promoters Used to Drive Rescue Constructs 194

Supplemental Table 2.3: Pharynx and Body Lengths of Strains Imaged Using Various Anesthetics at Different Time Points After Egg Laying During Development 194

List of Figures

Figure 2.1 11
Figure 2.2 13
Figure 2.3 17
Figure 2.4 19
Figure 2.5 26
Figure 2.6 33
Figure 3.1 39
Figure 3.2 43
Figure 3.3 46
Figure 3.4 50
Figure 3.5 55
Figure 3.6 61
Figure 3.7 63
Figure 3.8 66
Figure 3.9 68
Figure 3.10 70
Figure 3.11 73
Figure 3.12 76
Figure 3.13 79
Figure 3.14 82
Figure 3.15 84
Figure 3.16 86
Figure 4.1 99
Figure 4.2 101
Figure 4.3 104
Figure 4.4 108
Figure 4.5 110
Figure 4.6 113
Figure 4.7 117
Figure 4.8 120
Figure 4.9 123
Figure 4.10 127

Figure 4.11 129
Figure 4.12 132
Figure 4.13 136
Figure 4.14 140
Figure 4.15 143
Figure 4.16 145
Figure 4.17 148
Figure 4.18 151
Figure 4.19 156
Figure 4.20 163
Supplemental Figure 2.1 197
Supplemental Figure 2.2 199
Supplemental Figure 2.3 201
Supplemental Figure 2.4 203
Supplemental Figure 2.5 205

List of Symbols and Abbreviations

TGF-β:	Transforming Growth Factor-β
Sma/Mab:	Small and Male Abnormal
DIC:	Differential Interference Contrast
GFP:	Green Fluorescent Protein
Wt:	Wild-type
NEXTDB:	Nematode Expression Pattern Database
-:	Non-transgenic Animal
+:	Transgenic Animal
ECM:	Extracellular Matrix
Twp:	Twisted Pharynx
Int:	Intestinal Ring
SOP:	Sensory Organ Precursor (*Drosophila*)
PH:	Post-hatching
KO:	Knockout
COPAS:	Complex Object Parametric Analyzer and Sorter
tof:	Time of Flight
FPKM:	Fragments per Kilobase Exon per Million Mapped Reads
SAGE:	Serial Analysis of Gene Expression
FDR:	False Discovery Rate
RSAT:	Regulatory Sequence Analysis Tools
PFM:	Position Frequency Matrix
SELEX:	Systematic Evolution of Ligands by Exponential enrichment
ChIP:	Chromatin Immunoprecipitation

Chapter 1: *C. elegans* is a Versatile Model Organism

Studying fundamental aspects of human biology is an essential component of creating better treatments and cures for human disease. It is often unethical or impractical to study basic questions of biology in humans. The use of model organisms is therefore an advantageous method that can be used to build a framework of understanding to be applied to our species.

The roundworm *C. elegans* is a very tractable tool that is well suited for answering questions about key aspects of biology. The small size of this nematode (1mm), short life cycle (~3 days) and large number of progeny allows for cheap cultivation and maintenance in the laboratory. There are a number of aspects of *C. elegans* biology that make it a great organism for studying questions of developmental biology. The worm is transparent, which allows for easy microscopic imaging of cells, gene expression patterns and cellular localization of proteins. It was the second eukaryote to have its genome completely sequenced (Consortium, 1998) - the high quality annotation of which is convenient for large multi-gene studies. *C. elegans* has five pairs of autosomes and one pair of sex chromosomes allowing for easy genetic manipulation (Brenner, 1974; Nigon, 1949). The hallmark of *C. elegans* developmental biology is its mapped, invariant cell lineage producing 959 cells (Sulston et al., 1983). This has made it possible to solve complex questions of cell fate specification and organ development.

Here, I describe three independent studies undertaken to acquire better understanding of the processes of cell signaling and transcriptional regulation of gene expression. First, an investigation of how body size and organ size is regulated in *C. elegans* by a TGF-β signaling pathway. TGF-β signaling pathways are conserved amongst metazoans and used widely and repeatedly in different contexts to regulate aspects of organismal development (Gerhart, 1999). Coordinating growing cells and organs within an organism is a central requirement of multicellular organisms and applies to multiple aspects of human biology. The second and third studies aimed to further our understanding of endoderm gene regulation by GATA transcription factors. GATA transcription factors, like TGF-β signaling pathways, are found in organisms from simple slime molds to humans (Lowry and Atchley, 2000) and play critical roles during

organogenesis in vertebrates (Aronson et al., 2014; Costa et al., 2001; Viger et al., 2008). I conclude the thesis with a brief discussion of relationships between these projects and the importance of the conclusions from each study.

Chapter 2: TGF-β Signaling Can Act from Multiple Tissues to Regulate *C. elegans* Body Size (Dineen and Gaudet, *BMC Developmental Biology, 2014*)

2.1 Abstract

Background: Regulation of organ and body size is a fundamental biological phenomenon, requiring tight coordination between multiple tissues to ensure accurate proportional growth. In *C. elegans*, a TGF-β pathway is the major regulator of body size and also plays a role in the development of the male tail, and is thus referred to as the TGF-β/ Sma/Mab (for small and male abnormal) pathway. Mutations in components of this pathway result in decreased growth of animals during larval stages, with Sma mutant adults of the core pathway as small as ~60-70% the length of normal animals. The currently accepted model suggests that TGF-β/Sma/Mab pathway signaling in the *C. elegans* hypodermis is both necessary and sufficient to control body length. However, components of this signaling pathway are expressed in other organs, such as the intestine and pharynx, raising the question of what the function of the pathway is in these organs. **Results:** Here we show that TGF-β/Sma/Mab signaling is required for the normal growth of the pharynx. We further extend the current model and show that the TGF-β/Sma/Mab pathway can function in multiple tissues to regulate body and organ length. Specifically, we find that pharyngeal expression of the SMAD protein SMA-3 partially rescues both pharynx length and body length of *sma-3* mutants. **Conclusions:** Overall, our results support a model in which the TGF-β/Sma/Mab signaling pathway can act in multiple tissues, activating one or more downstream secreted signals that act non cell-autonomously to regulate overall body length in *C. elegans*.

2.2 Background

An important question in developmental biology is what controls growth at three levels: the organism, the organ and the individual cells (Reddy and Irvine, 2008). Organismal size appears to be regulated by multiple inputs including genetic pathways that are active during development to regulate cell number (cell proliferation and apoptosis) and cell size. Overall body size of an organism also responds to environmental cues such as nutrient availability and stress. Many of these environmental inputs

converge on the Tor signaling pathway, which regulates multiple downstream targets to ultimately control both cell size and cell division (Grewal, 2009). Similarly, the size of an organ can be determined by cell number and/or cell size, again controlled by genetic and environmental components. In Drosophila and mammals, the regulation of individual cell size is controlled in part by a conserved insulin signaling pathway that receives nutritional input and translates this information to regulate cellular metabolism (Hyun, 2013).

Growth of the nematode *C. elegans* occurs through both increase in cell number and increase in cell size. From hatching to adulthood, the number of somatic nuclei increases from 550 to 959 (Sulston et al., 1983), while cells also increase in size. As in other animals, *C. elegans* body size is regulated by nutrient status. For example, animals defective in feeding are significantly smaller than wild type animals (Mörck and Pilon, 2006; Smit et al., 2008). Genetic regulation of size involves at least two signaling pathways: a much less studied pathway that includes the MAP kinase SMA-5, and the major pathway involving TGF-β (Savage et al., 1996; Watanabe et al., 2005). These two pathways act non-redundantly in body size regulation and may also act independently of nutritional status (Roberts et al., 2010; Watanabe et al., 2007). Additionally, body size can be constrained by morphology defects in the extracellular cuticle surrounding the worm resulting in the Dpy phenotype (Brenner, 1974; Kusch and Edgar, 1986).

The TGF-β pathway is referred to as the Sma/Mab pathway because loss of function mutations lead to both small body length (Sma phenotype) and male tail defects (Mab phenotype). The pathway ligand, DBL-1, is expressed in a set of neurons, including some pharyngeal neurons (Morita et al., 1999; Suzuki et al., 1999). Binding of DBL-1 to the Type I/II receptors SMA-6 and DAF-4 activates the downstream effector SMADs SMA-2, -3 and -4, which function together with the Schnurri homolog SMA-9 to regulate transcription of target genes *in C. elegans*, none of which have yet been identified (Gumienny and Savage-Dunn, 2013). Previous work has suggested that TGF-β/Sma/Mab signaling acts solely in the hypodermis to control organismal length (Wang et al., 2002; Yoshida et al., 2001). However, components of the pathway, such as SMA-3 and SMA-6, are expressed in additional tissues, namely the pharynx and intestine (Wang et al., 2002;

Yoshida et al., 2001), where it has been suggested that they regulate innate immunity genes (Mallo et al., 2002; Roberts et al., 2010; Zugasti and Ewbank, 2009).

We have previously reported that the pharynges of *sma-2(e502)* and *sma-3(e491)* mutants are shorter in length than wild type (Raharjo et al., 2011). To test the hypothesis that pharynx length is regulated by TGF-β/Sma/Mab signaling in pharyngeal cells, we performed rescue experiments in which *sma-3(+)* was expressed under the control of different tissue-specific promoters. We found that expression of *sma-3* in the pharynx could partially rescue both pharynx length and body length of *sma-3* mutants, in contrast to expectations based on the prevailing model. Our findings suggest that TGF-β/Sma/Mab signaling can function in multiple tissues (hypodermis and pharynx) to control organ and overall body length.

2.3 Materials and Methods

2.3.1 *C. elegans* Strains

Standard nematode handling conditions were used (Brenner, 1974). Animals were grown at 20°C. Strains used were wild type N2, CB61 *dpy-5(e61) I,* DR1785 *mIn1[dpy-10(e128)]/unc-4(e120) II,* CB1482 *sma-6(e1482) II,* CB491 *sma-3(e491) III,* CS24 *sma-3(wk30) III* (kindly provided by Dr. Cathy Savage-Dunn, Queens College), MT468 *dpy-7(e88) unc-6(n102) X* and JM228 *ctIs40*[*dbl-1(+) sur-5::gfp*] *X; sma-3(e491) III; ivEx163[myo-2p::sma-3 marg-1p::sma-3 phat-1p::yfp elt-2p::tdTomato::His2B]*.

2.3.2 Plasmid Construction

To construct a *sma-3* minigene, we amplified *sma-3* cDNA from a library using primers oGD861 acggtaccATGAACGGATTACTGCATATGCATGGTC and oGD860 tagagctcTTATGTCATTGAATTTGGTTCCATCAAGTTCG; for all oligos, uppercase sequence corresponds to gene sequence; lowercase corresponds to restriction site-containing sequence or plasmid sequence that facilitates cloning. This 1.2 kb fragment was cut with *Kpn*I and *Sac*I and cloned into the *myo-2*-containing plasmid pSEM474 (Gaudet and Mango, 2002) to create a *myo-2p::sma-3(cDNA)* plasmid. We next amplified a 2.9 kb genomic *sma-3* fragment from N2 DNA using the same oligos and

digested with *Bgl*II and *Sal*I to isolate a 1 kb genomic fragment containing exons 2-8 (and introns 2-7). This genomic fragment was cloned into the *myo-2p::sma-3(cDNA)* construct to generate *myo-2p::sma-3* minigene. The *sma-3* minigene was sequenced to ensure that no mutations were introduced during cloning.

Other minigene constructs contained the same *sma-3* minigene cassette but with different promoter sequences, amplified with the following pairs of primers from either genomic N2 DNA or pRF4 (in the case of *rol-6*):

sma-3: oGD3 gctgaaatcactcacaacgatgg

oGD1230 cggggtaccTTGCTCTCATTTCAAAAAAACTAATTC

marg-1: oGD316 aactgcagATCAAAGTGCCGATCGAAGT

oGD317 ggggtaccGTTGGAGGAGCCATTGAGA

rol-6: oGD1047

gagactgcagGTTTTGATAAAATTGTGGTGTAGTCCATAATG

oGD1048

gagaggtaccCTGGAAATTTTCAGTTAGATCTAAAGATATATCC

The K07C11.4 promoter was cloned from the previously described reporter plasmid pSEM900 (Gaudet et al., 2004).

To examine *sma-3* expression, we amplified the entire *sma-3* gene, including ~1.2 kb of sequence upstream of the predicted ATG (the entire intergenic region), using the primers oGD956 (caactgcagCTTGCTAACTGTGTCCCCAACCATC) and oGD957 (catggtaccGTCATTGAATTTGGTTCCATCAAGTTCG). We digested this fragment with *Pst*I and *Kpn*I and cloned it into the GFP expression vector pPD95.77, creating an in-frame translational fusion between *sma-3* and *gfp*. To create a *sma-3p::sma-3* minigene, we isolated a *Pst*I-*Bgl*II fragment from the *sma-3::gfp* vector and cloned it into a *Pst*I-*Bgl*II cut *myo-2p::sma-3* minigene construct, effectively swapping the *myo-2* promoter for the *sma-3* promoter.

To test rescue of *dpy-7*, we used the same promoter fragments as above but replaced the *sma-3* minigene with genomic *dpy-7* sequence, amplified from N2 genomic DNA using the primers oGD864 (ccaaggtaccATGGAGAAGCCCAGTTCGGG) and oGD865 (ccaagagctcTTATTTCTTTCCATAACCACCACCAG), and digested with *Kpn*I and *Sac*I. The *dpy-7* promoter fragment was amplified using the primers oGD989

(aactgcagTGGCGCAAGAGGCAGTGC) and oGD990 (cggggtaccTTATCTGGAACAAAATGTAAGA).

2.3.3 Generating Transgenics by Microinjection

C. elegans transgenic lines were created using standard microinjection techniques (Mello et al., 1991). *sma-3* rescuing constructs were injected at 5-50 ng/µL, as noted in Tables and Figures, together with 30 ng/µL of either the intestinal reporter *elt-2p::tdTomato::His2B* (pJM371) or the body wall muscle reporter *myo-3p::wCherry* (kindly provided by Dr. Mei Zhen, University of Toronto) and pBlueScriptII (KS+) to a total DNA concentration of 100 ng/µL. Transgenic rescue was performed by injection directly into *sma-3* mutants, followed by screening for td-Tomato positive or wCherry-positive F1 transgenic animals, which were transferred to new plates. We specifically screened for transgenic larvae (rather than adults) to avoid biasing our selection for larger animals. We then similarly screened for F2 transgenics to establish stable lines. For rescue of *dpy-7*, transgenes were injected into wild type animals at 20 ng/µL together with 30 ng/µL of either *elt-2p::tdTomato::His2B* or *myo-3p::wCherry* and 50 ng/µL pBlueScriptII (KS+). Transgenic males were mated to *dpy-7 unc-6* mutants. Rescue of *dpy-7* was initially assayed in cross-progeny Unc males (*dpy-7 unc-6/0*). Non-Dpy non-Unc hermaphrodite cross-progeny were also isolated, and their Unc transgenic progeny were scored for the presence or absence of the Dpy phenotype.

2.3.4 Length Measurements

Young adult hermaphrodites were allowed to lay eggs for 2 hours at 20^{o}C after which the adults were removed and the progeny were incubated at 20^{o}C for 96 hours. Three methods were used to obtain length measurements. In the first two methods, 96 ± 1 or 120 ± 1 hour old animals were transferred to 2% agar pads on glass slides, anaesthetized with either 5 mM levamisole or 20mM sodium azide diluted in 1X M9 buffer and photographed at 40x magnification under DIC optics. Levamisole was used in all experiments except those which indicate otherwise. Images were captured using a Zeiss Axio Imager.Z1 microscope with a Zeiss AxioCam MRm camera and AxioVision (4.8.1) software. Pharynx and body lengths were measured in ImageJ using segmented

lines (Collins, 2007). Pharynx length was measured as the distance from the posterior of the buccal cavity to the pharyngeal-intestinal valve. Calibration was achieved using a Pyser-SGI micrometer slide. Sigma Plot 12.5 was used to perform Mann-Whitney rank sum tests for statistical significance. For the final method, 96 ± 1 hour old animals were imaged directly on growth plates, in the absence of anesthetic, under a Zeiss Stemi SV11 dissecting microscope with a Canon PC1210 camera. Length measurements and statistical analysis was performed as outlined above.

2.3.5 Egg to Egg Timing

Individual gravid hermaphrodites were picked to plates and checked periodically for egg laying. When an egg was observed to have been laid the time was recorded and the worm was picked off the plate along with any extra eggs. The laid eggs were allowed to develop at 20 ^{o}C and were checked at periodic times to see if the animals had begun to lay eggs on the plate.

2.4 Results

2.4.1 The TGF-β/Sma/Mab pathway regulates pharynx length

Previous reports indicated that TGF-β/Sma/Mab signaling in the hypodermis controls body length (Gumienny et al., 2010; Wang et al., 2002; Yoshida et al., 2001). In particular, the small body length of *sma-3* mutants was rescued to comparable levels by either hypodermal expression or by the native *sma-3* promoter (Wang et al., 2002), leading to the current model that hypodermal action of the TGF-β/Sma/Mab pathway is necessary and sufficient for regulation of body length. Interestingly, pharyngeal expression of the rescuing construct using the *myo-2* pharyngeal muscle promoter also resulted in a small but statistically significant increase in body length of *sma-3* mutants (Wang et al., 2002). Furthermore, pharynx lengths of Sma mutants at the L3 stage were found to be slightly but significantly smaller than the pharynx length of wild type N2. These two pieces of evidence suggest first, that TGF-β/Sma/Mab pathway signaling may regulate pharynx length and second, that signaling within the pharynx may contribute significantly to body length regulation.

We previously found that *sma-2(e502)* and *sma-3(e491)* mutants have adult pharynges that are 79 ± 4% and 76 ± 2% the length respectively of N2 pharynges (when measured 96 ± 1 hours after adult hermaphrodites were allowed to lay eggs for two hours at 20^{o}C) (Raharjo et al., 2011). We also find that *sma-3(wk30)* and *sma-6(e1482)* mutants have pharynges that are 81 ± 2% of N2 length (Figure 2.1, Supplemental Table 2.1). The reduced pharynx length in these TGF-β/Sma/Mab pathway mutants suggests that pathway activity is required for pharyngeal growth, consistent with expression of pathway components in the pharynx. However, another possibility is that pharynx length is reduced as a consequence of reduced body length, specifically, that growth of the pharynx might be constrained by the smaller hypodermis. To test this latter possibility, we measured pharynx length in two hypodermal collagen mutants with reduced body length, *dpy-5(e61)* and *dpy-10(e128)* (Levy et al., 1993; Thacker et al., 2006). The body lengths of the Dpy and Sma animals are comparable. For example, *dpy-5(e61)* and *sma-3(wk30)* are 63 ± 4% and 62 ± 6% respectively of N2 body length, not a statistically significant difference (p=0.225, Mann-Whitney Rank Sum Test on raw data). Pharyngeal lengths of *dpy-5* and *dpy-10* mutants are 89 ± 3% and 92 ± 3% of N2 length respectively, significantly greater than that of Sma mutants (p<0.001, Mann-Whitney Rank Sum Test on raw data) (Figure 2.1, Supplemental Table 2.1). These measurements imply that pharynx length may be partially reduced in response to smaller body length but is also positively influenced by TGF-β/Sma/Mab signaling. We therefore propose that some aspect of pharyngeal growth requires TGF-β/Sma/Mab signaling.

2.4.2 The TGF-β/Sma/Mab effector SMA-3 can act in the pharynx to regulate body length

sma-3 is reported to be expressed in the pharynx (as are other components of the pathway) (Wang et al., 2002; Yoshida et al., 2001), yet the individual cells in which it is expressed have not been described. We constructed a *sma-3p::sma-3::gfp* translational reporter to determine in which pharyngeal cells *sma-3* might function. As previously described (Wang et al., 2002), expression of the reporter was observed in hypodermis, intestine and pharynx. Within the pharynx, we observe expression in most or all pharyngeal muscles and marginal cells (Figure 2.2A). Given the expression pattern of

Figure 2.1. Mean pharynx and body length measurements ± standard deviation of wild type (Wt) N2 and body size mutants at 96 ± 1 hrs AEL. Complete data is provided in Supplemental Table 2.1. * denotes a statistically significant difference of $p<0.001$. All other differences in pharynx length between strains but not directly indicated on the graph are significant ($p<0.001$) with the exception of N.S. (not significant). All differences in body length between strains not directly indicated on the graph are significant ($p<0.05$), except where indicated by N.S.

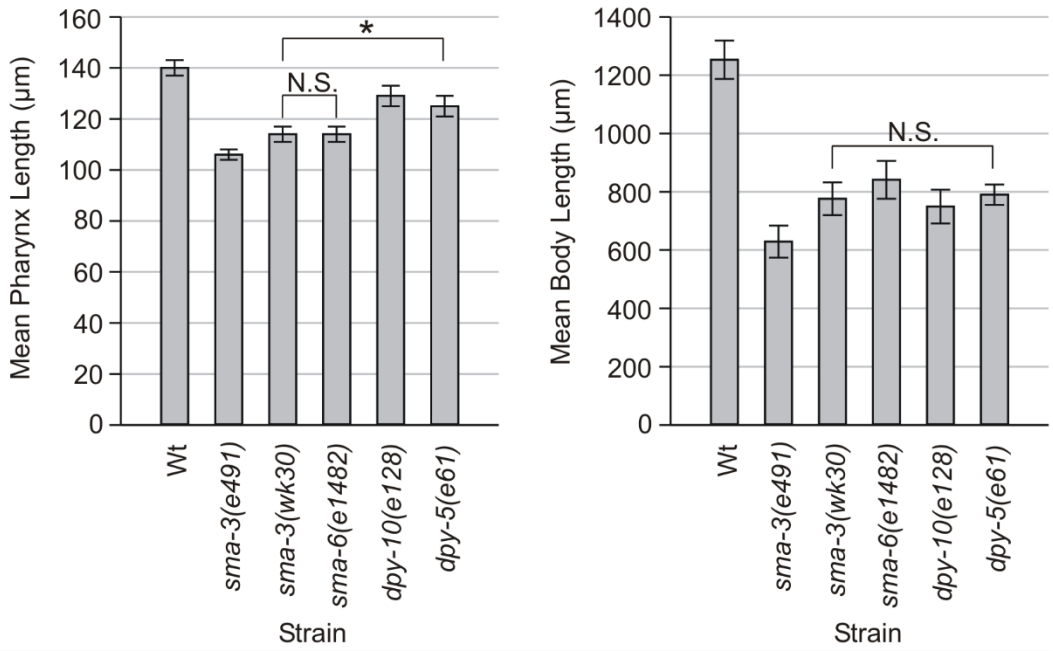
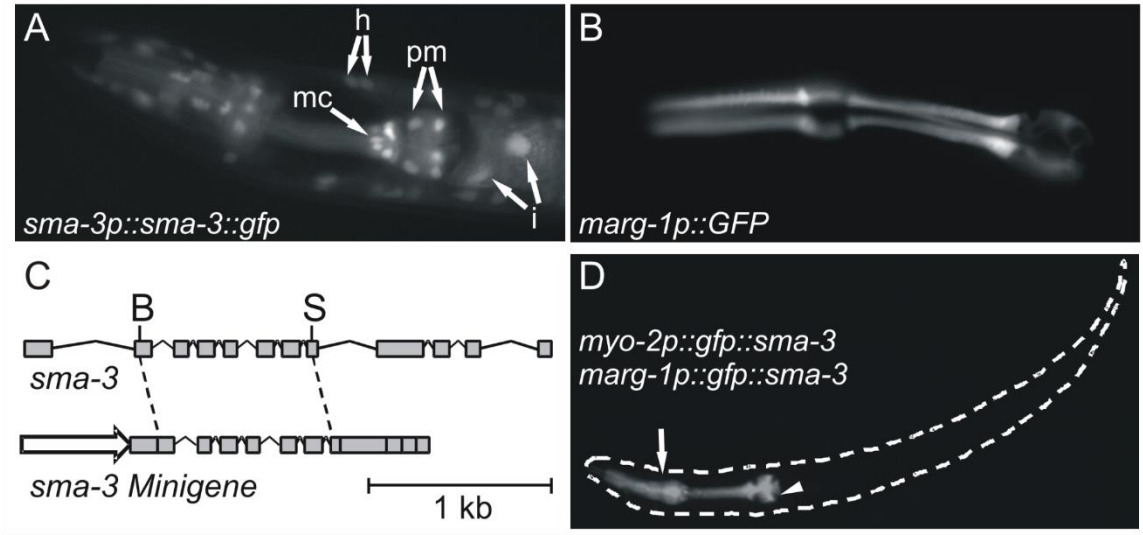

Mean Pharynx Length (µm)
160
140
120
100
80
60
40
20
0
*
N.S.
Wt
sma-3(e491)
sma-3(wk30)
sma-6(e1482)
dpy-10(e128)
dpy-5(e61)
Strain
Mean Body Length (µm)
1400
1200
1000
800
600
400
200
0
N.S.
Wt
sma-3(e491)
sma-3(wk30)
sma-6(e1482)
dpy-10(e128)
dpy-5(e61)
Strain

Figure 2.2. (A) Expression of the *sma-3p::sma-3::gfp* translational fusion. Expression is visible in the nuclei and cytoplasm of hypodermal cells (h), intestinal cells (i), pharyngeal muscles (pm) and pharyngeal marginal cells (mc). (B) The *F47B7.7 (marg-1)* transcriptional reporter is strongly expressed in pharyngeal marginal cells. (C) Top, a genomic *Bgl*II (B) - *Sal*I (S) *sma-3* fragment was cloned into a similarly digested *sma-3* cDNA clone to create the *sma-3* "minigene" (below) used for rescue experiments, under the control of various promoters (arrow); see text for details. (D) Pharyngeal expression of the *sma-3* minigene carrying an in-frame N-terminal GFP tag under the control of the *myo-2* and *marg-1* promoters. Expression is absent from pharyngeal gland cells (arrowhead) and weak expression is occasionally observed outside of the pharynx (arrow).


A
h
pm
mc
i
sma-3p::sma-3::gfp
B
marg-1p::GFP
C
B
S
sma-3
sma-3 Minigene
1 kb
D
myo-2p::gfp::sma-3
marg-1p::gfp::sma-3

sma-3 in pharyngeal cells and the decreased pharynx length of *sma-3* mutants, we next asked whether pharynx length could be rescued by pharyngeal expression of *sma-3*. We performed tissue specific rescue experiments in two different *sma-3* mutant strains. The *sma-3(wk30)* mutant contains an early stop codon in *sma-3,* is predicted to be a molecular null and behaves like a genetic null allele (Savage-Dunn et al., 2000). The *sma-3(e491)* mutant contains a missense mutation in the MH2 domain of SMA-3 that is predicted to be a loss of function and genetically behaves like a strong hypomorph (Savage-Dunn et al., 2000). However, the *sma-3(e491)* allele may have dominant neomorphic properties as both mean pharynx and body length are significantly smaller in heterozygous ($p<0.001$) and homozygous ($p<0.001$) animals compared to the *sma-3(wk30)* nulls (Supplemental Table 2.1, Supplemental Figure 2.1).

We attempted to rescue pharynx length by expressing a *sma-3(+)* 'minigene' under the control of different pharyngeal promoters (see Supplemental Table 2.2 for a list of all promoters used in rescue experiments and their tissue specificity): *myo-2*, which is expressed solely in pharyngeal muscles (Okkema et al., 1993) and *K07C11.4*, which is expressed in pharyngeal muscle, marginal cells and epithelia, the intestine, and in late stage somatic gonad (Gaudet et al., 2004). We also identified *marg-1/F47B7.7* as a marker for pharyngeal marginal cells based on our search of the Nematode Expression Pattern Database, NEXTDB (Kohara, 2001; NEXTDB The Nematode Expression Pattern Database). A transcriptional reporter containing 2 kb of sequence upstream of the predicted *marg-1* start codon recapitulates this pattern of expression, showing strong expression in all marginal cells and weak, variable expression in pharyngeal epithelial cells and arcade cells and in the excretory cell of adults (Figure 2.2B and data not shown).

The *sma-3* minigene used for rescue combines both *sma-3* cDNA and *sma-3* genomic sequence. On the one hand, the presence of introns can improve transgene expression (Gaudet et al., 1996; Okkema et al., 1993) (Figure 2.2C). On the other hand, large introns can often contain control elements in *C. elegans* (Gaudet and Mango, 2002; Shibata et al., 2000) and therefore these were excluded from the *sma-3* minigene to avoid affecting the tissue specific expression of the construct. The final *sma-3* minigene used in rescue experiments contains all 12 exons as well as six small introns (2-7). We first

constructed *myo-2p::gfp::sma-3* and *marg-1p::gfp::sma-3* constructs, in which the minigene is fused in-frame to GFP, to verify that the *sma-3* introns did not influence the pattern of expression. As expected, in all stages we observed strong expression of these translational fusions in the pharynx with notable lack of expression in the pharyngeal glands (Figure 2.2D). Extremely weak expression was also observed in one to four cells just outside of the pharynx in half of the animals (n=140, Supplemental Figure 2.2) but only when the exposure was increased dramatically. We noticed, however, that the GFP signal was often punctate, possibly reflecting aggregation of the fusion protein. We relied on our *sma-3* minigene (lacking GFP) because our GFP fusion had little rescuing activity (data not shown) whereas a previous report used a GFP::SMA-3 fusion to rescue *sma-3(wk30)* mutants (Wang et al., 2002). As a positive control, we tested for rescue of *sma-3(e491)* mutants when the minigene was expressed under the control of the *sma-3* promoter (*sma-3p::sma-3*). (For all rescue experiments, the data charted in figures and referenced in the text refers to line A of each transgenic strain from the Supplemental Tables; mean lengths of transgenics are presented as a percentage of non-transgenic siblings mean length). As expected, this transgene exhibited significant rescue of body length (155 ± 16% compared to 100 ± 8% for non-transgenic siblings, $p<0.001$), though not to full N2 levels, possibly reflecting the artificial nature of *C. elegans* transgenic arrays and *sma-3* minigene (Figure 2.3, Supplemental Table 2.3). We also found that pharynx length of *sma-3(e491)* mutants was significantly restored by the *sma-3p::sma-3* transgene (119 ± 6% compared to 100 ± 4% for non-transgenic siblings, $p<0.001$), consistent with the expectation that pharynx length is regulated by *sma-3*. In all experiments, we measured non-transgenic siblings as a control and note that none of the strains used in these experiments displayed any strong mosaic expression of the transgenic arrays.

We next tested whether the *myo-2p::sma-3* and *marg-1p::sma-3* transgenes could rescue *sma-3(e491)* pharynx length to N2 levels, either alone or in combination (Figures 2.3A, 2.4, Supplemental Table 2.3). *myo-2p::sma-3* rescued animals had pharynx lengths that were 110 ± 5% compared to 100 ± 3% for non-transgenic siblings and *marg-1p::sma-3* rescued animals were 107 ± 3% of N2 length compared to 100 ± 2% for non-transgenic siblings. The relative small effect of rescue on pharynx length by each of these

Figure 2.3. Mean pharynx and body length measurements ± standard deviation of *sma-3(e491)* (A, B) and *sma-3(wk30)* (C, D) animals from various *sma-3* minigene rescue experiments at 96 hrs AEL. Wild type (Wt) N2 is included for comparison. Vertical labels indicate tissue specific promoter-*sma-3* minigene fusion rescue constructs in each strain, except for *elt-3p::gfp::sma-3* which indicates the *elt-3p::gfp::sma-3* transgene (pCS223) was used. In each case, we measured animals carrying the transgenic array (+) and siblings that lacked the array (-), as indicated by presence of the transformation marker (either *elt-2p::tdTomato::His2B* or *myo-3p::wCherry*). For each transgene tested, a representative line is shown; complete data for multiple lines is provided in Supplemental Table 2.3. All transgenic animal means (+) were statistically significantly different from non-transgenic sibling means (-) ($p<0.001$) unless otherwise indicated. ** denotes a significant difference of $p<0.001$.

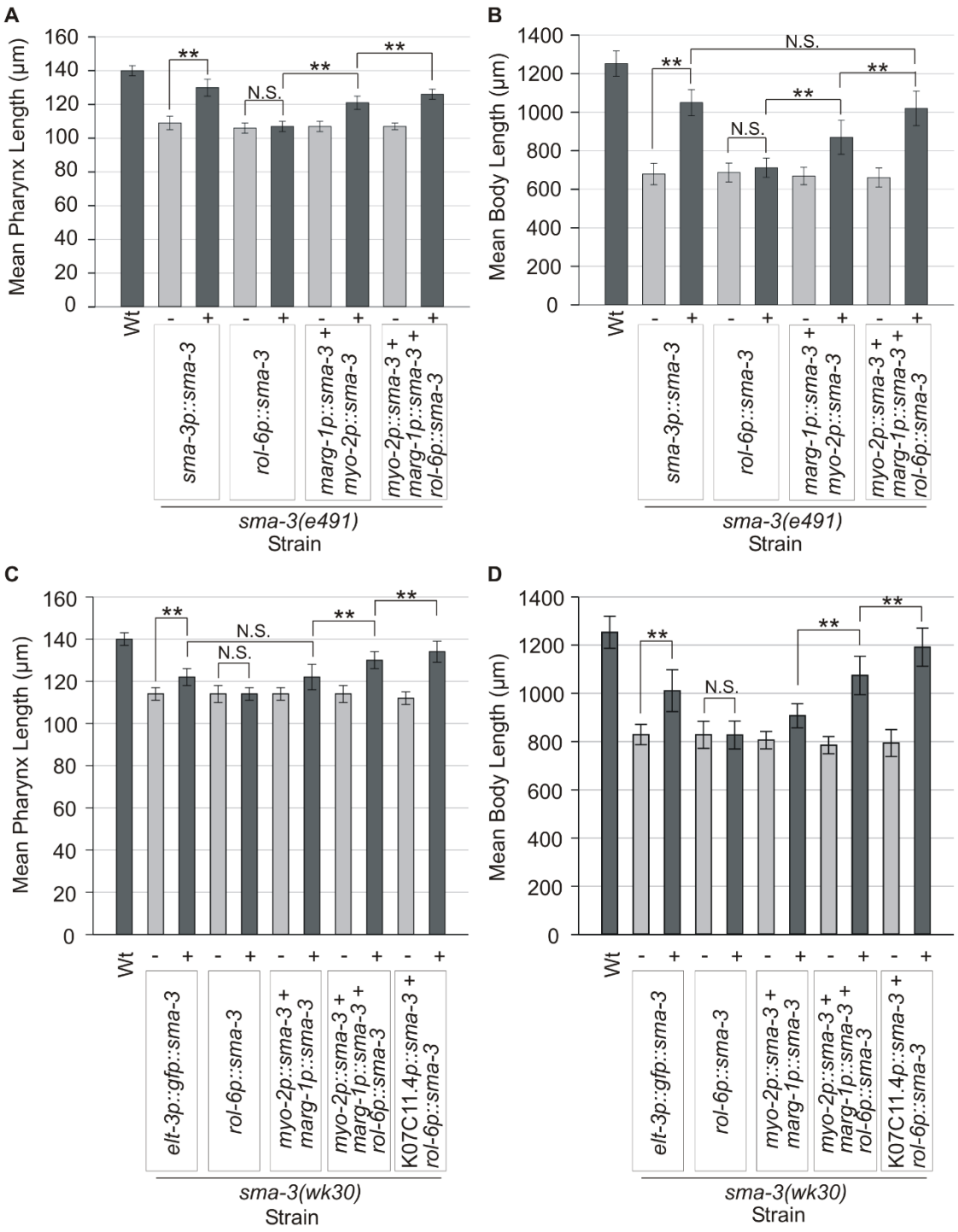
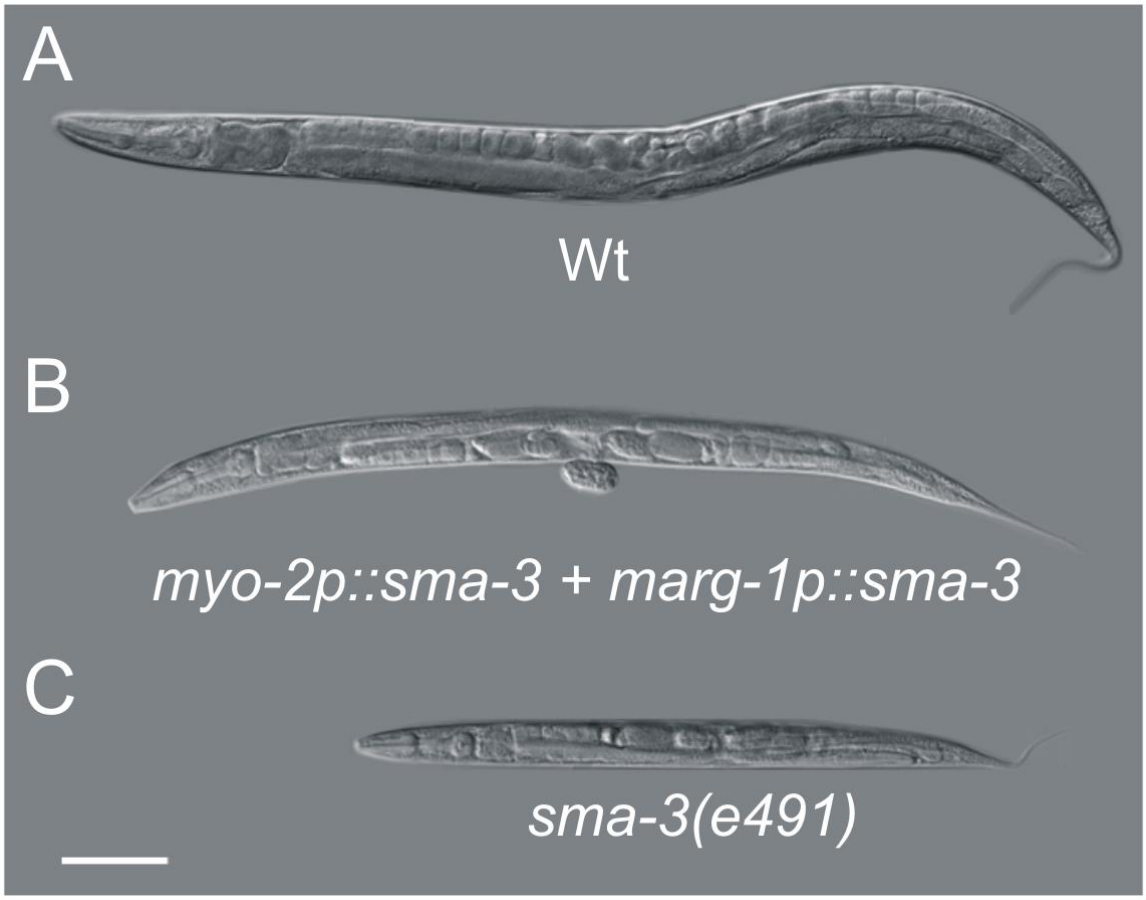

A
Mean Pharynx Length (µm)
160
140
120
100
80
60
40
20
0
**
N.S.
**
**
Wt
- + - + - + - +
sma-3p::sma-3
rol-6p::sma-3
marg-1p::sma-3 + myo-2p::sma-3
myo-2p::sma-3 + marg-1p::sma-3 + rol-6p::sma-3
sma-3(e491)
Strain
B
Mean Body Length (µm)
1400
1200
1000
800
600
400
200
0
N.S.
**
N.S.
**
**
Wt
- + - + - + - +
sma-3p::sma-3
rol-6p::sma-3
marg-1p::sma-3 + myo-2p::sma-3
myo-2p::sma-3 + marg-1p::sma-3 + rol-6p::sma-3
sma-3(e491)
Strain
C
Mean Pharynx Length (µm)
160
140
120
100
80
60
40
20
0
**
N.S.
N.S.
**
**
Wt
- + - + - + - + - + - +
elt-3p::gfp::sma-3
rol-6p::sma-3
myo-2p::sma-3 + marg-1p::sma-3
myo-2p::sma-3 + marg-1p::sma-3 + rol-6p::sma-3
K07C11.4p::sma-3 + rol-6p::sma-3
sma-3(wk30)
Strain
D
Mean Body Length (µm)
1400
1200
1000
800
600
400
200
0
**
N.S.
**
**
Wt
- + - + - + - + - + - +
elt-3p::gfp::sma-3
rol-6p::sma-3
myo-2p::sma-3 + marg-1p::sma-3
myo-2p::sma-3 + marg-1p::sma-3 + rol-6p::sma-3
K07C11.4p::sma-3 + rol-6p::sma-3
sma-3(wk30)
Strain

Figure 2.4. Nomarski differential interference contrast (DIC) images of sodium azide anesthetized 96 ± 1 hrs AEL (A) Wt (N2); (B) *sma-3(e491)* mutant with pharyngeally-expressed *sma-3* minigene *(sma-3(e491); ivEx163[myo-2p::sma-3 marg-1p::sma-3; elt-2p::tdTomato::His2B])*; (C) *sma-3(e491)* mutant non-transgenic sibling of B. Scale bar is 100µm.


A
Wt
B
myo-2p::sma-3 + marg-1p::sma-3
C
sma-3(e491)

pharyngeal transgenes was nonetheless statistically significant compared to their non-transgenic siblings ($p<0.001$, Mann-Whitney Rank Sum Test on raw data). The combination of *myo-2p::sma-3* and *marg-1p::sma-3* transgenes (20 ng/µL injection mix) produced animals with an average pharynx length of 113 ± 5% compared to 100 ± 3% for non-transgenic sibling controls, a significantly greater degree of rescue compared to either alone ($p<0.001$). The simple interpretation of these results is that TGF-β/Sma/Mab signaling acts within the pharynx to control pharynx length, as it also does in the hypodermis. However, we also observed an unexpected rescue of body length by pharyngeal expression of *sma-3* (Figure 2.3B, Supplemental Table 2.3). As with rescue of pharynx length, both the *myo-2p::sma-3* and *marg-1p::sma-3* transgenes showed some rescue of body length individually, while the combination of transgenes (130 ± 16% compared with 100 ± 7% for non-transgenic siblings, 20 ng/µL injection mix) exhibited a greater rescue than either transgene alone ($p<0.01$). In two variations of this rescue experiment, N2 animals, *sma-3(e491)* mutants and rescue strains were imaged in the absence of anesthetic under a dissecting microscope (Supplemental Table 2.3, Supplemental Figure 2.3) and imaged when anesthetized by sodium azide (Figure 2.4, Supplemental Table 2.3). These results also indicate significant but incomplete rescue of body length by pharynx specific *sma-3* minigene constructs. These findings suggest either that body length can be controlled by pharyngeal TGF-β/Sma/Mab signaling, or that our transgenic arrays are active in the hypodermis.

To rule out the possibility that the pharyngeal promoters might be active in the hypodermis, we performed two sets of experiments. First, as described above, we examined the expression of GFP-tagged versions of the transgenes, co-injected with the same transformation markers as in the rescue experiments. Strong expression was observed in the pharynx, however, the exposure was also significantly increased to rule out low levels of ectopic expression. As outlined above, in overexposed images we observed variable weak GFP expression in a few cells adjacent to the pharynx (Figure 2.2D, Supplemental Figure 2.2). Quantitating the intensity in a 16bit black and white file, we detect a ~25-30 fold difference between GFP reporter expression intensity in the pharynx and the faint adjacent cells, which was barely above background intensity. It is also possible that with the high exposure level that this weak signal is reflection of the

pharyngeal signal from the cuticle. We note that our injection mixes did not contain any other hypodermally expressed genes that might have influenced expression of the *sma-3* transgenes such as the hypodermal *rol-6* (pRF4) transformation marker (Kramer et al., 1990; Mello et al., 1991; Sassi et al., 2005). Instead, we used either an intestinal reporter or body wall muscle reporter to identify transgenic animals. A second experiment tested whether the combination of *myo-2* and *marg-1* promoters might be active in the hypodermis by testing rescue of the hypodermal mutant *dpy-7* to further rule out ectopic expression from the pharyngeal promoters.

The *dpy-7* gene encodes a collagen that is expressed in the hypodermis and by its nature is expected to act autonomously in the hypodermis to affect body length (and shape) (Gilleard et al., 1997; Johnstone et al., 1992). We tested whether the combination of the pharyngeal *myo-2* and *marg-1* promoters might be unexpectedly active in hypodermis by seeing if they could drive expression of *dpy-7(+)* to rescue the *dpy-7(e88)* mutant phenotype. The combination of *myo-2p::dpy-7* and *marg-1p::dpy-7* transgenes did not exhibit any rescue of the mutant phenotype (Table 2.1, Supplemental Figure 2.4). Most worms with the *rol-6p::dpy-7* construct appeared rescued; however, 9% displayed an intermediate phenotype between Dpy and wild type suggesting that partial rescue occurred in this case. In contrast, the *dpy-7p::dpy-7* transgene rescued 100% of mutant animals assayed. We therefore conclude from these two lines of evidence that the combination of the *myo-2* and *marg-1* promoters does not significantly activate expression of transgenes in the hypodermis. We propose instead that pharyngeal expression of *sma-3* can partially rescue the small body length of *sma-3* mutants.

2.4.3 TGF-β/Sma/Mab signaling can act in multiple tissues to control overall body length

To further test the tissue specificity of *sma-3* in body length regulation, we performed rescue tests of *sma-3(e491)* mutants using *rol-6p::sma-3* and *dpy-7p::sma-3* transgenes. Importantly, these constructs use the same promoter sequences used to rescue *dpy-7* as a control above, confirming that the promoters are functional in the hypodermis. Surprisingly, we found that neither transgene had rescuing activity in *sma-3(e491)* mutants when injected alone (Figure 2.3A, B, Supplemental Table 2.3, data not shown).

One possibility for the weaker rescue is that the relative dose of the transgene might be too low. We therefore increased the concentration of *rol-6p::sma-3* used in our injection mixes. However, increasing the *rol-6p::sma-3* transgene dose did not enhance rescue of *sma-3(e491)* mutants (Supplemental Table 2.3). We do note that others have also reported lack of rescuing activity with the *rol-6* promoter for LON-2 (but not with another hypodermal promoter), a protein proposed to bind to the DBL-1 ligand to function in the hypodermis as a negative regulator of the TGF-β/Sma/Mab pathway (Gumienny et al., 2007).

Another possibility is that our *rol-6p::sma-3* transgene is not functional or has reduced function. However, we find that our hypodermal *sma-3* transgene is functional when used in combination with the pharynx-expressed *sma-3* transgenes. We created transgenic lines carrying *myo-2p::sma-3, marg-1p::sma-3* and *rol-6p::sma-3* in the *sma-3(e491)* background and compared their pharyngeal and body length rescuing ability to the combination of the pharyngeal promoters alone. We found that this combination of pharyngeal and hypodermal transgenes leads to an average pharynx length of 118 ± 4% compared to 100 ± 2% for non-transgenic siblings, and an average body length of 154 ± 18% compared to 100 ± 7% for non-transgenic siblings (Figure 2.3A, B, Supplemental Table 2.3). This pharyngeal and hypodermal combination of transgenes resulted in significantly greater rescue of both pharynx and body length in *sma-3(e491)* mutants than seen with the pharyngeal transgenes alone ($p<0.001$) (Figure 2.3A, B, Supplemental Table 2.3). Furthermore, the combination of pharyngeal (*myo-2p::sma-3* and *marg-1p::sma-3)* and hypodermal (*rol-6p::sma-3)* transgenes was able to rescue body length of *sma-3(e491)* mutants to the same extent as the native *sma-3p::sma-3* transgene (154 ± 18% and 155 ± 16% respectively of their non-transgenic siblings, $p=0.173$), though the difference in pharynx lengths was still significant ($p<0.001$).

As an additional test of the generality of these results, we performed similar tissue specific rescue experiments with the *sma-3(wk30)* strain as all of the above tests were done with the *e491* allele. We independently assayed hypodermal rescuing activity of three constructs: our own *rol-6p::sma-3* minigene; as well as two constructs (kindly provided by Dr. Cathy Savage-Dunn) that have been reported to rescue *sma-3(wk30)* mutants, *dpy-7p::gfp::sma-3* (pCS226*)* and *elt-3p::gfp::sma-3* (pCS223) (Wang et al.,

2002). Similar to the results observed for the *sma-3(e491)* background, we did not see significant rescue of the pharynx or body length phenotypes of *sma-3(wk30)* with the *rol-6p::sma-3* transgene (Figure 2.3C,D, Supplemental Table 2.3). However, hypodermal expression using the *elt-3p::gfp::sma-3* transgene did significantly rescue pharynx length (107 ± 5% compared to 100 ± 3% for non-transgenic siblings, $p<0.001$) and body length (122 ± 12% compared to 100 ± 5% for non-transgenic siblings, $p<0.001$). Again, we found that the combination of the pharyngeal (*myo-2p::sma-3, marg-1p::sma-3*) and hypodermal (*rol-6p::sma-3*) transgenes provided significantly greater rescue of *sma-3(wk30)* mutants by 96 ± 1 hours after egg laying compared to the pharyngeal transgenes alone (114 ± 5% of non-transgenic sibling pharynx length compared to 107 ± 6% for pharyngeal transgenes alone, $p<0.001$ and 137 ± 12% of non-transgenic sibling body length compared to 113 ± 8% for pharyngeal transgenes alone, $p<0.001$) (Figure 2.3C, D, Supplemental Table 2.3). Thus, our results do not appear to depend strongly on allele-specific effects.

We also tested whether a different pharyngeal promoter could replace the combination of *myo-2* and *marg-1* in rescue of *sma-3(wk30)* mutants. We used the well-characterized pharyngeal promoter from *K07C11.4*, which is active in pharyngeal muscle, marginal cells and epithelial cells, as well as in the intestine and late-stage somatic gonad, but is not expressed in the hypodermis (Gaudet et al., 2004). On its own, K07C11.4*p::sma-3* had little rescuing activity (data not shown). However, the combination of K07C11.4*p::sma-3* and *rol-6p::sma-3* showed significant rescue of both pharynx length (120 ± 5% compared to 100 ± 3% for non-transgenic siblings, $p<0.001$) and body length (150 ± 15% of N2 length compared to 100 ± 7% for non-transgenic siblings, $p<0.001$) (Figure 2.3C,D, Supplemental Table 2.3). This K07C11.4*p::sma-3* and *rol-6p::sma-3* combination had significantly greater pharynx and body length rescue than that observed with the combination of *myo-2p::sma-3, marg-1p::sma-3,* and *rol-6p::sma-3* ($p<0.001$). Thus, we conclude the rescue of body length by pharyngeal *sma-3* minigene constructs is not a unique feature of *myo-2* and *marg-1* pharyngeal promoters.

We draw three conclusions from these results. First, the hypodermal promoter *rol-6* is capable of driving *sma-3* expression and this expression contributes to rescue of *sma-3* mutants, at least when present in an extrachromosomal array in combination with other

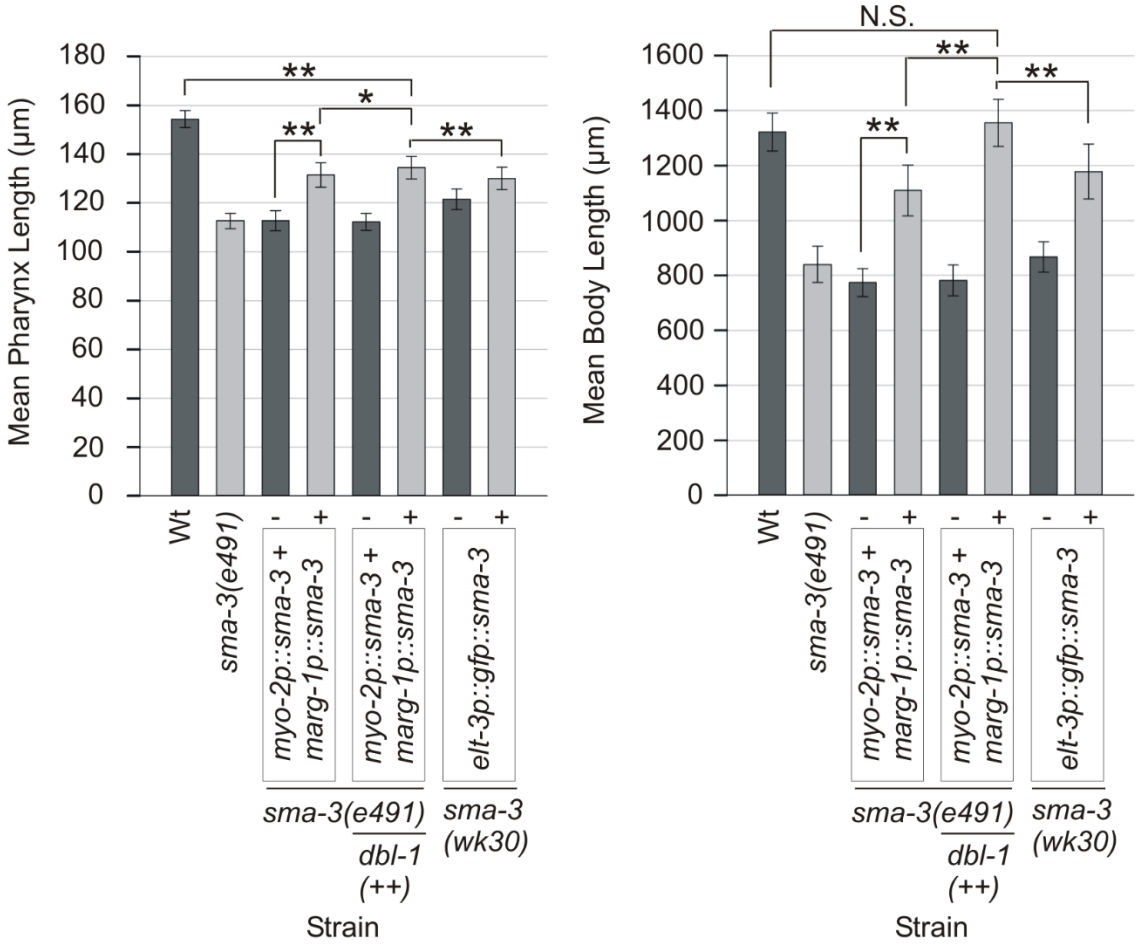
rescue constructs. Second, hypodermal expression of *sma-3* can influence pharyngeal length, just as pharyngeal expression can influence body length. Thirdly, *sma-3* can function in both of these tissues to promote normal growth.

During the course of these experiments it was noticed that some *sma-3* mutant animals were developmentally delayed compared to wild type animals by the time of imaging (96 ± 1 hours after egg laying). We were concerned that this could affect how we interpret the results so we characterized the mean egg-to-egg time for three strains: wild type N2, the *sma-3(e491)* mutant and the pharyngeal *sma-3(e491)* rescue strain *myo-2p::sma-3 + marg-1p::sma-3* (30 ng/µL). As seen in Table 2.2, 100% of wild type animals and 91% of transgenic pharyngeal rescue animals had begun egg laying by 96 hours. In contrast, only 68% of *sma-3(e491)* animals had begun laying eggs by 96 hours. This suggests that we are over-estimating the quantity of rescue with each of our transgenic strains. However, we do note that 35% of wild type worms had begun laying eggs by 72 hrs compared to 0% of the transgenic pharyngeal rescue worms. This result indicates that while these pharyngeal rescue transgenic animals develop somewhat ahead of their non-transgenic mutant siblings, they were also developmentally delayed relative to wild type. This result is consistent with the partial in-between Sma phenotype rescue observed with this strain. To account for this feature of the *sma-3* mutant phenotype, as all worms had begun egg laying by 114 hours, we repeated our length analysis at 120 hours.

We observed partial rescue of both pharynx length and body length with the pharyngeal promoters driving the *sma-3* minigene (*sma-3(e491)* background) as well as the *elt-3::gfp::sma-3* transgene (*sma-3(wk30)* background) (Figure 2.5, Supplemental Table 2.3) which is consistent with our previous results. Furthermore, a statistically significant difference is observed between *sma-3(e491)* mutants at 120 ± 1 hours and pharyngeal rescued transgenic animals at 96 ± 1 hours for both pharynx length (110 ± 5% compared to 100 ± 3%, $p<0.001$) and body length (122 ± 11% compared to 100 ± 8%, $p<0.001$) (Supplemental Table 2.3). These results confirm that pharyngeal signaling is partially rescuing pharynx length and body length in *sma-3(e491)* mutants.

As an extension of these results, we crossed in the *ctIs40[dbl-1(+) sur-5::gfp]* integrated multicopy array [14] to our pharyngeal rescue strain to create JM228:

Figure 2.5. Mean pharynx and body length measurements ± standard deviation of sodium azide anesthetized wild type (Wt) and *sma-3* mutant animals at 120 ± 1 hrs AEL from various *sma-3* minigene rescue experiments. Vertical labels indicate tissue specific promoter-*sma-3* minigene fusion rescue constructs in each strain, except for *elt-3p::gfp::sma-3* which indicates the *elt-3p::gfp::sma-3* transgene (pCS223) was used. In each case, we measured animals carrying the transgenic extrachromosomal array (+) and siblings that lacked the array (-), as indicated by a transformation marker (either *elt-2p::tdTomato::His2B* or *myo-3p::wCherry*). Horizontal labels indicate genetic background. All transgenic animal means (+) were statistically significantly different from non-transgenic sibling means (-) ($p<0.001$). ** denotes significant differences of $p<0.001$, * denotes significant differences of $p<0.05$.


Mean Pharynx Length (µm)
180
160
140
120
100
80
60
40
20
0
**
*
**
**
Wt
sma-3(e491)
myo-2p::sma-3 + marg-1p::sma-3
myo-2p::sma-3 + marg-1p::sma-3
elt-3p::gfp::sma-3
- + - + - +
sma-3(e491) dbl-1 (++)
sma-3 (wk30)
Strain
Mean Body Length (µm)
1600
1400
1200
1000
800
600
400
200
0
N.S.
**
**
**
Wt
sma-3(e491)
myo-2p::sma-3 + marg-1p::sma-3
myo-2p::sma-3 + marg-1p::sma-3
elt-3p::gfp::sma-3
- + - + - +
sma-3(e491) dbl-1 (++)
sma-3 (wk30)
Strain

ctIs40[dbl-1(+) sur-5::gfp] X; sma-3(e491) III; ivEx163[myo-2p::sma-3 marg-1p::sma-3 phat-1p::yfp elt-2p::tdTomato::His2B]. In a wild type background, the *ctIs40* insertion results in a Lon phenotype due to the overexpression of the TGF-β/Sma/Mab pathway ligand *dbl-1* (Suzuki et al., 1999). However, in the newly constructed strain, the downstream *sma-3(e491)* mutation should be epistatic to *dbl-1* overexpression. As expected, all worms with the *dbl-1(+)* array that did not get the extra chromosomal pharyngeal *sma-3* rescue array (as determined by the lack of extra chromosomal array reporter expression) had small mean pharynx length and body length (Figure 2.5, Supplemental Table 2.3). Surprisingly, the animals that carried the extra chromosomal pharyngeal *sma-3* rescue array and *dbl-1(+)* array insertion had fully rescued body length compared to wild type ($p=0.132$) but not pharynx length ($p<0.001$) (Figure 2.5, Supplemental Table 2.3). The increase in rescue that is observed with the elevated dose of *dbl-1* ligand validates our conclusion that it is TGF-β/Sma/Mab pathway signaling in the pharynx that is rescuing body length in these animals and not an independent function of SMA-3. We conclude from this experiment that pharynx length is regulated by TGF-β/Sma/Mab pathway signaling from both within and outside the pharynx.

Expression of *sma-3* in multiple tissues might rescue *sma-3(e491)* and *sma-3(wk30)* mutants for two reasons. Each tissue could make a distinct contribution to growth, for example, by expressing distinct secreted signals that act on different downstream components. Alternatively, the effect on rescue could be quantitative rather than qualitative whereby simply increasing the dose of *sma-3* in a single tissue may be sufficient to increase rescue. As noted above, increasing the relative concentration of hypodermal *sma-3* transgenes had no observable effect on rescue of *sma-3(e491)* mutants. Likewise, reducing the dose of *sma-3* in each of three tissues (pharyngeal muscle, pharyngeal marginal cells and hypodermis) only slightly reduced rescuing activity (Supplemental Figures 2.1, 2.5, Supplemental Tables 2.1, 2.3). While we cannot interpret whether this effect reflects a qualitative difference (e.g. tissue-specific targets) or quantitative difference (amount of downstream signal produced by the tissues) or a combination of both, the results support the conclusion that *sma-3* can act in multiple tissues to control growth.

2.5 Discussion

2.5.1 The TGF-β/Sma/Mab pathway regulates pharynx length

In a previous study on the morphology of the pharynx, we identified a number of Sma mutants with decreased pharynx lengths compared to wild type, including members of the TGF-β/Sma/Mab signaling pathway *sma*-2 and *sma-3* (Raharjo et al., 2011). Here we report that mutants of TGF-β/Sma/Mab pathway components have significantly smaller pharynx lengths compared to Dpy mutants of similar body length. The pharynges of Dpy mutants are also significantly smaller than wild type suggesting that growth of this organ is limited by the length of the entire animal. Thus, pharynx length appears to be determined partly by positive TGF-β/Sma/Mab signaling but is also dependent on the overall body length of the animal. It was interesting that signaling in the pharynx never resulted in full rescue of pharynx length even when body length was completely rescued. This result implies that TGF-β/Sma/Mab signaling from outside the pharynx must also play a role in determining pharynx length.

It is important to note the similarity in growth control of the pharynx and the hypodermis, where both organs utilize positive TGF-β/Sma/Mab signaling to regulate organ and overall body length. Furthermore, pharynx and body length can be restricted by morphology defects in components of the surrounding extracellular cuticle. Previous study of mutations in extracellular matrix (ECM) components (including the cuticle collagen *dpy-7*) and membrane proteins revealed a role in pharynx morphology and a twisted pharynx (Twp) phenotype (Axäng et al., 2007; Jafari et al., 2010). The authors observed bending of contractile arrays in pharyngeal muscles and hypothesized that during normal pharyngeal growth, the defective ECM must restrict these arrays. It is unclear if Twp animals have reduced pharynx lengths but if so it would be a noteworthy parallel between pharynx growth and overall body growth. It would be interesting to see if there are any common ECM or membrane components as downstream targets of pharyngeal and hypodermal TGF-β/Sma/Mab pathway signaling.

The hypodermis undergoes post-embryonic endoreduplication and it has been suggested that a primary function of TGF-β/Sma/Mab pathway signaling is to positively regulate this event to achieve body size regulation (Flemming et al., 2000). On the other

hand, analysis of potential downstream targets of the pathway did not reveal a large number of cell cycle regulators (Roberts et al., 2010) and it is still not known which direct downstream targets of the TGF-β/Sma/Mab pathway are most critical for regulating body size. In contrast, the pharynx does not undergo any cell number or ploidy changes during post embryonic development (Sulston et al., 1983). It would be interesting to see if this pathway mediates cell and tissue growth via different downstream targets in the pharynx vs the hypodermis.

2.5.2 Activity of the TGF-β/Sma/Mab pathway in multiple tissues can control body length

The results described here support a model in which downstream effectors of TGF-β/Sma/Mab signaling can act in both the pharynx and hypodermis to influence overall body length in *C. elegans*. Multiple lines of evidence support this model. First, expression of *sma-3* under the control of different pharyngeal promoters (either *myo-2* or *marg-1* alone or in combination) is able to partially rescue body length of *sma-3* mutants. These promoters are not active in the hypodermis, based on the lack of expression of GFP reporters and inability to rescue the hypodermal collagen mutant *dpy-7(e88)*. Second, we find that simultaneous expression of *sma-3* in both the hypodermis and pharynx provides stronger rescue of body length than when *sma-3* is expressed in either tissue alone suggesting that both can contribute to normal growth in an additive manner. Third, overexpression of the *dbl-1* ligand with only pharyngeal signaling results in complete rescue of body length indicating that while likely insufficient in wild type situations, pharyngeal signaling is capable of regulating body length. Finally, components of the TGF-β/Sma/Mab pathway, including SMA-3, are expressed in the pharynx and hypodermis, consistent with the proposed function of this pathway in these tissues. Interestingly, at least one other TGF-β/Sma/Mab component has been demonstrated to act in the pharynx. Body length (and width) of *daf-4* mutants, the Type II TGF-β receptor, can be rescued by expression of *daf-4* in pharyngeal muscle under the control of the *myo-2* promoter (Inoue and Thomas, 2000), consistent with our findings for *sma-3*.

Our results differ from previous reports that hypodermal expression of TGF-β/Sma/Mab components is sufficient for body length rescue (Gumienny et al., 2010;

Wang et al., 2002; Yoshida et al., 2001). This difference may only represent variation in the degree of rescue, as we do observe partial rescue with the hypodermal *elt-3p::gfp::sma-3* transgene, though not to the extent previously reported. It is possible that these disparities may also reflect differences in generation of transgenic lines. Likewise, hypodermal expression of other components (*sma-6, sma-10* and *drag-1*) is sufficient for complete rescue (Gumienny et al., 2010; Tian et al., 2010; Yoshida et al., 2001); though this does not preclude contributions from other tissues. We did not observe any rescue with the *rol-6p::sma-3* and *dpy-7p::sma-3* transgenes when injected alone, however, cuticle collagen gene expression (including *rol-6*) is known to cycle in relation to molts (Park and Kramer, 1994) and as such the *rol-6p* and *dpy-7p* hypodermal promoters may not drive sufficient expression of the rescuing transgene at necessary times in development to achieve rescue. Additionally, the *dpy-7p::sma-3* transgene appeared to have integrated into the genome. The lack of rescue observed in this strain may be explained by an integration event that disrupted an important gene as this strain appeared sick relative to the other transgenic strains generated. Previous work demonstrated that a *sma-3* transgene under control of the *myo-2* promoter weakly but significantly rescued growth of *sma-3(wk30)* mutants, similar to what we observe (Wang et al., 2002). However, no tests were performed with the combination of *myo-2* and other pharyngeal promoters, i.e. simultaneous expression in both marginal cells and muscles was not tested as we did here, which resulted in an obvious partial rescue of body length.

The rescuing activity of pharyngeal and hypodermal promoters when used in combination was consistently more robust than either promoter alone. In the *sma-3(e491)* background, this combination of transgenes rescued body length to the same extent as the native *sma-3* promoter construct suggesting that both tissues can contribute to body length regulation. Interestingly, the *K07C11.4* promoter, which drives expression in the pharynx and late stage somatic gonad did not rescue by itself but in combination with the *rol-6p::sma-3* transgene almost completely rescued the *sma-3(wk30)* body length phenotype to N2 levels. We do note that complete rescue of the *sma-3* body length phenotype to N2 length was only observed when the *dbl-1* ligand was over expressed. It is unclear what underlies the differences in rescue activity of the various extrachromosomal arrays tested here. One source of variation could be the copy number

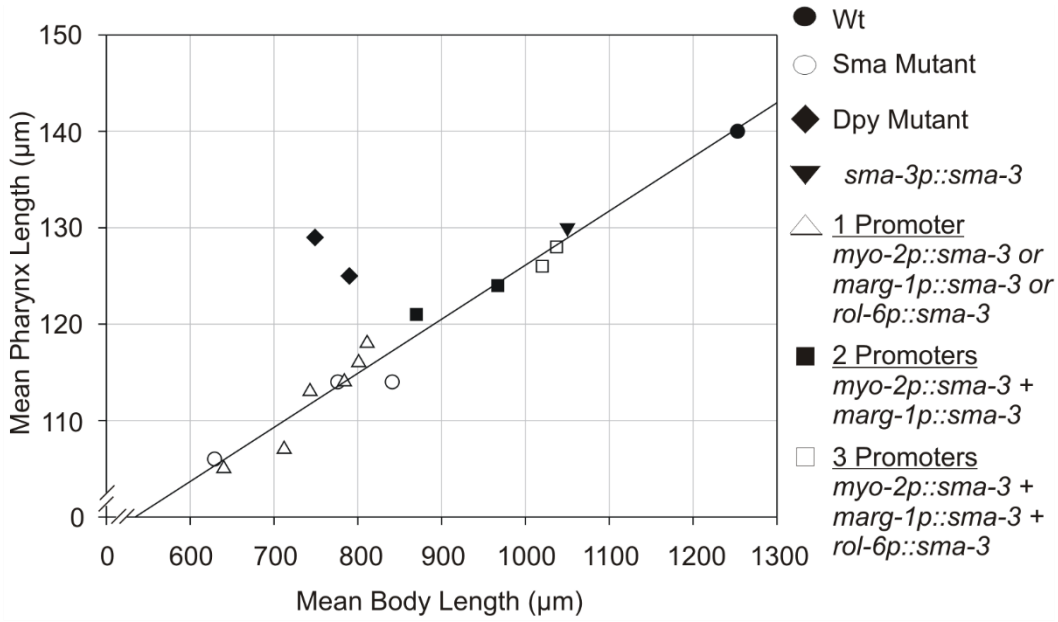
of each transgene present in each extrachromosomal array and the level of expression of the *sma-3* minigene from each array. Furthermore, the neomorphic properties of the *sma-3(e491)* allele may interfere with the rescuing activity of the *sma-3* minigene. This could account for the observation of almost complete rescue of body length to N2 levels with the *K07C11.4p::sma-3* and *rol-6p::sma-3* transgene combination in the *sma-3(wk30)* background but the lack of complete rescue with the *sma-3p::sma-3* transgene in the *sma-3(e491)* background.

2.6 Conclusions

Given the developmental delay observed in animals homozygous for *sma-3* mutant alleles and the artificial nature of *C. elegans* extrachromosomal arrays, it is difficult to quantitate exactly how much contribution TGF-β/Sma/Mab signaling in the pharynx makes to regulation of body length. Certainly the only case where full rescue of body length was achieved was with high levels (presumed to be greatly in excess of wild type levels) of the *dbl-1* ligand present. Taken together, using our results presented here and those of previous studies on tissue specific regulation of body length (Inoue and Thomas, 2000; Wang et al., 2002; Yoshida et al., 2001), we make the following conclusions about pharynx length and body length regulation by TGF-β/Sma/Mab signaling. TGF-β/Sma/Mab signaling in the pharynx is capable of contributing to pharynx length and body length regulation but this signaling is not sufficient or necessary to facilitate wild type pharynx length or body length.

Coordination of growth is an interesting feature of all of the animals examined here. We note a strong linear correlation between pharynx and body length, except in Dpy mutants, in which signaling is presumably normal but body length is reduced due to defects in hypodermal collagen (Figure 2.6). Consistent with signaling between tissues, animals in which manipulation of TGF-β/Sma/Mab signaling results in uncoupled hypodermal and pharynx lengths were not observed (i.e. we have not seen small animals with big pharynges or vice versa). Furthermore, organs of Dpy mutants such as the pharynx and gonad often appear compressed compared to those of TGF-β/Sma/Mab pathway mutants which appear more proportional to the overall body length of the animal

Figure 2.6. Linear relationship between pharynx length and body length in mutants of the TGF-β/Sma/Mab pathway and strains carrying various *sma-3* rescuing extrachromosomal arrays in the *sma-3(e491)* mutant background. One promoter refers to cases where only a single *sma-3* minigene rescue construct was used (either *myo-2p::sma-3*, *marg-1p::sma-3*, or *rol-6p::sma-3*). Two and three promoters refer to strains carrying all indicated transgenes. Wild type (Wt) N2 is included for comparison. The correlation coefficient for the linear model is $R^2 = 0.97$. Notably, *dpy-5* and *dpy-10* mutants (black diamonds; not included in the linear model) do not fall on the line.


Mean Pharynx Length (µm)
150
140
130
120
110
0
0
600
700
800
900
1000
1100
1200
1300
Mean Body Length (µm)
Wt
Sma Mutant
Dpy Mutant
sma-3p::sma-3
1 Promoter
myo-2p::sma-3 or
marg-1p::sma-3 or
rol-6p::sma-3
2 Promoters
myo-2p::sma-3 +
marg-1p::sma-3
3 Promoters
myo-2p::sma-3 +
marg-1p::sma-3 +
rol-6p::sma-3

(Brenner, 1974). This suggests that TGF-β/Sma/Mab signaling in the pharynx and hypodermis may be coordinating growth of many tissues within the animal. Future directions should focus on identifying what signaling molecules could be functioning downstream of this pathway to mediate proportional growth between the tissues of *C. elegans*. Pursuit of these molecules may ultimately lead to a deeper understanding of the role of TGF-β Sma/Mab signaling growth control. This pathway has recently been shown to function in a non-canonical manner to regulate pharyngeal gland cell morphology (Ramakrishnan et al., 2014), a finding completely un-predicted from the current understanding of this pathway's role in *C. elegans* biology. Understanding all of the functions of the TGF-β signaling pathway in the pharynx will be critical to put together what appears to be a complex role in growth regulation in *C. elegans.*

Chapter 3: *pho-1* Spatial Patterning in the *C. elegans* Intestine is Regulated Indirectly by the DNA Binding Protein LIN-14

3.1 Abstract

The regulation of gene expression patterns, both during development and in response to the environment, is a basic aspect of biology. Gene expression can be controlled at many levels, the most central of which is at the level of transcription. The goal of this research is to understand the molecular mechanisms that control transcription in the *C. elegans* intestine. The mature *C. elegans* intestine is composed of 20 cells, arranged in nine intestinal rings (int-I to IX) along the anterior/posterior (A/P) axis of the animal. All intestinal cells are clonally derived from a single progenitor cell (E), which is specified by the redundant GATA transcription factors END-1 and END-3. At the 2-4E cell stage, END-1 and END-3 activate expression of the GATA factors ELT-2 and ELT-7, which drive expression of terminally differentiated genes in the intestine throughout the life of the animal. Some intestine specific genes are expressed only in a subset of the ints; for example, the acid phosphatase encoded by *pho-1* is expressed only in the posterior 14 cells of the intestine (int-III to IX), even though *pho-1* is controlled by ELT-2. Such observations lead to the question: how does gene expression in the intestine become spatially patterned? ELT-2 is expressed uniformly in all cells of the intestine, suggesting that patterned genes could be regulated by a combination of ELT-2 with some other *trans*-acting factor(s). Previous studies (Fukushige et al., 2005; Huang et al., 2007) have observed that the Wnt/β-catenin asymmetry pathway is necessary for repressing *pho-1* expression in the anterior six cells (int-I and II), as well as specifying the anterior fate of these cells. At the moment, it is unclear if spatial patterning of *pho-1* expression by this pathway occurs at the level of cell fate or by direct action on the *pho-1* promoter. A previous student used RNAi and promoter analysis to conclude that the heterochronic gene *lin-14* represses the anterior expression of *pho-1* (Yan, 2007). LIN-14 is expressed in the intestine (and other tissues) from late embryogenesis until the end of the L1 larval stage (Ambros and Horvitz, 1984). LIN-14 is required for the binucleation of int-III to VII (and sometimes VIII and IX) at the end of the L1 stage (Ambros and Horvitz, 1984). Here, I repeat the experiments using RNAi, *lin-14* mutants, promoter bashing and careful quantitation of reporter expression including how *pho-1* patterning forms. This analysis

revealed that the *pho-1* spatial expression pattern is generated after hatching by the end of the L1 larval stage in correlation with the patterned binucleation that is occurring in the intestine at this time. RNAi mediated knockdown of *lin-14* function results in increased *pho-1* reporter expression in the anterior six cells. Assessment of the *pho-1* reporter in the *lin-14(n179)* temperature-sensitive mutants shows that LIN-14 is required for *pho-1* repression in the anterior intestine. Repression of anterior *pho-1* expression by LIN-14 is not direct despite three putative LIN-14 binding sites in the *pho-1* promoter. In summary, the results suggest a novel role for the heterochronic gene *lin-14* in spatial patterning of gene expression in the intestine.

3.2 Background

3.2.1 Overview

During embryogenesis, an organism is engaged in the process of constructing tissues that must be capable of performing specific and often vital functions. This process of cell fate specification requires coordinating and instructing certain cells to express all the necessary genes required to create the correct morphology and carry out the specific molecular functions of a particular organ. Instruction can come in the form of signals from other cells, as well as from within the cells themselves based on their lineage. This information must be coordinated spatially and temporally to ensure that genes are expressed in the proper cells of the organ and at the proper time.

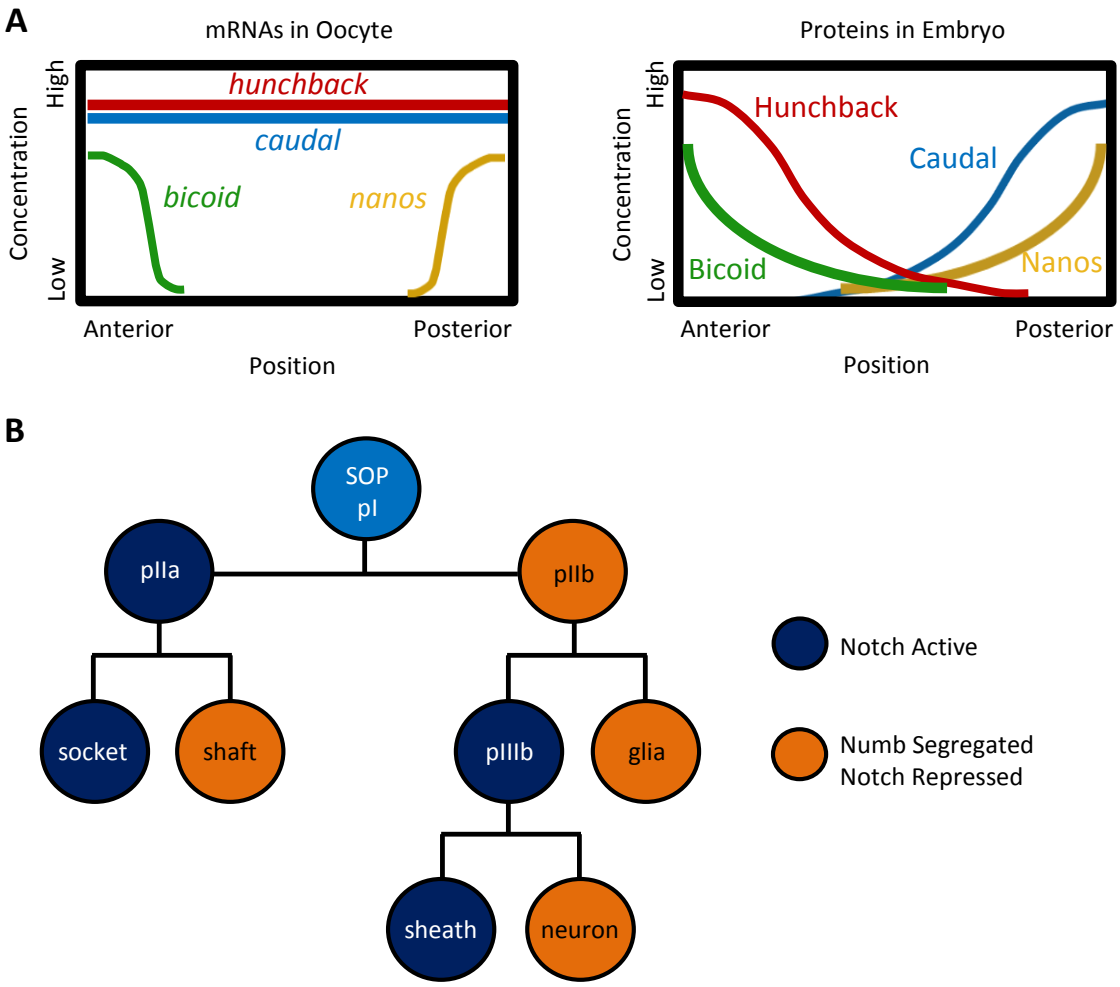
There are two major mechanisms that are used to regulate cell fate: autonomous specification and conditional (non-autonomous) specification. Autonomous specification refers to a case where the fate of a cell is established by the transcription factors that are passed down within a lineage and/or expressed within the cell itself. In contrast, conditional or non-autonomous specification is the determination of cell fate by extrinsic factors, typically in the form of cell-cell signaling. Both of these mechanisms can be used separately and in tandem to regulate cell fate decisions in organisms. How cell fate is regulated, and how this in turn regulates gene expression in different spatial and temporal patterns is an interesting biological problem, the solutions to which will have

implications for understanding the critical steps in various developmental disorders in humans.

One widespread system used in conditional specification to create spatial patterns of gene expression is the production of morphogen or signal gradients. Morphogens are molecules that are secreted from a source, depleted by a sink, and therefore become expressed in gradients (Nahmad and Lander, 2011). A morphogen is capable of activating different target genes at different concentrations to regulate multiple cell fates. During development, some cells are exposed to high levels of a morphogen/signal while others are exposed to an intermediate level, and still others receive little or none at all. Target genes are then activated in cells depending on whether the levels of morphogen/signal are higher than threshold requirements. For example, Drosophila embryos require the action of two gradients, *bicoid* from the anterior and *nanos* from the posterior, to produce specific spatial patterns of gene expression along the A/P axis (Driever and Nüsslein-Volhard, 1988; Lehmann and Nüsslein-Volhard, 1991; Nüsslein-Volhard and Wieschaus, 1980). The mRNA of these genes is maternally deposited into the oocyte and tethered to the cytoskeleton at the respective anterior and posterior ends (Figure 3.1A) (Johnstone and Lasko, 2001). When the proteins are translated after fertilization, they diffuse from the poles of the egg creating localized gradients. Bicoid represses translation of *caudal* and *nanos* in the anterior, while Nanos represses translation of *hunchback* in the posterior, resulting in an anterior gradient of Hunchback and a posterior gradient of Caudal. In the anterior, high levels of the Bicoid and Hunchback transcription factors leads to specification of head, while lower levels are required for thorax formation. In the posterior, Caudal is required for the specification of the abdomen.

Another common mechanism by which cells become patterned to express different genes is via asymmetric divisions within a cell lineage, resulting in daughter cells with different fates (Horvitz and Herskowitz, 1992). These types of divisions rely upon unequal segregation of cell fate determining factors to the daughter cells via proper orientation of the mitotic spindle (Rhyu and Knoblich, 1995). In the Drosophila mechanosensory bristle lineage, for example, a repeated series of asymmetric divisions coordinated in part by *delta/notch* signaling results in a socket cell, shaft cell, sheath cell, neuron and glia, each of which express a different set of terminal genes

Figure 3.1. A. A/P axis formation in Drosophila is mediated by mRNA and protein gradients of *bicoid*, *nanos*, *hunchback*, and *caudal* genes, (adapted from Johnstone and Lasko, 2001). B. The sensory bristle lineage consists of multiple delta-notch regulated asymmetric divisions (adapted from Rebeiz et al. 2011).


A
mRNAs in Oocyte
Concentration
High
Low
hunchback
caudal
bicoid
nanos
Anterior
Posterior
Position
Proteins in Embryo
Concentration
High
Low
Hunchback
Caudal
Bicoid
Nanos
Anterior
Posterior
Position
B
SOP
pI
pIIa
pIIb
socket
shaft
pIIIb
glia
sheath
neuron
Notch Active
Numb Segregated
Notch Repressed

(Gho et al., 1999; Muskavitch, 1994). The mechanosensory bristle lineage is an example of the integration of both autonomous and conditional specification systems to regulate cell fate specification (Frise et al., 1996; Guo et al., 1996; Hartenstein and Posakony, 1990; Posakony, 1994; Rhyu et al., 1994). The sensory organ precursor cell (SOP) undergoes an asymmetric cell division along the A/P axis to produce the posterior daughter cell pIIa and the anterior daughter pIIb (Figure 3.1B). The orientation of this division is controlled by activity of the Wnt receptor Frizzled (Gho et al., 1999). Notch signaling is active in pIIa but repressed by Numb, which is asymmetrically sequestered to pIIb, creating Notch dependent (pIIa) and Notch independent (pIIb) cell fates. The pIIa cell also produces and asymmetrically segregates Numb to one daughter cell resulting in socket cell specification by Notch activity and shaft cell fate in the Notch repressed daughter. Similarly, pIIb asymmetrically localizes Numb to one of its daughters leading to specification of the glial cell, while the Notch active daughter undergoes one more asymmetric division producing the sheath (Notch dependent) and neuron (Notch independent) cells (Gho et al., 1999).

Development of *C. elegans* is mediated by both autonomous specification events and conditional specification to determine cell fates. The combination of lineage specific factors and cell signaling events regulate spatial and temporal gene expression patterns during development. The endoderm is an example of a lineage in this nematode that uses intrinsic transcription factors as well as inductive signaling to specify cell fate. The intestine is ideal to use to identify how spatial patterning is controlled because this simple organ is clonally derived from a single blastomere, the E cell (Sulston et al., 1983) but not all intestinal cells express the same terminal differentiation genes. Given that the specification and development of the intestine, including the core transcriptional hierarchy, has already been elucidated, it is an appropriate model to study the mechanisms that direct spatial patterns of gene expression.

3.2.2 Anatomy and Function of the *C. elegans* Intestine

The intestine of *C. elegans* is essentially a long cylindrical tube running along the A/P axis of the worm, attached to the pharyngeal valve cells in the anterior and the rectal valve cells in the posterior. This organ is typically composed of 20 epithelial cells

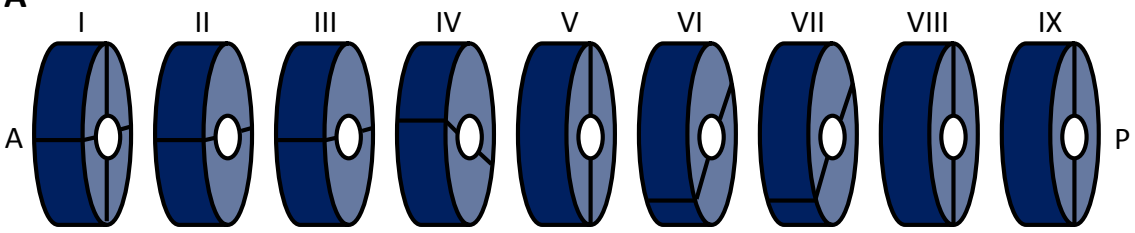
arranged into nine intestinal rings (ints) around a central lumen, although a sporadic extra division can result in 21 cells (Leung et al., 1999; Sulston et al., 1983) (Figure 3.2). Int-I is the most anterior ring and is formed of four cells, while all other intestinal rings consist of two cells each and are thus bilaterally symmetrical (Figure 3.2A). All intestinal cells have a brush border of microvilli on their apical lumenal side, are connected laterally via adherens junctions and gap junctions and secrete a basal lamina (Sulston and Horvitz, 1977). The microvilli of the anterior four cells are about half the length of those associated with posterior cells and the gut lumen is more expanded (Sulston et al., 1983).

At the end of embryogenesis, the intestine is composed of 20-21 cells and this number does not change during larval or adult stages. During the L1 molt, cells of int-III through int-VII and occasionally int-VIII and int-IX undergo binucleation (Sulston and Horvitz, 1977). Furthermore, every intestinal cell undergoes endoreduplication at each larval molt, which results in an adult organ typically composed of 20 intestinal cells and a total of 30-34 nuclei, each with a ploidy of 32n (Hedgecock and White, 1985).

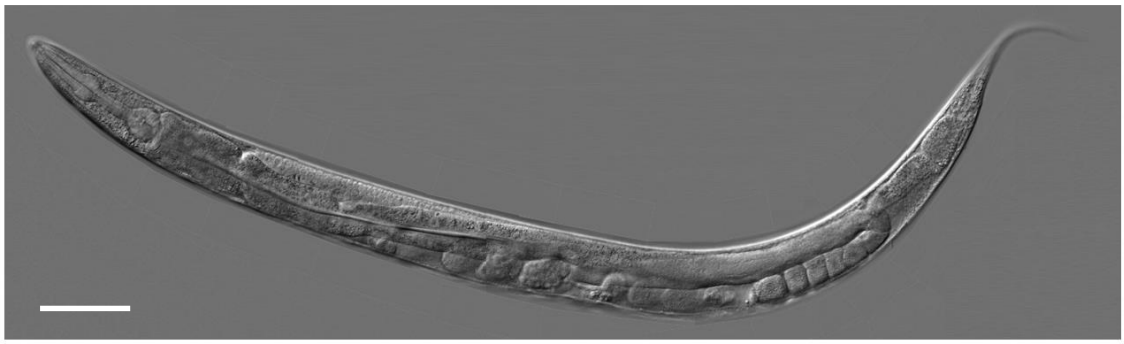
The main function of the *C. elegans* intestine is to digest and absorb nutrients from ingested bacterial cells. The intestine produces and secretes many enzymes into the lumen as well as assembling membrane bound proteins involved in nutrient transport (McGhee, 2007). The epithelial cells of the intestine are the site of synthesis and storage of macromolecules such as yolk for other cells in the worm (Kimble and Sharrock, 1983). These two core functions of digestion and storage are potentially separated along the A/P axis as expression of some enzymes are restricted to anterior cells while yolk and lipid vacuoles are more prominent in the posterior (Britton et al., 1998; McGhee, 2007). Also, it has been shown that there are more organelles and membrane bound vacuoles present in int-I and int-II compared to posterior rings (Borgonie et al., 1995; McGhee, 2007). In addition to the primary function, an often overlooked but vital secondary function of the intestine is to act as a first line of defense in the immune system. The epithelial cells actively prevent and respond to infections from ingested bacteria and viruses as *C. elegans* does not have typical mobile immune cells (Pukkila-Worley and Ausubel, 2012).

Figure 3.2. A. The 20 cells of the *C. elegans* intestine, arranged in nine rings, each with 2 cells except for the 4 found in int-I. A refers to anterior and P to posterior. B. Composite DIC image of a Wt worm.

A



B

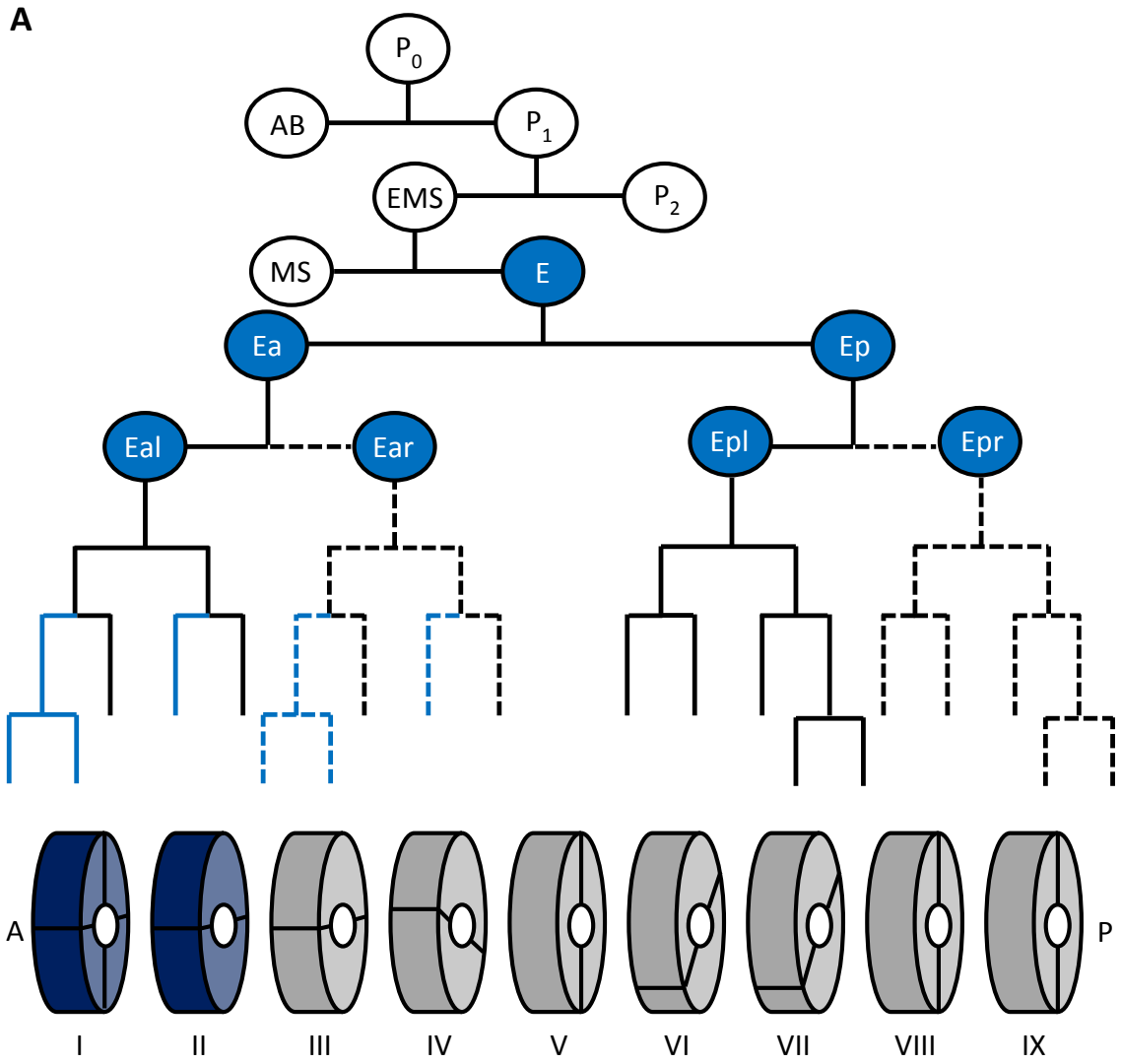


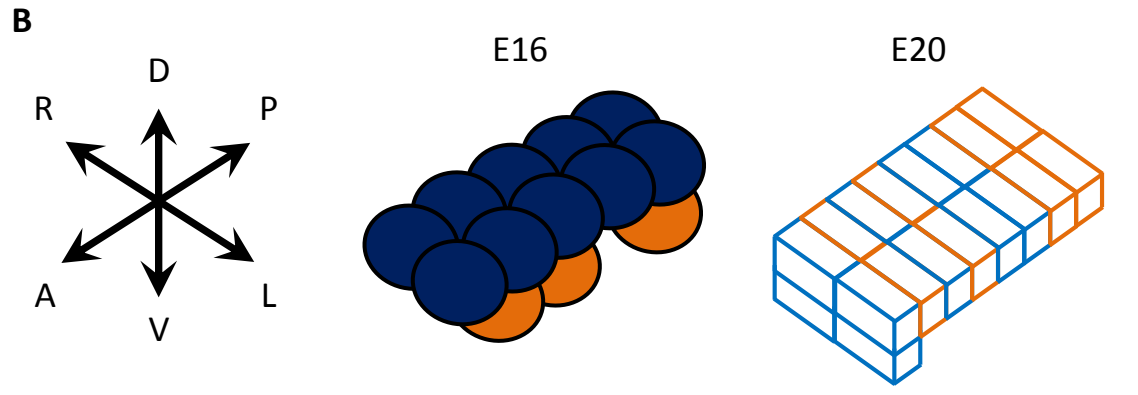
3.2.3 Morphogenesis of the *C. elegans* Intestine

As stated previously, all 20 intestinal cells arise from a single blastomere during embryogenesis, the E cell (Figure 3.3A). The progressive stages of intestinal development refer to the number of E cell descendants present (E2, E4 etc). The E progenitor divides along the A/P axis to produce two E cells (Ea and Ep) that lead migration into the embryo during gastrulation (Leung et al., 1999). Once internal in the embryo, Ea and Ep each undergo a left/right division resulting in the 4E cell stage (Leung et al., 1999). The next two divisions (3rd and 4th) that produce 8E and 16E cells are in the A/P axis plane, although some of the 4th divisions are skewed slightly to the dorsal/ventral (D/V) axis in the embryo (Leung et al., 1999). At the 16E cell stage, there are ten E cells located dorsally, and the other six found ventrally (Figure 3.3B). The left/right pair of cells most anterior/dorsal and the left/right pair most posterior/ventral all undergo one more round of division resulting in 20E cells (Leung et al., 1999). Occasionally one of the left/right pair of posterior and dorsal most cells (int-VII) at the 16E cell stage will irregularly divide during embryogenesis leading to 21E cells (Leung et al., 1999; Sulston et al., 1983).

It is during the 16E cell stage that these cells begin to show signs of cellular and epithelial polarization (Leung et al., 1999). The nuclei of both the dorsal and ventral rows of cells begin to move towards what will eventually be the apical lumen, at the midline between the left and right pairs of cells. At the same time, cytoplasmic organelles, lipid droplets and yolk move towards what will become the basal surface opposite the lumen and the cells elongate along this left/right axis (Leung et al., 1999). Shortly after cellular polarization, lumen formation begins at the midline between adjacent left/right pairs of cells with gaps between the membranes of these cells indicating cell separation. Two intercalation events, one at 16E and one at 20E create a single tube of bilaterally symmetrical left/right pairs of cells (with the exception of int-I which is radially symmetrical) out of the dorsal and ventral layers of cells observed in the early16E stage (Leung et al., 1999) (Figure 3.3B). Cells from the left rows only intercalate with cells on the left and similarly cells from the right rows intercalate only with right located cells. After intercalation is complete, the formation of adherens junction complexes around the lumen is observed indicating epithelial polarization.

Figure 3.3. A. The clonal endoderm lineage in *C. elegans* has three A/P asymmetric divisions, the first, third and fourth. B. At the E16 cell stage, there are 10 dorsally located cells and 6 ventrally located cells, arranged in left/right pairs along the A/P axis. These cells intercalate as the anterior/ dorsal most and posterior/ventral most cells divide to produce 20 cells in a single epithelial layer.


A
P0
AB
P1
EMS
P2
MS
E
Ea
Ep
Eal
Ear
Epl
Epr
A
P
I
II
III
IV
V
VI
VII
VIII
IX


B
D
R
P
A
L
V
E16
E20

Another set of cell movements occurs during the E16 to E20 stages where the cells of int-II, int-III and int-IV perform asymmetrical circumferential cell rotations across the plane of bilateral symmetry (Hermann et al., 2000; Leung et al., 1999). This produces a twist of between 90^{o} to 180^{o} in the intestine, a feature that later influences morphogenesis of the somatic gonad (Hermann et al., 2000; Leung et al., 1999; Sulston et al., 1983). When viewed from the posterior, the cells on the right of these ints migrate circumferentially around the lumen of the intestine in a counterclockwise direction, dorsally and to the left (Hermann et al., 2000). The cells on the left also move counterclockwise (as viewed from P), ventrally and to the right. The cell migrations of the intestinal twist are under control of Notch signaling, specifically the LAG-2 and APX-1 delta related ligands signaling through the LIN-12/Notch receptor and the LAG-1/Suppressor of Hairless transcription factor (Hermann et al., 2000). Asymmetric LIN-12 expression between the left and right rows are established by a LAG-2-LIN-12 mediated reduction of LIN-12 expression only in the left cells at the 4E cell stage. This leads to expression of the basic helix-loop-helix transcription factor *ref-1* in the left row of cells during the 4E and 8E stages (Neves and Priess, 2005). Expression of *ref-1* in these intestinal cells also depends upon ELT-2 which physically binds to LAG-1 (Neves et al., 2007). Later at the 16E stage, an APX-1-LIN-12 notch signaling event drives expression of *ref-1* in the right cells. Both Notch signaling events and REF-1 expression is required for the intestinal twist movements to occur (Hermann et al., 2000; Neves and Priess, 2005). Additionally, the restriction of this morphogenic movement to the anterior ints-II through IV is controlled by the non-canonical Wnt/β-catenin asymmetry pathway (Schroeder and McGhee, 1998). This pathway functions to regulate levels of the TCF transcription factor POP-1 in many lineages (Lin et al., 1998; Phillips et al., 2007; Siegfried et al., 2004) and is described in detail below. Disruption of POP-1 asymmetry results in extension of the intestinal twist posterior to ints-V and VI (Hermann et al., 2000).

3.2.4 Specification of the *C. elegans* Intestine

Many cell fates in the worm are produced by lineage autonomous mechanisms and while this is an important feature of intestinal development, the intestine also

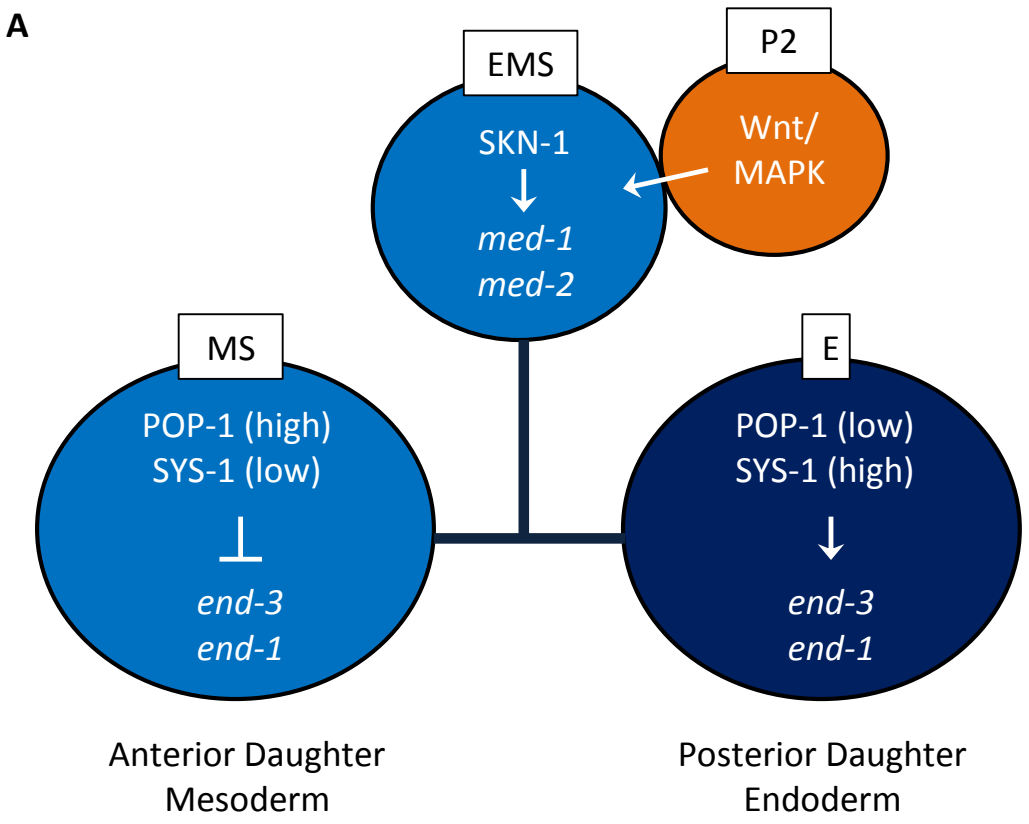
depends on external cues to direct proper specification. The intestine founder blastomere E is born 35 min after the first division as the posterior daughter cell of EMS (Leung et al., 1999; Sulston et al., 1983) and induced to form intestine by an adjacent blastomere (Goldstein, 1992) (Figure 3.4).

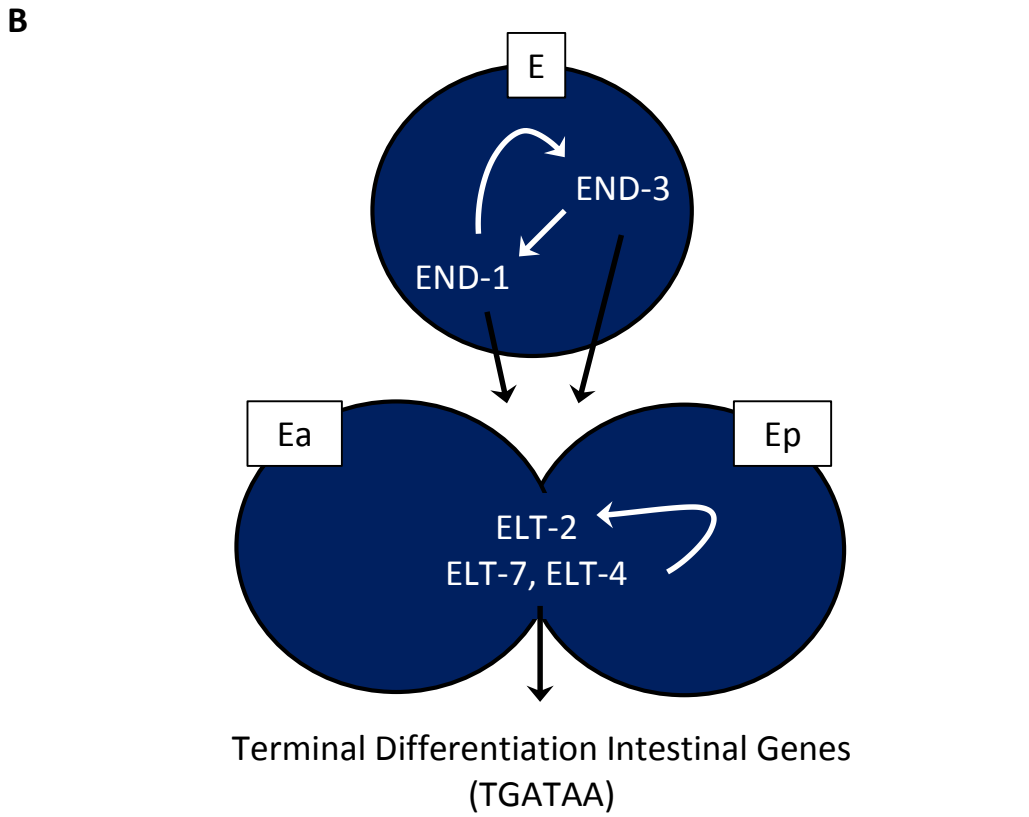
Specification of the EMS cell and its daughter cell E depends on maternally deposited factors (Bowerman et al., 1993). The mRNA for the bZIP homeodomain transcription factor SKN-1 is maternally provided to the embryo but translation of this message is restricted to the P1 blastomere and its descendents, including EMS (Maduro and Rothman, 2002; McGhee, 2013). In the EMS cell, SKN-1 directly activates transcription of two GATA transcription factors MED-1 and MED-2, all of which work together with an external signal to activate transcription of another two GATA transcription factors, END-1 and END-3 in the E cell (Goldstein, 1992; Lin et al., 1998; Maduro et al., 2001) (Figure 3.4). The activity of END-1 and END-3 in the E cell has been proposed to be sufficient for endoderm specification (Maduro et al., 2001; Owraghi et al., 2010; Zhu et al., 1997; Zhu et al., 1998).

SKN-1 activity in the E cell is necessary for transcriptional activation of the *end-1* and *end-3* genes but it is not sufficient (Korswagen, 2002). The redundant MED-1 and MED-2 GATA transcription factors support activation of *end-1* and *end-3*, but are not necessary. In addition to SKN-1 activity, an external Wnt/MAPK signaling event is also necessary for E cell specification and is provided by the P2 cell, which is in direct contact with EMS in the four cell embryo (Goldstein, 1992; Rocheleau et al., 1999; Thorpe et al., 1997) (Figure 3.4). Either removal of the P2 cell prior to induction or mutations in the Wnt/MAPK signaling pathway components such as *mom-2* and *mom-4* result in ectopic mesoderm (MS) at the expense of intestine (Goldstein, 1992; Rocheleau et al., 1997; Thorpe et al., 1997).

This Wnt/MAPK signaling event is a mechanism by which binary cell fate decisions are implemented to distinguish two different cell fates of sister cells during asymmetric divisions in *C. elegans* (Bertrand and Hobert, 2010). Levels of POP-1 (TCF/LEF transcription factor) and its divergent β-catenin coactivator SYS-1 are asymmetrically distributed in the daughter cells of many lineages, resulting in differential

Figure 3.4. A. Specification of the endoderm lineage requires maternal SKN-1 activation of *med-1* and *med-2* in the EMS cell. Wnt/MAPK signalling from P2 to EMS creates an asymmetric distribution of POP-1 and SYS-1 resulting in *end-1/end-3* activation in the E cell and endoderm specification. B. END-1 and END-3 activate the GATA factors ELT-7, ELT-2 and ELT-4 in the E lineage. ELT-2 and ELT-7 activate terminal differentiation intestinal genes by binding to TGATAA sites in their promoters.


A
EMS
P2
SKN-1
Wnt/
MAPK
med-1
med-2
MS
E
POP-1 (high)
SYS-1 (low)
POP-1 (low)
SYS-1 (high)
end-3
end-1
end-3
end-1
Anterior Daughter
Mesoderm
Posterior Daughter
Endoderm


B
E
END-3
END-1
Ea
Ep
ELT-2
ELT-7, ELT-4
Terminal Differentiation Intestinal Genes
(TGATAA)

outputs of gene expression and cell fates (Lin et al., 1998; Phillips et al., 2007; Siegfried et al., 2004). When POP-1 levels are low and SYS-1 levels are high, most POP-1 will be bound to SYS-1 and will function as an activator. In the other scenario, when POP-1 is abundant and SYS-1 is not, POP-1 primarily acts as a repressor of target genes. In the case of endoderm specification, P2 signaling to the posterior of the EMS cell leads to an asymmetric division that gives rise to a posterior daughter E and an anterior daughter MS (Figure 3.4). The signal from P2 consists of a MAPK signal (MOM-4, LIT-1, WRM-1) and a Wnt signal (MOM-2, MOM-5, APR-1) (Kaletta et al., 1997; Rocheleau et al., 1999; Thorpe et al., 1997). The MAPK signal from MOM-4 (MAPKKK) is transduced by LIT-1 (MAPK) and the divergent β-catenin WRM-1 (Lo et al., 2004; Rocheleau et al., 1999; Yang et al., 2011). WRM-1 binds to POP-1, enabling an interaction between LIT-1 and WRM-1, with subsequent LIT-1 phosphorylation of POP-1. These series of interactions and modifications result in POP-1 export from the E nucleus by the 14-3-3 protein PAR-5 and therefore low levels of this TCF/LEF transcription factor in the E nucleus (Lo et al., 2004; Yang et al., 2011). At the same time, the Wnt ligand MOM-2 signals through the Frizzled receptor MOM-5 to inhibit SYS-1 degradation by the APC protein APR-5, resulting in increased SYS-1 levels in the E nucleus. This active export of POP-1 from the nucleus allows the coactivator SYS-1 to bind POP-1 to directly activate *end-1* and *end-3* transcription (Huang et al., 2007; Lin et al., 1998; Lo et al., 2004; Maduro and Rothman, 2002; Phillips et al., 2007; Shetty et al., 2005; Yang et al., 2011). In contrast, the Wnt/MAPK signal is not perceived by the anterior daughter MS cell. Thus, POP-1 levels in the MS nucleus are high, SYS-1 levels are low, resulting in most POP-1 remaining unbound to SYS-1 and direct repression of *end-1/end-3* transcription by POP-1. The absence of *end-1/end-3* transcription in MS leads to MED-1 and MED-2 specifying the anterior (mesoderm) fate (Maduro et al., 2001). In addition to the important role in endoderm specification, the Wnt/β-catenin asymmetry pathway has also been shown to regulate POP-1/SYS-1 levels during the first, third and fourth divisions of the E lineage, which are all A/P divisions (Lin et al., 1998).

END-1 and END-3 are critical for endoderm specification: double mutants do not have an intestine and are 100% lethal (Owraghi et al., 2010). The transcription of both END-1 and END-3 is transient in the E lineage, turning off by the 8E cell stage (Raj et

al., 2010). These transcription factors activate expression of another GATA transcription factor ELT-2 at the 4E cell stage, which drives expression of terminal genes in the intestine (Fukushige et al., 1998; McGhee et al., 2007; McGhee et al., 2009; Zhu et al., 1998). Homozygous null mutants of *elt-2* arrest after embryogenesis at the L1 stage with 20 specified intestinal cells and a gut-obstructed (Gob) phenotype, suggesting that the intestine is non-functional (Fukushige et al., 1998). ELT-4 and ELT-7 are two other GATA transcription factors activated by END-1 and END-3 but neither shows any clear mutant phenotype (McGhee et al., 2009). It has thus been proposed that ELT-2 is the primary regulator of all terminal gene expression in the intestine from the 2E-4E cell stage throughout the life of the animal (McGhee et al., 2007; McGhee et al., 2009).

3.2.5 Spatial Patterning of Terminal Gene Expression in the *C. elegans* Intestine

As described above, the transcriptional hierarchy of the *C. elegans* intestine utilizes multiple GATA transcription factors both for endoderm specification and for activation of terminal differentiation genes. The role of GATA transcription factors in endoderm specification is conserved among other organisms, including vertebrates (Stainier, 2002; Zorn and Wells, 2009). GATA transcription factors are part of the Cys4 superfamily of zinc stabilized DNA binding domains (zinc finger) with four cysteine residues that interact with a Zn^{2+} atom (Clarke and Berg, 1998). These transcription factors bind to the WGATAR motif via hydrophobic interactions in the major groove coupled with a number of non-specific interactions with the sugar-phosphate backbone (Bates et al., 2008; Omichinski et al., 1993). ELT-2 contains a single zinc finger domain that is necessary for binding of WGATAR sequence elements (Hawkins and McGhee, 1995).

ELT-2 activates most terminal differentiation genes in the intestine to drive differentiation and to maintain a functioning organ. This GATA transcription factor is strongly expressed in all intestinal cells and positively autoregulates its own expression (Fukushige et al., 1998; Fukushige et al., 1999). ELT-2 binds to the *cis*-regulatory sequence of TGATAA found overrepresented in the promoters of several hundred intestine specific genes to activate transcription (McGhee et al., 2007; Pauli et al., 2006). However, there are examples of many genes that are expressed only in a subset of

intestinal cells along the A/P axis of the worm, leading to the question of how various spatial patterns of gene expression are controlled.

Spatially patterned genes can be classified based on the cells in which they are expressed. For example, at all stages the genes *pho-1* and *F57F4.4* are both expressed in the posterior 14 cells of the intestine (int-III through int-IX), but not in the anterior six (int-I and int-II) (Beh et al., 1991; Fukushige et al., 2005; NEXTDB The Nematode Expression Pattern Database) (Figure 3.5A, B). Interestingly, *F55G11.2* is expressed at all stages in the complementary pattern of only the anterior most six cells of the intestine (int-I and int-II). The expression pattern of *ZK1193.2* is distinctive in that expression of this gene seems to be restricted to cells of int-II, int-III and int-IV, the cells that undergo the intestinal twist. In addition, there are numerous examples of genes that are expressed in either the anterior or posterior half of the intestine, as well as those that are expressed in an anterior or posterior gradient.

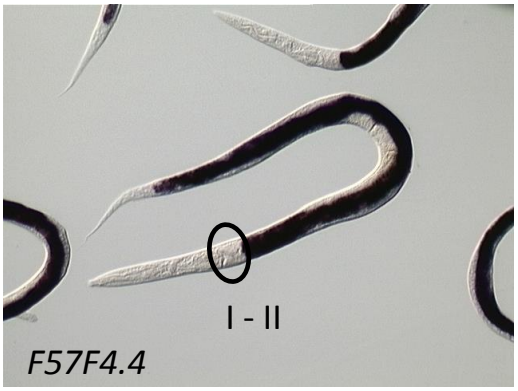
The simplest mechanism to mediate these spatial patterns would be by control of transcription by a combination of ELT-2 and various co-activators or repressors. This is most likely how many genes are regulated in the intestine and has already been shown to be the case for genes that respond to developmental or environmental stimuli such as *mab-3*, *ref-1,* and *ftn-1* (Neves et al., 2007; Romney et al., 2008; Yi and Zarkower, 1999). While this evidence suggests that ELT-2 regulates transcription of intestinal genes in combination with other *trans*-acting factors, very little data exists on how this mechanism potentially controls spatial patterning and what these other *trans*-acting factors might be.

Previous analysis of the transcriptional control of two terminally differentiated intestinal genes, *ges-1* and *pho-1,* identified potential roles for unknown anterior and posterior *trans*-acting factors. Reporter expression of the *ges-1* gut esterase gene is restricted from all intestinal cells to only the anterior six cells of int-I/int-II upon deletion of a 50 bp region from the promoter (ges-1ΔB) (Schroeder and McGhee, 1998). This result indicates the existence of an additional *cis*-acting element which appears to be required for activation of posterior transcription of *ges-1*. Expression of the acid phosphatase PHO-1 has previously been shown to be negatively regulated by the Wnt/ β-catenin asymmetry pathway and requires direct ELT-2 binding to a TGATAA site in its promoter for expression (Fukushige et al., 2005; Huang et al., 2007). As outlined above,

Figure 3.5. The expression patterns of various terminal intestinal genes. A. *F55G11.2* B. *F57F4.4*. Adapted from NextDB.

A

B



POP-1 and SYS-1 are asymmetrically localized during the first, third and fourth divisions in the E lineage and drive asymmetric A/P divisions. When Huang et al. (2007) increased SYS-1 levels in anterior cells, they observed a cell fate switch of int-I and int-II to more posterior intestinal rings, which ultimately de-repressed *pho-1* expression in these cells. It is currently unclear if this altered regulation of *pho-1* expression by POP-1/SYS-1 is by direct action at the level of the promoter or indirectly via alteration of anterior cell fate.

Interestingly, transcription of the *pho-1* gene in the anterior int-I/int-II cells also appears to be under control of the heterochronic gene *lin-14* (Yan, 2007). Knockdown of *lin-14* function by dsRNA mediated interference results in increased frequency and intensity of reporter expression in the anterior most six cells of the intestine (Yan, 2007), signifying a repressive function for this heterochronic gene. Taken together, this evidence suggests that activators and repressors function with ELT-2 to spatially regulate transcription along the A/P axis in the intestine. The expression or function of these potential *trans*-acting factors may be controlled by activity of the Wnt/ β-catenin asymmetry pathway in the E lineage.

3.2.6 LIN-14 and the Heterochronic Pathway of *C. elegans*

The heterochronic pathway in *C. elegans* is responsible for controlling the timing of developmental events during larval stages of the life cycle (Resnick et al., 2010). This pathway regulates the divisions of various lineages such as the vulval and hypodermal seam cells, ensuring the proper divisions occur at each larval stage. Loss of function mutations in genes of this pathway, such as *lin-14*, typically result in precocious phenotypes where later larval stage divisions are observed at the expense of earlier divisions (Ambros and Horvitz, 1984). This leads to a number of phenotypes including a protruding vulva and animals being defective for egg laying.

The novel DNA binding protein LIN-14 is expressed in most cells (including all intestinal cells), starting in late embryogenesis, peaking at the beginning of the L1 stage and disappearing by the end of L1 (Ruvkun and Giusto, 1989). This gene is required for normal L1 developmental events and possibly some events during the L2 stage (Ambros and Horvitz, 1984; Ambros and Horvitz, 1987; Hristova et al., 2005; Ruvkun and Giusto, 1989). LIN-14 protein levels are down-regulated by binding of the 22 nt miRNA *lin-4* to

the *lin-14* 3'UTR, which prevents protein production after translation is initiated (Olsen and Ambros, 1999; Wightman et al., 1991; Wightman et al., 1993). The onset of feeding at hatching triggers post-embryonic development of larval growth, cell division and expression of *lin-4* (Ambros, 2000). Mutations disrupting *lin-4* function or those that prevent *lin-4* from binding to the *lin-14* 3'UTR result in retarded developmental phenotypes, where LIN-14 protein persists and the animal undergoes multiple L1 developmental stages. LIN-14 has been shown to bind to the consensus sequence GAACRY to directly repress transcription of the Insulin/Insulin-Like Growth Factor Gene *ins-33* and it has been postulated that it functions with tissue specific factors to regulate transcription (Hristova et al., 2005).

One of the L1 developmental timing events controlled by LIN-14 is the binucleation of the intestinal cells described previously in section 3.2.2. Loss of function mutations in *lin-14* result in no binucleation of intestinal cells, while gain of function mutations lead to supernumerary nuclei (Ambros and Horvitz, 1984). This is particularly interesting because the cells that do not normally undergo binucleation are also the cells that do not express *pho-1*. Since the pan-intestinal LIN-14 protein is positively regulating binucleation in posterior intestinal rings but negatively regulating *pho-1* transcription in the anterior int-I/int-II rings, there must be one or more factors regulating LIN-14 function in a spatial manner. Additionally, the L1 binucleations and endoreduplications do not occur in loss of function mutants of the cyclin-D *cyd-1* and cyclin dependent kinase *cdk-4*, both of which are required for the transition from G1 to S phase (Boxem and van den Heuvel, 2001; Park and Krause, 1999). RNAi mediated knockdown of *cdk-4* results in the classic heterochronic protruding vulva phenotype (Park and Krause, 1999) and anterior *pho-1* expression (Yan, 2007), indicating that LIN-14 regulation of *pho-1* spatial patterning is intertwined with that of binucleation.

Furthermore, evidence for a link between the Wnt/β-catenin asymmetry pathway and the heterochronic pathway has been previously found (Ren and Zhang, 2010). The authors discovered that a loss of function mutation in *lit-1* (and other Wnt/β-catenin asymmetry pathway components) suppresses heterochronic retarded phenotypes in the hypodermis (Ren and Zhang, 2010). This evidence suggests that LIN-14 may work with

the Wnt/β-catenin asymmetry pathway to regulate spatial patterning in the intestine, in addition to its well-known developmental timing role.

3.3 Materials and Methods

3.3.1 *C. elegans* Strains

Standard nematode handling conditions were used (Brenner, 1974). Animals were grown at 20°C. Strains used were wild type N2, JM139 *caIs67[pho-1p::gfp, rol-6(su1006)] V*, JM149 *caIs71[elt-2p::gfp, rol-6(su1006)]*, DR441 *lin-14(n179) X*, and *Ex[pho-1p(lin-14KO)::gfp, ttx-3p::rfp, rol-6(su1006)]*.

3.3.2 Promoter Sequence Alignment

Putative LIN-14 consensus binding sites were identified manually by searching through the *C. elegans pho-1* promoter. Clustal Omega (Sievers et al., 2011) was used to align the *pho-1* promoter sequences from *C. elegans, C. brigssae* and *C. remanei*.

3.3.3 Plasmid construction

The *pho-1p(*LIN-14KO*)::gfp* construct was generated using a synthesized *pho-1*(LIN-14KO) promoter as a 439 bp insert in pCR2.1. This was double digested with BamH1 and Sbf1 to give a 439 bp insert that was inserted into the pJM355 (eGFP) backbone to create ~5kb plasmid. The new plasmid was verified by sequencing.

3.3.4 Generating transgenics by microinjection

C. elegans transgenic lines were created using standard microinjection techniques (Mello et al., 1991). The *pho-1p(lin-14KO)::gfp* construct was injected at 20 ng/µL or 60 ng/µL, together with 5 ng/µL of *ttx-3p::rfp* (pJM356) and 50 ng/µL of *rol-6(su1006)* (pRF4). dsRNA microinjections were carried out in the same manner.

3.3.5 Image Acquisition and GFP Intensity Quantitation

Young adult hermaphrodites were allowed to lay eggs for 2 hours at 20^{o}C after which the adults were removed and the progeny were incubated at 20^{o}C for various

times. Animals were transferred to 2% agar pads on glass slides, anaesthetized with 5 mM levamisole diluted in 1X M9 buffer and photographed at 20x or 40x magnification under Normarski Differential Interference Contrast (DIC) and fluorescence optics. Images were captured using a Zeiss Axioplan2i microscope with a Hamamatsu Orca ER digital camera and AxioVision (version 4.8.1) software. Length calibration was achieved using a Pyser-SGI micrometer slide. GFP expression intensity was measured as the total pixel intensity within a box drawn around each intestinal ring, subtracting the same box from the background of the slide away from the worm. Sigma Plot 12.5 was used to perform Mann-Whitney rank sum tests for statistical significance. Confocal stacks were taken on a Zeiss Elyra microscope with Zen software.

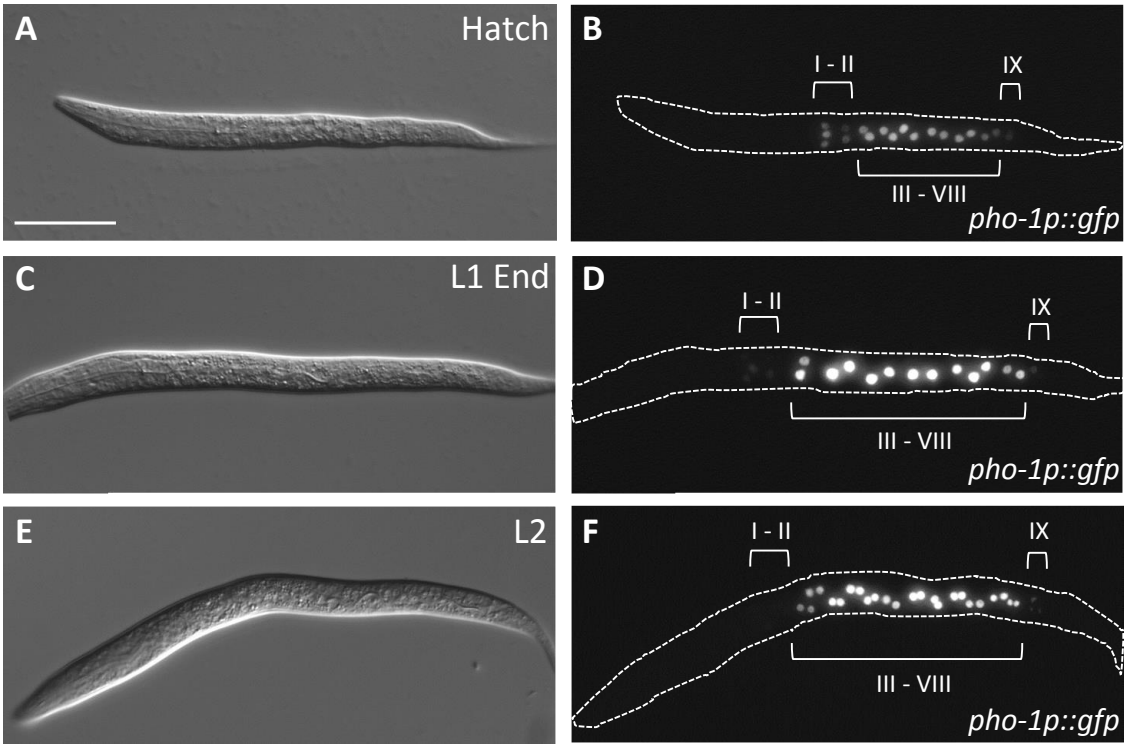
3.4 Results

3.4.1 The *pho-1* Spatial Expression Pattern Forms During the L1 Larval Stage

It has previously been reported that the *pho-1* expression pattern in the intestine is present at all stages of development (Beh et al., 1991; Fukushige et al., 2005; NEXTDB The Nematode Expression Pattern Database). The expression of an integrated multi-copy transcriptional *pho-1p::gfp* nuclear reporter array at 20^{o}C was quantitated at various stages. This reporter strain JM139 was observed to grow slightly slower than Wt or healthy strains; hence multiple time points during early development were assayed to identify when the first larval molt occurred. This was reasoned to be an important time point to identify as it was hypothesized that there may be a connection between the two apparent roles of LIN-14 in the intestine, regulating binucleation and *pho-1* expression. Four time points were assayed during the L1 larval stage; one hour after hatching, five hours after hatching, 15 hour after hatching and 20 hours after hatching. It was determined that the first larval molt in this strain on average took place approximately 20 hours after hatching based on the presence of worms with fully binucleated intestinal cells and worms in the process of shedding their cuticle.

At one hour after hatching, *pho-1p::gfp* expression is visible in all 20 nuclei of the intestine with comparable intensity observed between int-I, int-III and int-IV (Figure 3.6B), particularly clear in confocal stacks taken at the start of the L1 stage (Figure 3.7A,

Figure 3.6. Representative images of *pho-1p::gfp* expression at 20°C. A-B. 1 hour post hatching (PH), C-D. 20 hours PH, not binucleated. E-F. 20 hours PH, binucleated. G. Quantitation of mean GFP Pixel intensities in the anterior intestine shows a steady decline in int-I/int-II expression and a sharp increase in int-III/int-IV expression during the L1 stage. ** denotes statistical significance of $p < 0.001$. All images taken at 40x magnification with a GFP exposure of 450 ms. Scalebar is 100 µm.


A
Hatch
B
I - II
IX
III - VIII
pho-1p::gfp
C
L1 End
D
I - II
IX
III - VIII
pho-1p::gfp
E
L2
F
I - II
IX
III - VIII
pho-1p::gfp

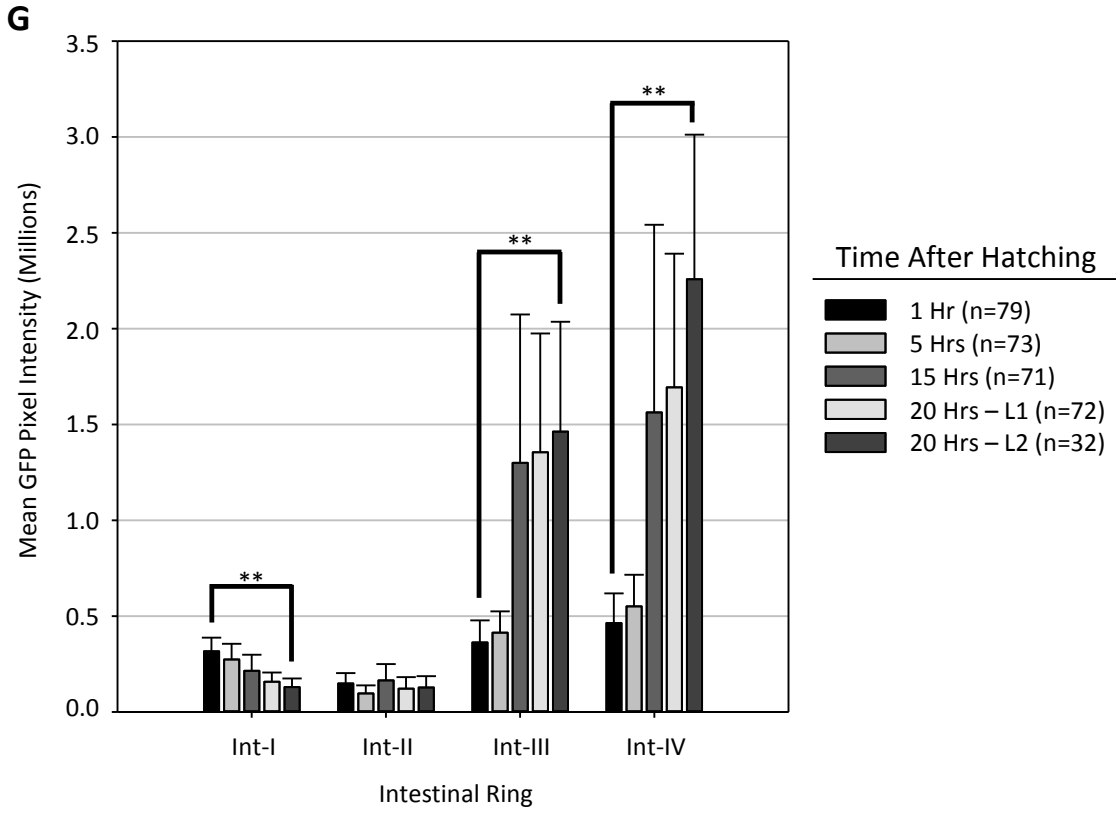
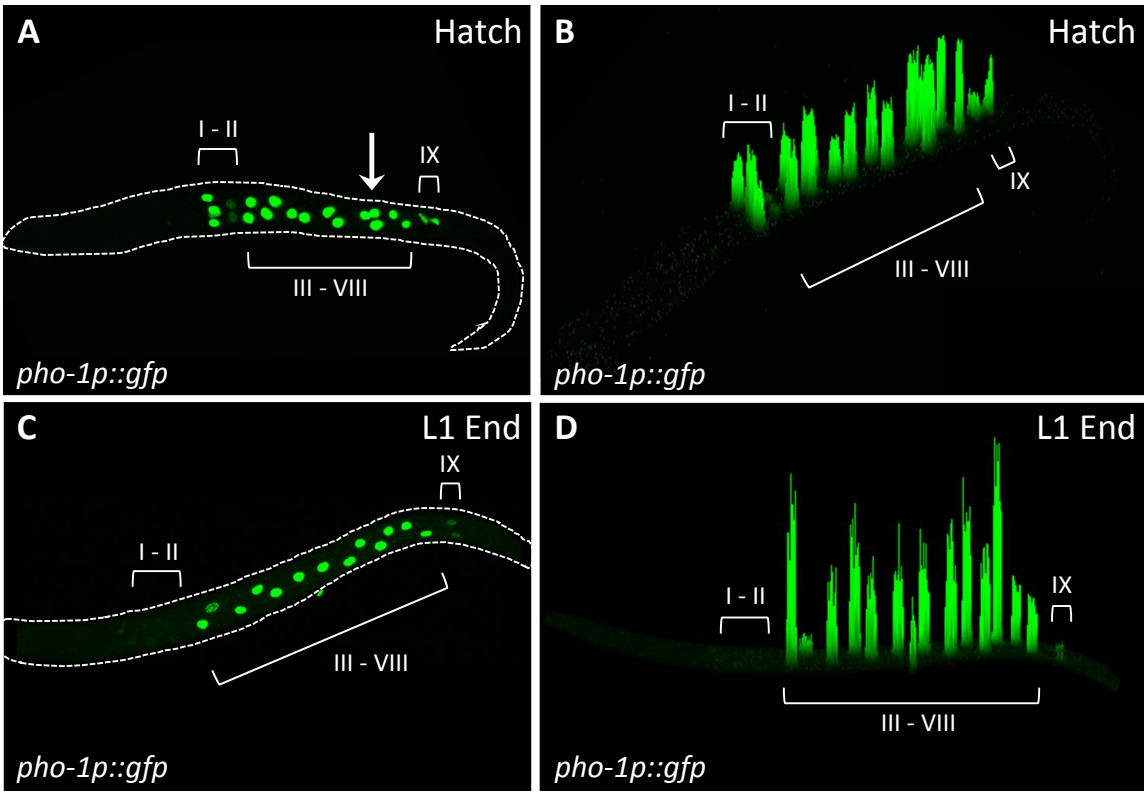

G
Mean GFP Pixel Intensity (Millions)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
**
**
**
Int-I
Int-II
Int-III
Int-IV
Intestinal Ring
Time After Hatching
1 Hr (n=79)
5 Hrs (n=73)
15 Hrs (n=71)
20 Hrs – L1 (n=72)
20 Hrs – L2 (n=32)

Figure 3.7. Confocal stack images of *pho-1p::gfp* expression at 20ºC. A. 1 hour PH. C. End of L1. B and D. 2.5D projections of same confocal stacks where each vertical green line indicates the pixel intensity at that location of the image.

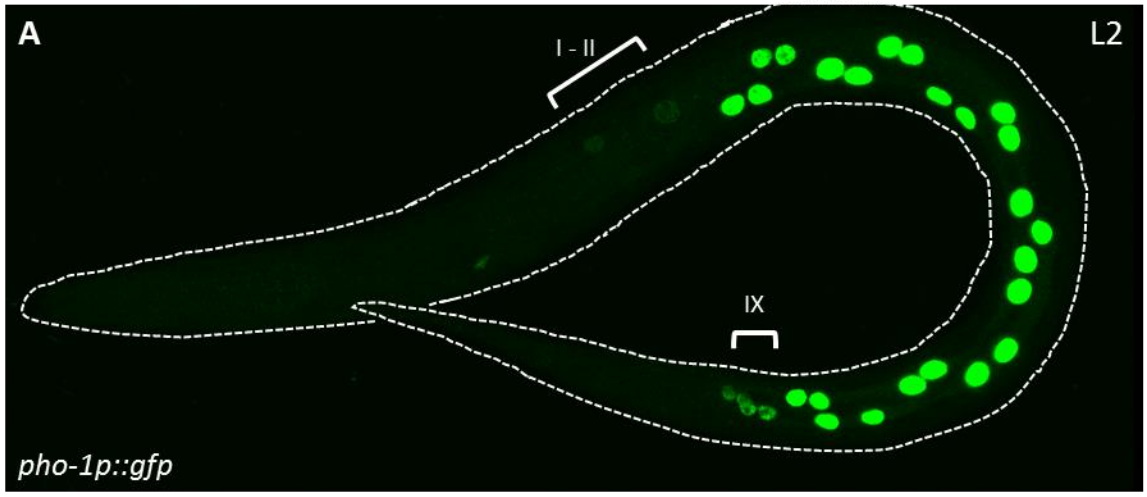

A
Hatch
I - II
IX
III - VIII
pho-1p::gfp
B
Hatch
I - II
IX
III - VIII
pho-1p::gfp
C
L1 End
IX
I - II
III - VIII
pho-1p::gfp
D
L1 End
I - II
IX
III - VIII
pho-1p::gfp

B). This result is in contrast with previous studies that found little to no expression in the anterior six intestinal cells of int-I/int-II at any stage of development (Fukushige et al., 2005). However, the animals in this study contained a GFP reporter as opposed to the *lacZ* reporter and were not starved when imaged, two factors that could account for this difference. At this first time point, int-II and int-IX were observed to have weaker expression of the GFP reporter relative to the nuclei of other intestinal rings. Interestingly, int-I expression was observed to have declined while expression in the posterior rings greatly increased at later time points during the first larval stage (Figure 3.6D, F; 3.7C, D). By the end of the L1 stage, the intensity of GFP expression in int-I was approximately half of the intensity it was just after hatching (Figure 3.6G). In contrast, the GFP expression intensities of the more posterior rings int-III and int-IV were about 3.7 times higher at the end of the L1 stage then at the start. These results indicate that expression of *pho-1* is initially present in the most anterior intestinal cells of int-I but is progressively turned off during the first larval stage. Close inspection of the GFP expression intensity in int-II during the first larval stage revealed only low levels of expression that did not change during the first larval stage (Figure 3.6; 3.7). Similarly, expression in int-IX was always observed to be present during all assayed time points of the L1 stage but was consistently weaker in intensity than the other posterior intestinal rings.

The observed spatial patterning within the intestine at the end of the L1 stage was maintained at the start of the L2 stage, with low levels of GFP expression present in int-I/int-II and high levels found in the binucleated posterior intestinal rings int-III to int-VIII (Figure 3.6F, 3.8). Again, the most posterior ring int-IX was observed to have weaker expression than the other rings. This pattern of *pho-1* expression was also found to be maintained during the L3, L4 and young adult stages (Figure 3.9).

As a control, intestinal expression of an integrated multi-copy transcriptional *elt-2p::gfp* nuclear reporter array at 20^{o}C was also quantitated. The JM149 *elt-2p::gfp* strain grew at normal rate and underwent the first larval molt approximately 15 hours after hatching or five hours prior to JM139. At the start of the L1 stage one hour after hatching, *elt-2p::gfp* expression could be seen in all 20 intestinal nuclei with slightly stronger intensity observed in the anterior and posterior regions (Figure 3.10B). GFP

Figure 3.8. A. Confocal stack image of *pho-1p::gfp* expression at 20°C in L2 larva. B. 2.5D projection of same confocal stack where each vertical green line indicates the pixel intensity at that location of the image.


A
L2
I - II
IX
pho-1p::gfp

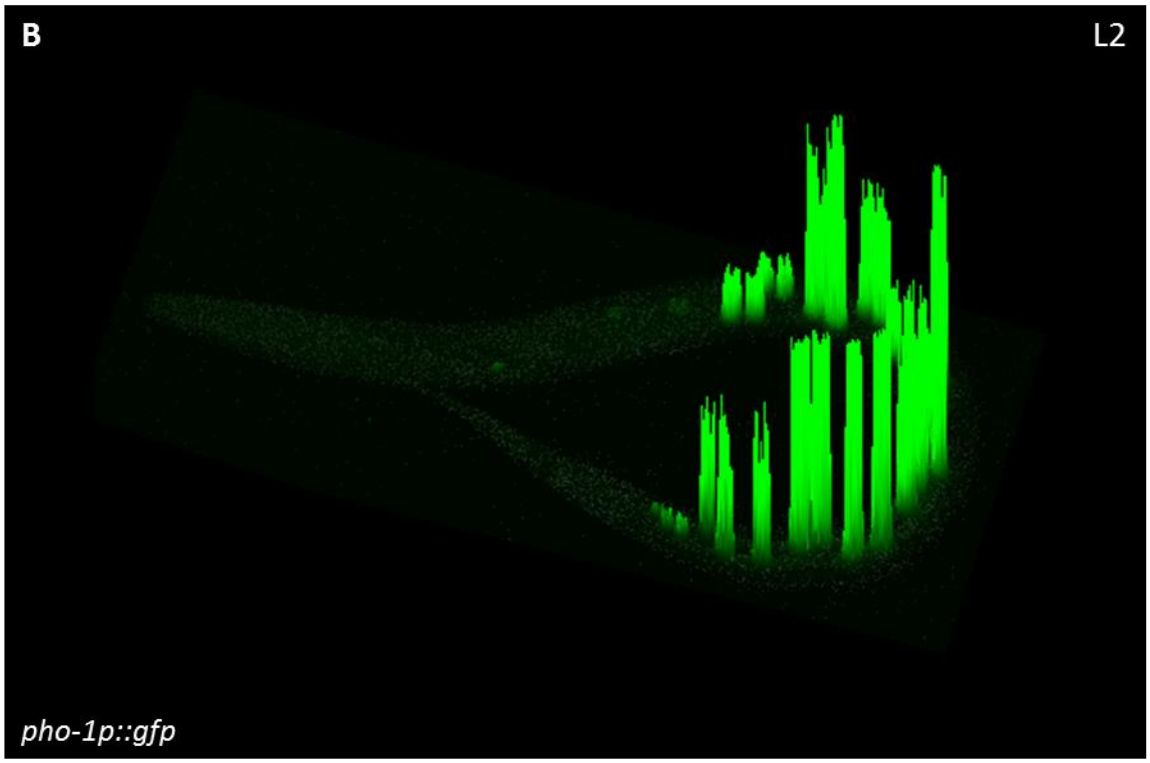
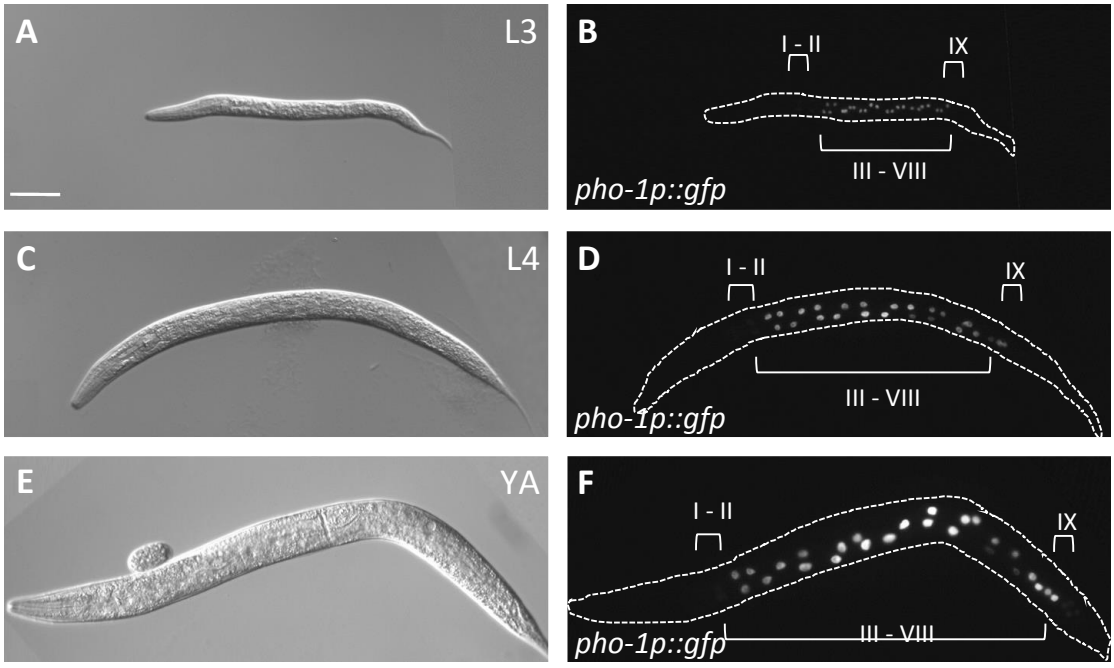

B
L2
pho-1p::gfp

Figure 3.9. Representative images of *pho-1p::gfp* expression at 20ºC. A-B L3 stage. C-D. L4 stage. E-F. Young adult stage. G. Quantitation of mean GFP Pixel intensities in the anterior intestine shows maintenance of the spatial patterning of *pho-1*. ** denotes statistical significance of $p < 0.001$. All images taken at 20x magnification with a GFP exposure of 100 ms. Scalebar is 100 µm.


A
L3
B
I - II
IX
III - VIII
pho-1p::gfp
C
L4
D
I - II
IX
III - VIII
pho-1p::gfp
E
YA
F
I - II
IX
III - VIII
pho-1p::gfp

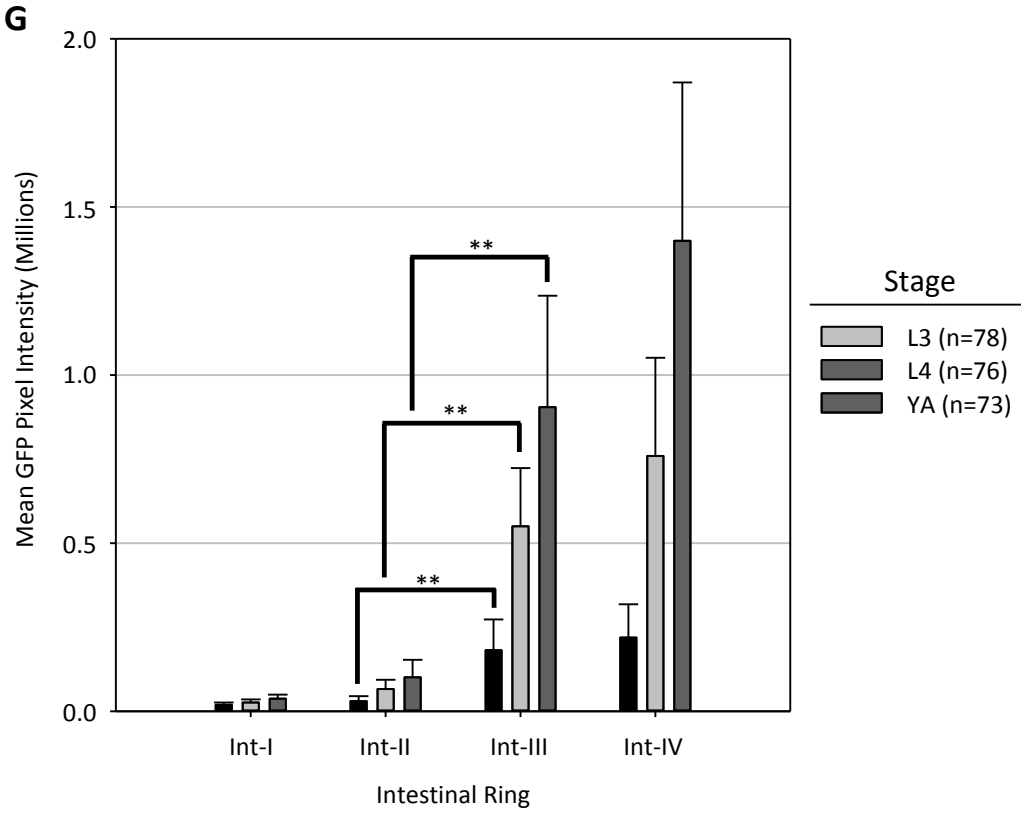
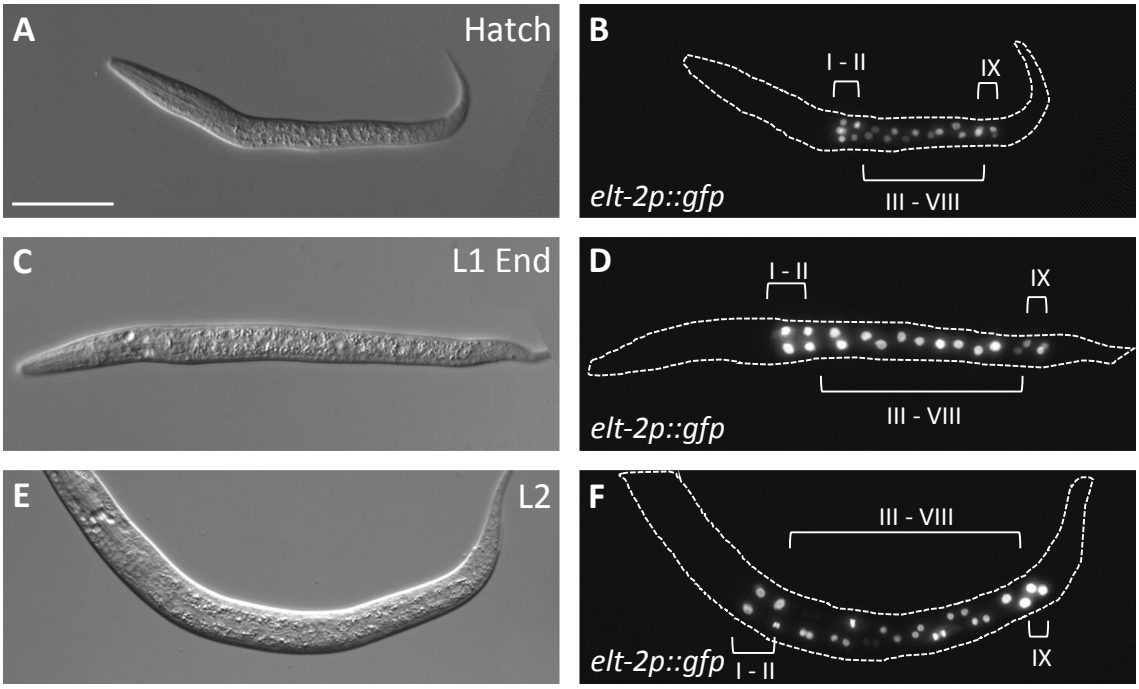
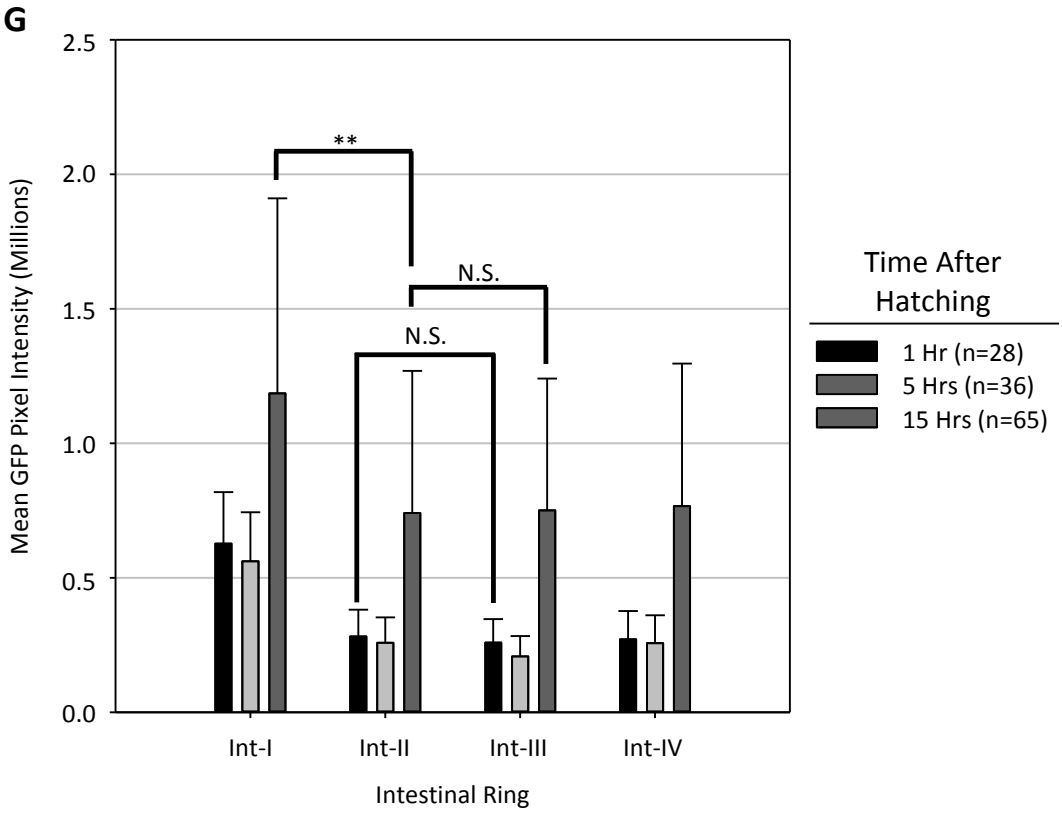

G
2.0
1.5
1.0
0.5
0.0
Mean GFP Pixel Intensity (Millions)
**
**
**
Int-I
Int-II
Int-III
Int-IV
Intestinal Ring
Stage
L3 (n=78)
L4 (n=76)
YA (n=73)

Figure 3.10. Representative images of *elt-2p::gfp* expression at 20ºC. A-B. 1 hour PH. C-D. 15 hours PH, not binucleated. E-F. 15 hours PH, binucleated. G. Quantitation of mean GFP pixel intensities in the anterior intestine shows no change in GFP intensity pattern between the anterior four intestinal rings during the L1 stage. ** denotes statistical significance of $p < 0.001$, N.S. indicates no significant difference. All images taken at 40x magnification with a GFP exposure of 100 ms. Scalebar is 100 µm.


A
Hatch
B
I - II
IX
elt-2p::gfp
III - VIII
C
L1 End
D
I - II
IX
III - VIII
elt-2p::gfp
E
L2
F
III - VIII
elt-2p::gfp
I - II
IX


G
Mean GFP Pixel Intensity (Millions)
2.5
2.0
1.5
1.0
0.5
0.0
**
N.S.
N.S.
Int-I
Int-II
Int-III
Int-IV
Intestinal Ring
Time After Hatching
1 Hr (n=28)
5 Hrs (n=36)
15 Hrs (n=65)

intensity in all intestinal nuclei had increased by the end of the first larval stage and start of the L2 stage but was still ubiquitous in the intestine with slightly more expression at the poles of the intestine (Figure 3.10D, F).

Taken together, these finding suggest that the spatial expression pattern of *pho-1* along the A/P axis of the intestine is established during the first larval stage. *pho-1* expression initially is observed in all cells but becomes restricted to the posterior 14 cells of int-III to int-IX by the end of the first larval stage. Once established, this pattern is maintained throughout the life of the animal.

3.4.2 Analysis of Regulatory Sites in the *pho-1* Promoter

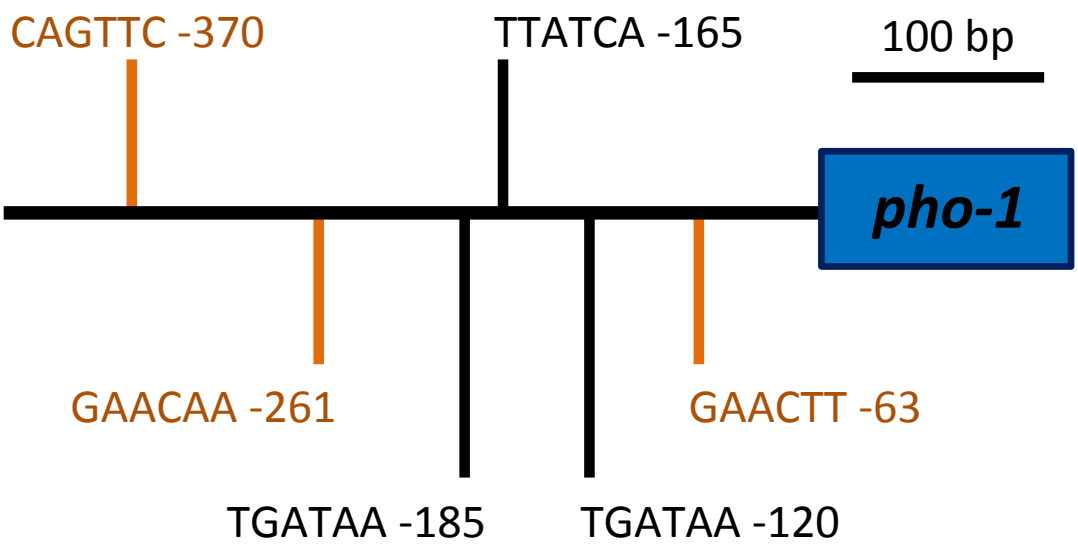
The formation timeline of *pho-1* spatial expression pattern during the first larval stage fits well with a role for LIN-14 in regulating this event. As outlined above, RNAi mediated knockdown of *lin-14* function results in reduced *pho-1* patterning and protein levels of LIN-14 peak at hatching. Thus, LIN-14 is a good candidate to be involved in repressing *pho-1* transcription in the anterior int-I/int-II cells by direct action on the *pho-1* promoter.

The consensus binding sequence for LIN-14 was identified as GAACRY (Hristova et al., 2005) where R is any purine (adenine or guanine) and Y is any pyrimidine (cytosine or guanine). Searching for this site in the 425bp *C. elegans pho-1* promoter (defined as everything from the first ATG to the upstream 5' gene) yielded three putative binding sites. All three potential LIN-14 binding sites contained the core GAAC motif with two scoring five out of six bases in the consensus and one that matched four out of six (Figure 3.11A, B). The *C. elegans pho-1* promoter was aligned to the promoters of the *pho-1* homologs in *C. briggsae* and *C. remanei* since *cis*-regulatory elements are more likely to be genuine functional binding sites for transcription factors when they are conserved amongst related species (Nelson and Wardle, 2013). Of the three putative LIN-14 sites, only the 3' most proximal downstream site (-66bp) is conserved between all three species (Figure 3.11C).

Additionally, previous work had already shown that only the 214bp immediately upstream of the *pho-1* gene is required to produce the spatially patterned expression within the intestine (Fukushige et al., 2005). This includes the three conserved TGATAA

Figure 3.11. A. The 425 bp long *pho-1* promoter contains three conserved TGATAA sites and three putative GAACRY LIN-14 binding sites. B. None of the LIN-14 sites match the consensus binding sequence completely. C. Clustal Omega alignment of the *pho-1* promoters for *C. elegans* (CEL), *C. remanei* (CRE) and *C. Briggsae* (CBG) reveals conservation of all three TGATAA sites but only conservation of the most downstream putative LIN-14 site.

A



B

LIN-14 Site: GAACRY
-370 bp: GAACTG 4/6
-261 bp: GAACAA 5/6
-63 bp: GAACTT 5/6

C

```
CBG_pho-1       ------------------------------------------------------------TTTT- 4
CRE_pho-1       ------------------------------------------------------------TTTTC 5
CEL_pho-1       ATTTGTGTCCAACATATAAGTTTGGGGAAGCCTAAAAACAAAAAAAAAGCAGCAGTTCGA 60
                                                                            **

CBG_pho-1       -ACCGCTCCT--TCA-----------------TTTTGTACA------------------T 26
CRE_pho-1       AACTGCAACTACTCA-----------------TCT-GTACA------------------T 29
CEL_pho-1       GATGGATCTTCCTCAGATGTGACAAGTACGACTATTGTAGAGTTACTGAAGATGCAAAAT 120
                 *  *    *  ***                * * *** *                   *

CBG_pho-1       T--ACCA-----------TT-----GAAAC---AGTTAC-----------------TGAATA 50
CRE_pho-1       TTTAATA-----------TTTGCTCAAAAC---AGTTGC-----------------TGAATA 60
CEL_pho-1       TGCATCAGCCGCGACTAGTTGTAGAAAAACCAAAATTCCGGGAACAAACCAACTTGAATA 180
                *  *  *           **      ****   * ** *             ******

CBG_pho-1       AACTTCGTTTGTCCATGTAAATTGACTTAAAGTTTTAGACACTTATTATAGAAGCAAATT 110
CRE_pho-1       AAGGGGTTTTGTCATTGTAAATTGTATTAAAGTTTTAGACATTTCTCTTC----CAAATT 116
CEL_pho-1       AAAGTCGAATTTCTATGTAAACTGTACTAAAGAATTGGACACTCATCAGTG--GCA---T 235
                **       * **  ****** **   *****  ** **** *  *        **   *

CBG_pho-1       GATGATAAAACAGCATTGATCACACTATCGAAAAAAGAAGA----GTATATTGAAGTGGC 166
CRE_pho-1       GATGATAAAACAGCATTGATCAT--TATCAAAAAA----GA----GAATATTGAAGCGAC 166
CEL_pho-1       GATGATAAAGAAGCATTGATCAT--TATCAAAAAAAGAAGGAGACGAATATTG--GCATC 291
                ********  **********   ****  *****    *     * ******  *   *

CBG_pho-1       AGAGAATGGTAGTCAACTGATAAAATTGCATTACTCTA----------CGCACTTTTCTT 216
CRE_pho-1       AG----TGGTAGTCAACTGATAAAAC-GCATTACTACA----------T---CTTTCCAT 208
CEL_pho-1       GA------GTAGCCAACTGATAAAAG-ACATTACTACAATTTCTCCGGCATGCTCCGGTT 344
                      **** *** ********   *******  *                **     *

CBG_pho-1       TTTTATTTCCTCCCTGCTCGGAACTTCTCTCTATTTTCTGAAAACTCGAAATCACAAAAC 276
CRE_pho-1       TTTTATTTCGTCTCT-----GAACTC-------TTTTCTTGAAACTCGAAATCACAATTT 256
CEL_pho-1       TTTTGTTTCGTCTCT-----GAACTTGATTTTCATTGCACAAAACTCTAAATCACA-TTT 398
                **** **** ** **     *****         ** *   ****** ********

CBG_pho-1       TCC---CTTTTTTCAGGG-TAAATTTCACT 302
CRE_pho-1       TTTAAATTGTTTTCAGGG-TAACTTTCACT 285
CEL_pho-1       TCCG---CATTTTCAGGGGTAACTTTTACT 425
                *       ********* *** *** ***
```

ELT-2 Site LIN-14 Site

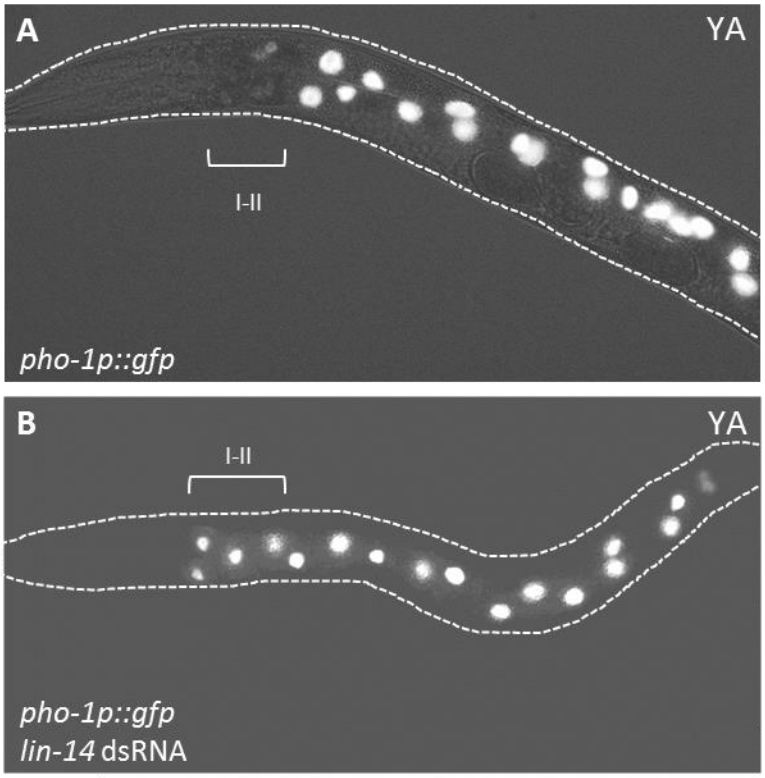
cis-regulatory elements to which ELT-2 binds, as well as the single conserved LIN-14 binding site. Interestingly, it seems that only the downstream most ELT-2 binding site (-123bp, 57bp from the putative LIN-14 element) is necessary for intestinal expression. Mutation of this TGATAA motif was shown to abolish almost all reporter expression whereas the upstream sites appear to simply boost expression levels (Fukushige et al, 2005). This points to the -66bp putative LIN-14 binding site as potentially important in mediating the *pho-1* spatial pattern of expression in the intestine.

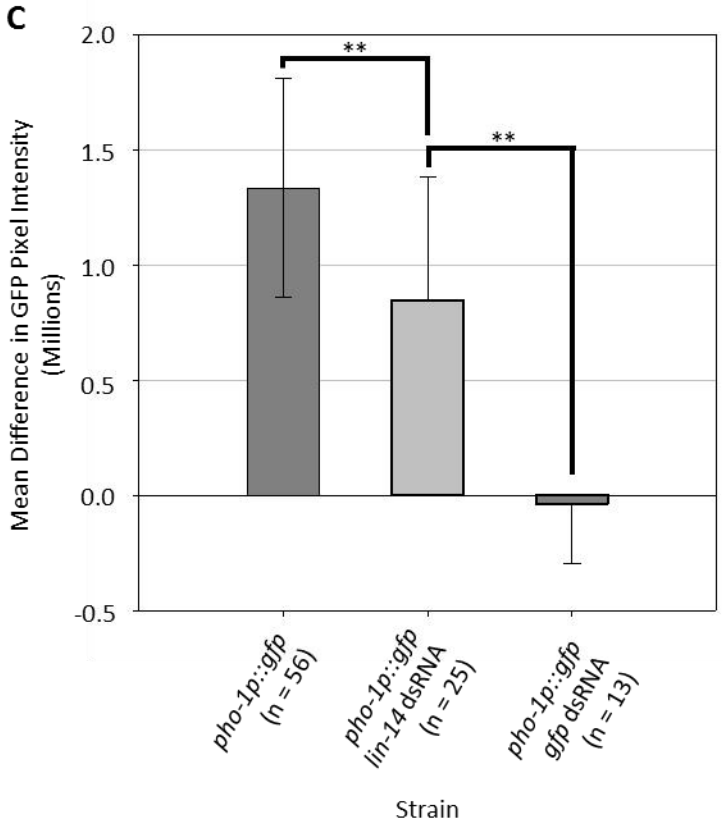
3.4.3 LIN-14 is Required for Repression of *pho-1* Expression in the Anterior Intestine

As detailed in section 2.2.5, the spatial pattern of *pho-1* expression in the posterior int-III to int-IX cells of the intestine was previously shown to be disrupted by RNAi against *lin-14* (Yan, 2007). This experiment was repeated and confirmed as shown in Figure 3.12. Expression in *pho-1p::gfp* young adult worms was observed in the expected pattern - little to no expression observed in int-I/int-II and strong expression observed in the binucleated posterior intestinal rings III through IX. In contrast, the F1 progeny of hermaphrodite worms injected with dsRNA against *lin-14* had the well described *lin-14* loss of function phenotype of a protruding vuvla coupled with failure to undergo intestinal binucleation. Additionally, there was noticeably increased expression of *pho-1* in the anterior int-I/int-II nuclei (Figure 3.12B). Quantitation of the difference in GFP expression intensity between int-III + int-IV and int-I + int-II confirms this. The difference in GFP expression between the more posterior int-III + int-IV and the anterior int-I + int-II is smaller in the *lin-14* dsRNA treated animals compared to the control untreated animals (Figure 3.12C). dsRNA against *gfp* was used as a control for the *lin-14* dsRNA treatment and shows no difference between the anterior and posterior regions as no GFP expression was observed in these animals.

As an extension of this experiment, the same integrated multi-copy *pho-1p::gfp* reporter array was crossed into the *lin-14(n179)* mutant background. The *lin-14(n179)* mutation is an A23013G change encoding a glycine (G) substitution for a conserved arginine (R303) residue in both LIN-14 isoforms (Reinhart and Ruvkun, 2001). It is a temperature sensitive loss of function mutation with animals showing no phenotype at the

Figure 3.12. Representative images of *pho-1p::gfp* expression at the young adult stage. A. No dsRNA B. *lin-14* dsRNA C. Quantitation of increased *pho-1p::gfp* intensity in the anterior intestine as the difference in mean GFP pixel intensity between int-III/int-IV and int-I/int-II. ** denotes p < 0.001. All images taken at 20x magnification with a GFP exposure of 100 ms.


A
YA
I-II
pho-1p::gfp
B
YA
I-II
pho-1p::gfp
lin-14 dsRNA

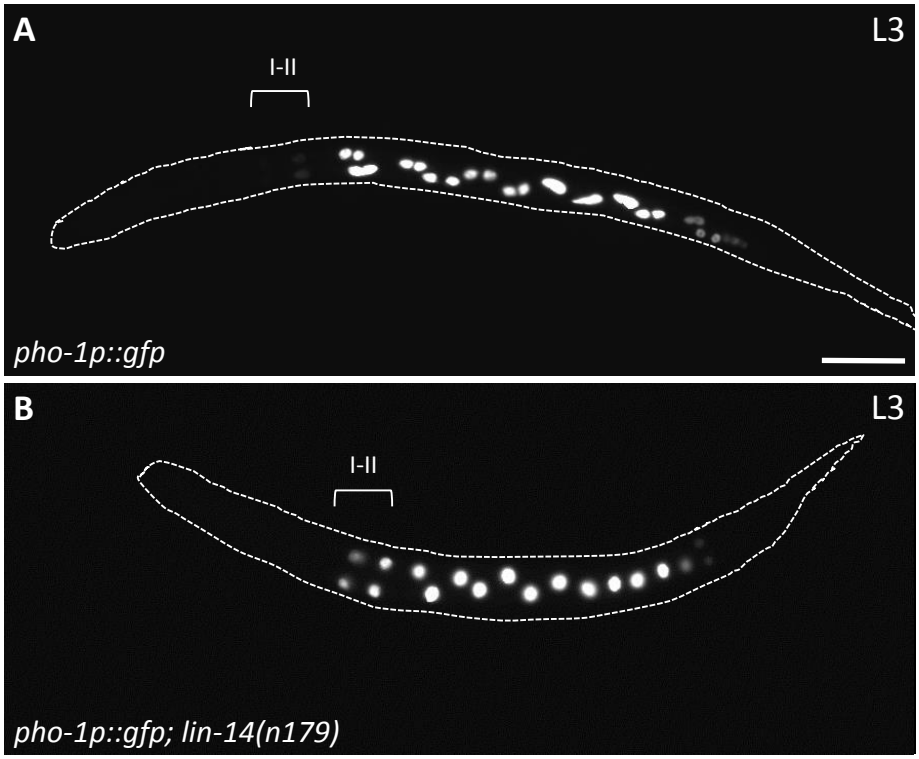

C
2.0
1.5
1.0
0.5
0.0
-0.5
**
**
Mean Difference in GFP Pixel Intensity (Millions)
pho-1p::gfp (n = 56)
pho-1p::gfp lin-14 dsRNA (n = 25)
pho-1p::gfp gfp dsRNA (n = 13)
Strain

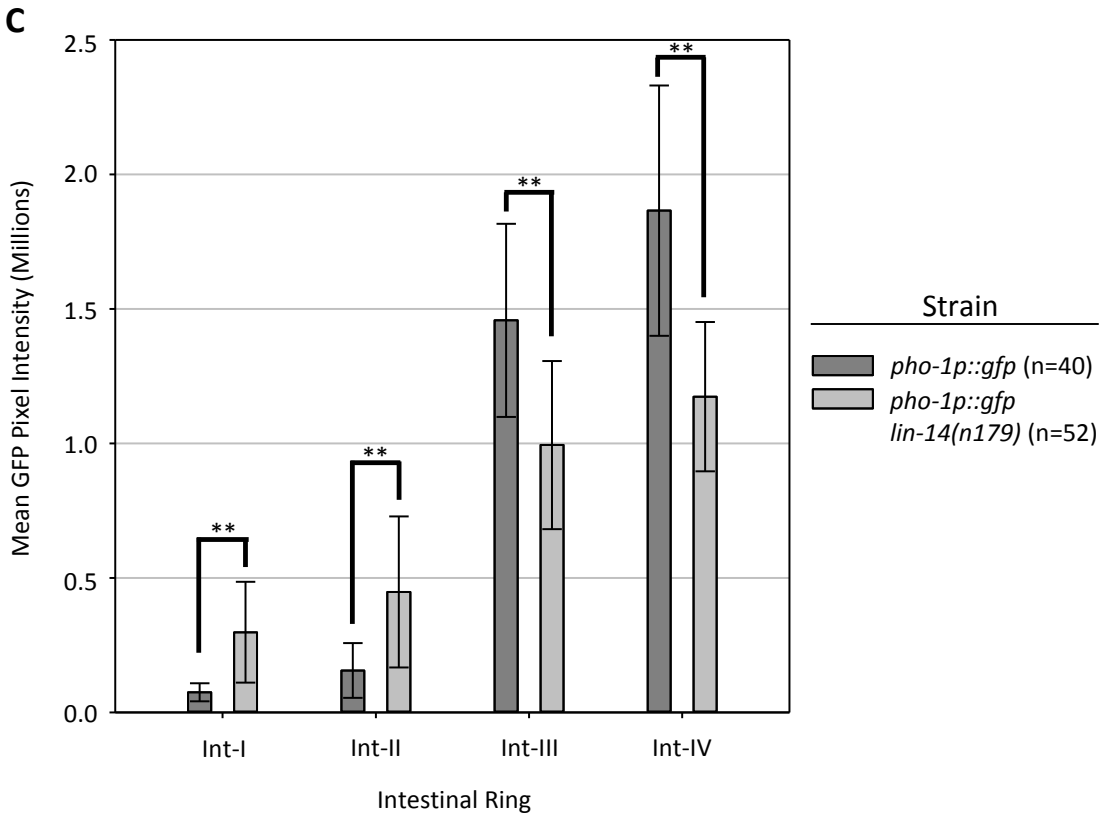
permissive temperature of 15ºC but at restrictive temperatures of 20ºC and above, animals have multiple lineage defects resulting in the classic heterochronic loss of function phenotypes (Ambros and Horvitz, 1987).

Expression of the *pho-1p::gfp* reporter in the *lin-14(n179)* background was examined for increased anterior expression at multiple stages of development at the restrictive temperature of 25ºC. This strain was compared to the original JM139 reporter strain to identify differences in spatial GFP expression at 25ºC. Phenocopying the dsRNA phenotype, the *lin-14(n179)* mutant strain was observed to have increased anterior expression in int-I/int-II relative to int-III/int-IV at multiple stages of development, particularly at the L3 stage (Figure 3.13). This phenotype was observed to vary in penetrance/severity as some animals displayed very little phenotype while others had quite obvious anterior expression. By quantitating the intensity of GFP expression in each intestinal ring, it is clear there is a statistically significant increase of int-I/int-II expression in the *lin-14(n179)* mutant compared to control (Figure 3.13C). However, the GFP intensity in these cells is still not as high as it is in the more posterior cells of int-III/int-IV. The intensity of int-III/int-IV GFP expression in the mutant was reduced compared to the control, suggesting that the reduced spatial patterning phenotype may be a reflection of an equalization of expression levels along the A/P axis of the intestine rather than simply increased anterior expression. Conversely, *lin-14(n179)* mutants do not undergo intestinal binucleation so this reduction of int-III/int-IV expression intensity may be an artifact of comparing two large nuclei (mutant) with four smaller nuclei (control) (Figure 3.13).

Based on the RNAi and mutant phenotypes, it can be concluded that LIN-14 is required to maintain the normal *pho-1* spatial expression pattern in the intestine. These results suggest that *pho-1* transcription in the anterior six cells of the intestine is repressed by LIN-14 action. It is interesting to note that this abolition of spatial patterning is maintained even after protein levels of LIN-14 are undetectable in later developmental stages (after L2). It is unclear what mechanism LIN-14 may be acting by and if it is regulating transcription of *pho-1* by directly binding to the *pho-1* promoter.

Figure 3.13. Representative images at the L3 stage of A. *pho-1p::gfp* expression. B. *pho-1p::gfp*; *lin-14(n179)* expression. C. Quantitation of increased *pho-1p::gfp* intensity in the anterior intestine as the difference in mean GFP pixel intensity between int-III/int-IV and int-I/int-II. ** denotes p < 0.001. All images taken at 20x magnification with a GFP exposure of 100 ms. Scalebar is 100 µm.


A
L3
I-II
pho-1p::gfp
B
L3
I-II
pho-1p::gfp; lin-14(n179)


C
Mean GFP Pixel Intensity (Millions)
2.5
2.0
1.5
1.0
0.5
0.0
**
**
**
**
Int-I
Int-II
Int-III
Int-IV
Intestinal Ring
Strain
pho-1p::gfp (n=40)
pho-1p::gfp
lin-14(n179) (n=52)

3.4.4 LIN-14 Does Not Act Directly on the *pho-1* Promoter to Mediate Spatial Patterning

There are three potential LIN-14 binding sites in the 425bp *pho-1* promoter and all three were knocked out to determine if LIN-14 is interacting with them to repress anterior *pho-1* expression. The LIN-14 binding sites were disrupted by substituting purine bases for pyrimidines and vice versa as well as switching the number of hydrogen bonds (Figure 3.14B). The core GAAC motif was mutated to TCCA, where for example, the three hydrogen bond forming purine G was replaced with a two hydrogen bond forming pyrimidine T.

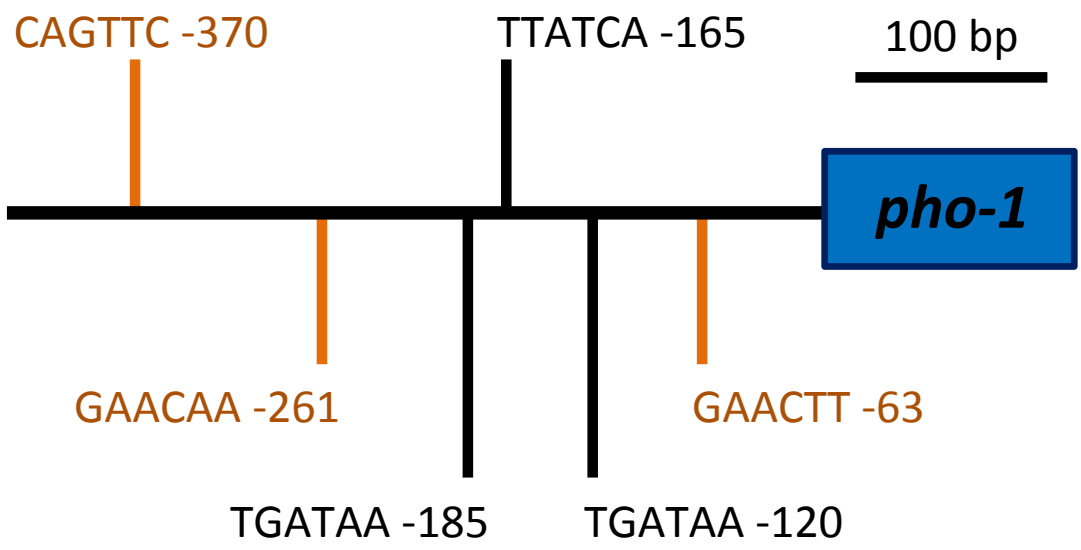
Transgenic strains containing multiple copies of the extra-chromosomal LIN-14 knockout (KO) *pho-1* promoter driving *gfp* were generated and analyzed for increased anterior intestinal expression. A total of six independently created transgenic lines with two different injection doses of the reporter construct were obtained. Despite the number of lines available, for all lines most worms were observed to have only weak and/or mosaic expression (data not shown). This is likely due to the silencing of the multi-copy array (Hsieh and Fire, 2000). The minority of worms that did have stronger expression within the intestine were never observed to have anterior expression in int-I/int-II (Figure 3.15). This suggests that LIN-14 regulates spatial patterning of *pho-1* transcription by an indirect mechanism.

3.4.5 Intestinal Development is not as Invariant as Previously Thought

The mature intestine formed during embryogenesis is widely accepted by the field to contain 20 intestinal cells, some of which go on to binucleate, producing an organ of 20 cells with 30-34 nuclei. It is supposedly rare (but not impossible) to observe an extra cell near the location of int-VII. This extra cell was observed in the *pho-1p::gfp* reporter strain at a higher rate than expected (Figure 3.7A arrow). Furthermore, while int-VIII/int-IX binucleation occurs sporadically, int-I/int-II are reported to never undergo binucleation (Sulston and Horvitz, 1977). However, this *pho-1p::gfp* reporter strain also displayed frequent int-II binucleation (Figure 3.16). It is unclear if these occurrences represent natural variation within the *C. elegans* intestinal development or are the result of unknown mutations in the genetic background of the strain.

Figure 3.14. A. The 425 bp long *pho-1* promoter. B. Diagram of *pho-1p*(LIN-14 KO)*::gfp* promoter containing indicated mutations to putative LIN-14 binding sites.

A



B

LIN-14 Site: GAACRY KO Site
-370 bp: GAACTG TCCATG
-261 bp: GAACAA TCCAAA
-63 bp: GAACTT TCCATT

Figure 3.15. Representative image of *pho-1p*(LIN-14 KO)*::gfp* expression. A. GFP channel. B. RFP channel. C. Merge with DIC. Scalebar is 50 µm.

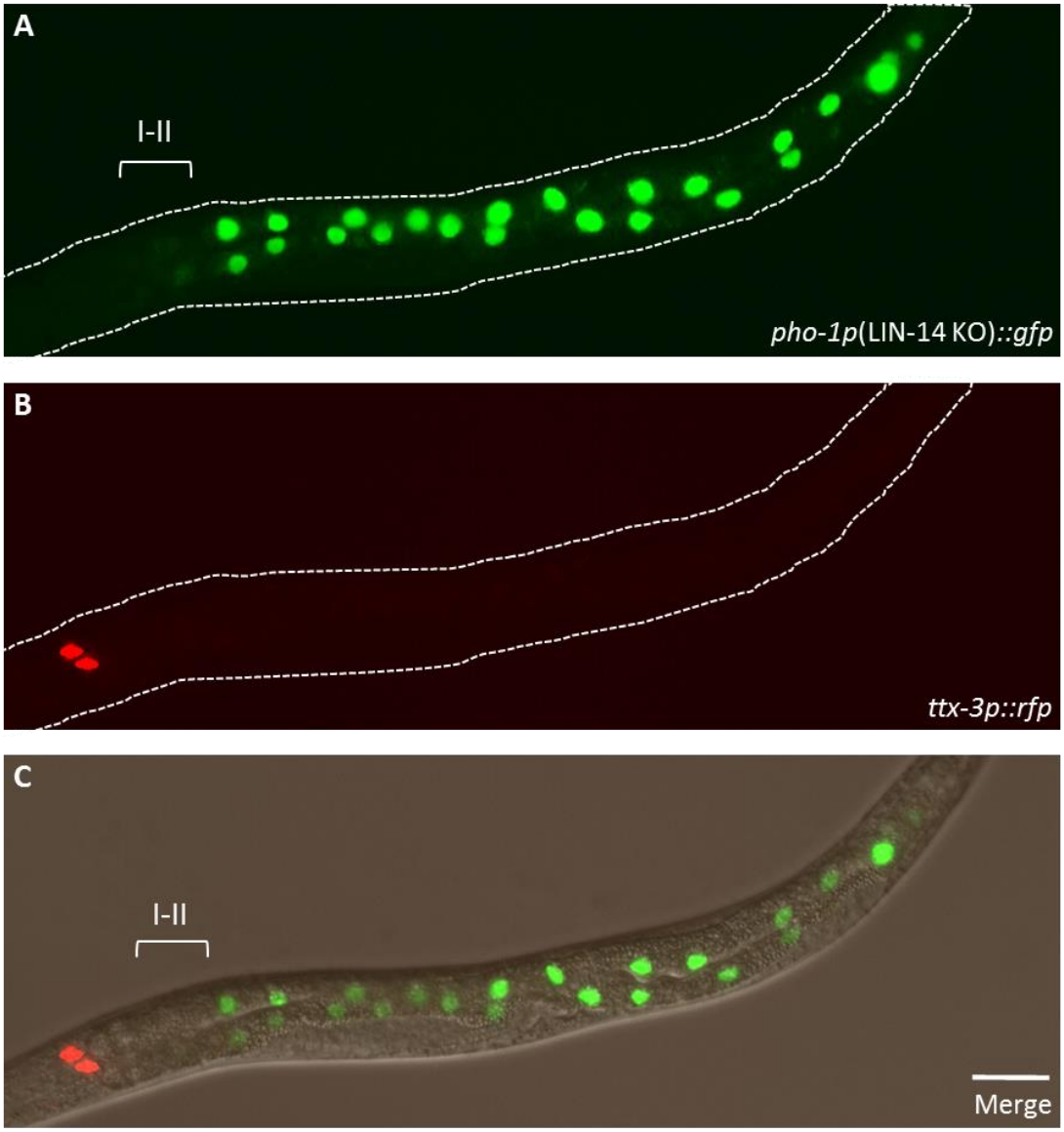
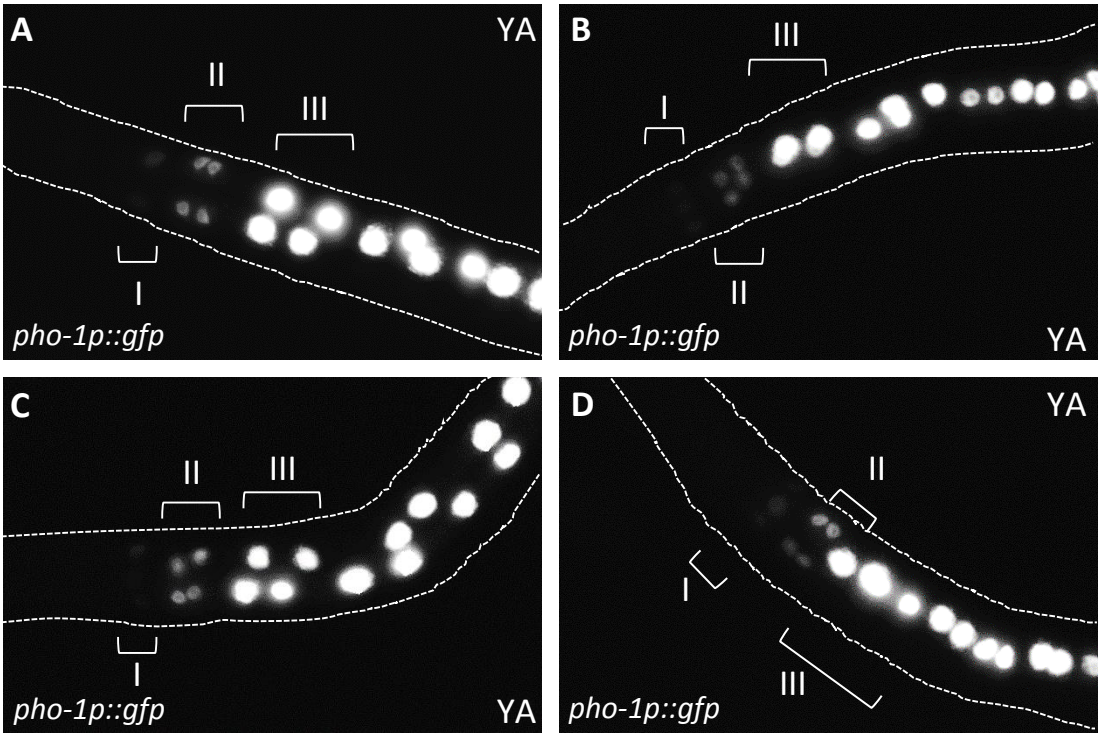

A
I-II
pho-1p(LIN-14 KO)::gfp
B
ttx-3p::rfp
C
I-II
Merge

Figure 3.16. Representative images of *pho-1* expression in young adults. Weak expression of int-II suggests that these cells have are binucleated.


A
YA
II
III
I
pho-1p::gfp
B
III
I
II
pho-1p::gfp
YA
C
II
III
I
pho-1p::gfp
YA
D
YA
II
I
III
pho-1p::gfp

3.5 Discussion

3.5.1 *pho-1* Expression is Spatially Patterned During the First Larval Stage

The spatial expression pattern of *pho-1* in the intestine had previously been reported to be present at all life stages of the animal (Fukushige et al., 2005). In contrast, the conclusion drawn from this present study is that *pho-1* spatial expression is initially un-patterned at the time of hatching and patterning is only completed near the end of the first larval stage. This difference could be due to the details of the experiments where the worms analyzed here were well fed upon hatching, compared to being starved for an unknown number of hours after hatching in the previous study. However, starving the hatched L1 worms and then analyzing the expression pattern of *pho-1* did not recapitulate the more complete spatial patterning observed in previous studies. It is possible these contrasting results are caused by differences in reporter constructs as the previous study used a lacZ reporter and here a GFP reporter was used. Futhermore, the stability of GFP could suggest anterior *pho-1* transcription is already being repressed before hatching.

Interestingly, int-IX *pho-1* expression was consistently weaker than other expressing intestinal rings at all stages of development. It is possible that this observed weaker expression is related to the repression in the anterior intestine since the *C. elegans* intestinal lineage is close to symmetrical. The weak expression observed in the anterior intestine after patterning (and int-II during patterning) is possibly a result of leaky expression from the multi-copy array. However, weak int-IX expression was typically stronger than int-I/int-II suggesting that endogenous *pho-1* is being expressed in these cells. Furthermore, actual *pho-1* patterning could be established earlier than the end of the first larval stage as the GFP reporter may persist longer in these cells beyond any PHO-1 protein. In this scenario, anterior *pho-1* expression would already be declining during late embryogenesis and be mostly patterned by early in the L1 stage. Future studies with a PHO-1 specific antibody should be able to resolve the temporal aspect of *pho-1* spatial patterning.

3.5.2 LIN-14 Regulates *pho-1* Spatial Patterning Indirectly

A number of lines of evidence support a role for LIN-14 in regulating *pho-1* spatial patterning in the intestine. First, the timing of this event correlates with LIN-14 expression, which peaks at the end of embryogenesis and persists until the end of the first larval stage. Secondly, disruption of the *lin-14* gene resulted in a more uniform expression intensity across the A/P axis of the intestine compared to controls. Increased anterior expression was never observed to be as intense as the more posterior expression, suggesting that other factors are contributing to the specific levels of transcription in these cells. This hypothesis is also supported by the fact that LIN-14 expression is ubiquitous in all intestinal rings, yet it appears to solely repress anterior *pho-1* expression and not posterior.

One possibility that would explain different apparent functions of LIN-14 along the A/P axis of the intestine could be the presence of *trans*-acting factors located in a subset of these cells. These factors could work in combination with ELT-2 and LIN-14 to regulate *pho-1* transcription. However, despite a conserved LIN-14 binding site within the previously identified minimal *pho-1* promoter (Fukushige et al., 2005), mutation of all three potential LIN-14 sites did not affect spatial patterning of the gene. Another possibility is that the reported LIN-14 binding site is incorrect and therefore the promoter knockout reporter constructs used here were too specific. Analysis of a deletion series of the *pho-1* promoter could help resolve this alternative hypothesis. However, the results described here indicate that LIN-14's regulation of *pho-1* spatial patterning likely occurs by an indirect mechanism, an idea which is further supported by the fact that the phenotype is present at later stages of development well after LIN-14 protein has disappeared. This is reminiscent of LIN-14's role in vulval precursor cell (VPC) specification during the L3 stage, well after LIN-14 protein is depleted (Li and Greenwald, 2010). There are 16 previously identified candidate targets of LIN-14, but the only known direct target of LIN-14 is the insulin gene *ins-33* (Hristova et al., 2005). Additionally, only one of these candidates encoded another transcription factor suggesting that LIN-14 does not act by regulating a few key factors that mediate effects on a larger number of genes. Thus, the general mechanism by which LIN-14 is involved

in various developmental processes remains unknown; yet it is clear that this protein can affect developmental events well after it has disappeared.

The unexpectedly high rate of binucleation of anterior int-II cells (which has not been reported until now) complicated the investigation of intestinal gene expression patterning. It is unclear if this is a naturally occurring event of wild-type *C. elegans* intestinal development or reflects an unknown mutant allele in the background of the *pho-1p::gfp* reporter strain. Binucleation of int-II cells was also observed in the *elt-2p::gfp* strain but at a noticeably lower frequency. *pho-1p::gfp* intensity appeared to be stronger in these unusual int-II nuclei indicating a potential correlation between *pho-1* expression and the number of nuclei in a cell. On the other hand, it is possible that int-II cells not expressing the reporter were also binucleated, which would suggest no link between the two events. It is certainly an attractive hypothesis that the mechanism by which LIN-14 promotes posterior cell binucleation is fundamentally entwined with that of *pho-1* transcriptional regulation. It would be interesting to see if other known regulators of intestinal binucleation also are involved in *pho-1* spatial patterning.

Chapter 4: Deciphering the Contributions to Intestinal Gene Expression in *C. elegans* by the GATA Transcription Factors ELT-2 and ELT-7

4.1 Abstract

The *C. elegans* intestine is produced from a single blastomere, which is specified by the redundant GATA transcription factors END-1 and END-3. END-1/END-3 are transiently expressed during development and activate transcription of the GATA factors ELT-2 and ELT-7. These two GATA factors activate terminal differentiation genes important for organ function. Null mutations in *elt-2* result in larval arrest after hatching due to a non-functional intestine. In contrast, *elt-7* null mutants have no obvious phenotype. The present model of how intestinal gene expression is mediated holds that ELT-2 regulates most intestinal genes, with ELT-7 functioning redundantly. The aim of this study was to better understand what the roles of ELT-2 and ELT-7 are in the intestine. RNAseq was performed on starved L1 larvae of different genetic backgrounds, including the *elt-2* single mutant, *elt-7* single mutant, and the *elt-7; elt-2* double mutant. The results show that ELT-2 appears to activate most intestinal genes independently of ELT-7. Furthermore, no genes were observed to be expressed at very low levels in the *elt-7* mutants, whereas there are examples of genes that critically depend on ELT-2 for expression. There are a number of genes found to be regulated redundantly by either ELT-2 or ELT-7 but only a few genes that appear to require both factors in an additive manner. There is no evidence of synergistic action by the ELT-2 and ELT-7 proteins. Genes that appear to be regulated by ELT-7 and/or ELT-2 have significantly more *cis*-regulatory TGATAA sites in their proximal promoters. Analysis of these sites hints at the possibility that these two GATA factors have slightly different binding preferences. Two noteworthy results are described, first that both ELT-2 and ELT-7 are implicated in the repression of intestinal genes and second, that *elt-7* mutants complete embryogenesis more rapidly than wild-type animals do. Finally, there are almost one hundred other transcription factors expressed in the intestine that are differentially expressed to some extent in the various genetic backgrounds. That is, intestinal gene regulation is likely to be more complex than simply activation by these two GATA factors. Taken together, the results support the hypothesis that ELT-2 is the primary GATA transcription factor regulating intestinal transcription, while ELT-7 has a mainly redundant role during

embryogenesis. A model of intestinal gene regulation is described in which ELT-2 outcompetes ELT-7 for binding to promoters of target genes that it exclusively regulates. In contrast, genes regulated by both ELT-2 and ELT-7 have enough sites for both factors to bind. Future work should aim to rigorously test this model.

4.2 Background

4.2.1 Three GATA Transcription Factors are Expressed in the Intestine After Hatching

Specification of the *C. elegans* E cell by the GATA transcription factors END-1 and END-3 ultimately results in the activation of three other GATA factors: ELT-2, ELT-7 and ELT-4. These three GATA factors are expressed exclusively in the intestine beginning in embryogenesis and continuing throughout the life of the worm. Expression of the *elt-7* gene occurs first at the 2E cell stage (Fukushige et al., 1998; McGhee, 2007; Zhu et al., 1998) (T. Wiesenfahrt, personal communication), while transcription of the *elt-2* gene is activated by a combination of END-1, END-3 and ELT-7 about the time of the 4E cell stage (Fukushige et al., 1998; Zhu et al., 1998). Maintenance of ELT-2 and ELT-7 expression in the intestine is mediated by a positive auto-regulatory loop where these factors activate their own transcription and each other's (Maduro and Rothman, 2002; McGhee, 2013). Expression of the *elt-4* gene on the other hand does not turn on until about the 1.5 fold stage of embryogenesis (Fukushige et al., 2003). The *elt-4* locus is located just upstream of *elt-2*, essentially encoding only a duplication of the single *elt-2* zinc finger, and lacks any clear function in *C. elegans* (Fukushige et al., 2003). This GATA factor does not appear to regulate any intestinal genes, has no phenotype when completely knocked out, is not present in other *Caenorhabditae* species and it has been hypothesized that this gene will eventually disappear from the *C. elegans* genome (Fukushige et al., 2003). Similarly, *elt-7* null mutants have no obvious phenotype and double null *elt-7; elt-4* mutants have no effect on differentiation or the function of the intestine (McGhee et al., 2009). Null mutations in *elt-2* produce worms that hatch but immediately arrest with a gut obstructed (Gob) phenotype (Fukushige et al., 1998). This suggests that of the three GATA factors whose expression persists from embryogenesis through adulthood, ELT-2 is the only necessary GATA transcription factor.

4.2.2 Regulation of Intestinal Gene Expression

A substantial number of intestinally expressed genes have been identified both in gene specific studies (Fukushige et al., 2005; Marshall and McGhee, 2001) and in large scale screening efforts (Blazie et al., 2015; McGhee et al., 2007; McGhee et al., 2009; Pauli et al., 2006). Many genes that are expressed exclusively or primarily in the intestine have been analyzed at the promoter level for *cis*-regulatory elements and the *trans*-acting factors that bind to them. Genes such as *mtl-1*, *mtl-2*, *ges-1*, *pho-1*, *ref-1*, *tmy-1* and *ftn-1* have all been shown to require ELT-2 binding to GATA sites in their promoters in order to activate transcription (Anokye-Danso et al., 2008; Fukushige et al., 2005; Marshall and McGhee, 2001; Moilanen et al., 1999; Neves et al., 2007; Roh et al., 2014; Romney et al., 2008). Additionally, extensive analysis of *elt-7*; *elt-4* double null mutants revealed no phenotype in a variety of assays including percentage of adults laying eggs, number of eggs laid, percentage of embryos hatching, adult size and cycle time for defecation (McGhee et al., 2007). A six base pair sequence of TGATAA has been identified as the only significantly over-represented *cis*-regulatory element found in the promoters of intestinal genes (McGhee et al., 2009). ELT-2 binds strongly to these TGATAA sequences and is capable of ectopically activating expression of intestinal genes when ectopically expressed itself (Fukushige et al., 1998). Therefore, ELT-2 seemingly is able to mediate the differentiation and maintenance of the intestine as an organ, in the absence of the other two GATA factors. It is hypothesized that ELT-2 plays some role in the regulation of most if not all intestinal gene expression (McGhee et al., 2009).

This leads to the question of what is the role of ELT-7 in the intestine if this GATA transcription factor does not appear to be necessary to create and maintain a functional intestine. ELT-7 has been reported to be involved in regulating some intestinal genes, those that are reported to be regulated by both ELT-2 and ELT-7 (Murray et al., 2008; Sommermann et al., 2010). The Sommermanan et al. (2010) study assessed the regulation of eight intestinally expressed (in several cases other tissues as well) genes and found that seven are down-regulated in *elt-7*; *elt-2* double mutants, but are not affected in either single mutant. The authors concluded that their results indicated a “profound synergy” between ELT-2 and ELT-7 to facilitate making of a functional intestine, although this is easily misinterpreted. The authors describe observations of synergy

between the loss of function mutations in *elt-2* and *elt-7* and are not suggesting that these two proteins function synergistically to activate intestinal gene transcription. It has also been reported that ELT-7 can drive transdifferentiation of other tissues such as the pharynx (Riddle et al., 2013). Ectopic ELT-7 activated expression of ELT-2 as well as intestinal gene reporters, producing cells that resemble those in the intestine instead of the pharynx (Riddle et al., 2013). This feature of ELT-7 activity during embryogenesis is also seen with ELT-2, although ELT-7 can mediate this effect during a much larger temporal window. It should be noted, however, that in all cases of ELT-7 mediated transdifferentiation, ELT-2 was rapidly activated and accumulated in these cells. Thus, it is not strictly clear if it is the action of ELT-7 or ELT-2 or both factors together that mediates such a phenomenon.

GATA transcription factor hierarchies and redundancy are a hallmark of endoderm development in many organisms including *Drosophila* (Murakami et al., 2005), zebrafish (Tseng et al., 2011) and mice (Carrasco et al., 2012; Zheng et al., 2013). There are two evolutionarily related families of GATA transcription factors in vertebrates: the GATA-1/2/3 family that functions as a sequential hierarchy of GATA factors in hematopoietic stem cell differentiation (Shimizu and Yamamoto, 2005) and the GATA4/5/6 family that functions in endoderm development (Carrasco et al., 2012; Zheng et al., 2013) and heart development (Mohun and Sparrow, 1997; Zheng et al., 2003). In mice, it has been reported that GATA2 binds to promoters in undifferentiated erythroid cells but switches off to GATA1 as differentiation occurs (Suzuki et al., 2013). In mammals, GATA-4 and GATA-6 have been shown to function completely redundantly in pancreas development (Carrasco et al., 2012) and liver development (Zheng et al., 2013). Remarkably, one of the *C. elegans* GATA transcription factor that specifies the E cell, END-1, can initiate endoderm development in *Xenopus* (Shoichet et al., 2000). This features suggests conservation of GATA factor function between these organisms. In nematodes, there has been extensive duplication and divergence of GATA factors (Gillis et al., 2008). In *C. elegans* there are eleven GATA factors, ten of which belong to the GATA-4/5/6 family and only one, *elt-1*, is a member of the GATA-1/2/3 family (Gillis et al., 2008). Intriguingly, all ten GATA-4/5/6 family members in *C. elegans* have a single zinc finger whereas vertebrate family members have two. ELT-2 is

most closely related to the vertebrate GATA-4 transcription factor whereas ELT-7 is less similar to vertebrate factors, orthologous to human GATA-2 (Yook et al., 2012). Orthologs of the *elt-7* gene appear to be present in the nematodes *Pristionchus pacificus* of the Diplogastridae family and the parasitic *Haemonchus contortus*.

The goal of this study was to identify how intestinal gene expression is regulated by ELT-2 and ELT-7, ultimately to ascertain the contribution that each GATA factor makes in the intestine. The experiments aimed to discover the genes that are regulated by each factor, either exclusively by one or the other, or in combination. Based on all the evidence to date, it is hypothesized that ELT-2 regulates more intestinal genes than ELT-7, which is partially redundant with ELT-2.

4.3 Materials and Methods

4.3.1 *C. elegans* Strains

Standard nematode handling conditions were used (Brenner, 1974). Animals were grown at 20°C unless otherwise indicated. Strains used were wild type N2, JM147 *elt-2(ca15) X; caEx3[elt-2(+),sur-5p::gfp, rol-6(su1006)]*, JM199 *elt-7(tm840) V; elt-2(ca15) X; caEx3[elt-2(+),sur-5p::gfp, rol-6(su1006)]*, JM222 *elt-7(tm840) V* (outcrossed from JM199) and JR2132 *wIs126[elt-7p::gfp::lacZ]*.

4.3.2 COPAS Biosorting and RNA Extraction

Worms from starved NGM plates were chunked to fresh 150 mm diameter plates (OP50) and allowed to develop at 20^{o}C until most were gravid adults. The gravid adults were washed off the plates (15 ml of 1X M9 buffer into 50 ml Falcon tubes) and centrifuged five minutes at ≤ 1000 rpm (20^{o}C). The supernatant was removed and the worms were washed in another 50 ml of 1X M9 buffer. Animals were re-suspended in 5 ml of dH_2O and a 17.5 ml alkaline bleach solution was added (5ml 6% hypochlorite with 12.5 ml 1M NaOH). Worms were incubated in the bleach solution four to six minutes until approximately 50% of gravid adults had burst open. This solution was then added to an equal volume of an alkaline bleach stop solution (0.95 M Tris-HCl with 0.05 M Tris Base), vortexed briefly and centrifuged five minutes at ≤ 1000 rpm (20^{o}C). The

supernatant was removed, the embryos were washed with 50 ml of 0.45 μm filtered 1X M9 buffer, centrifuged again, then re-suspended in 50 ml of filtered 1X M9 buffer in a 75 cm^2 tissue flask and allowed to hatch overnight at room temperature on a shaker.

The next morning (12-14 hours after bleaching), starved L1 worms were passed through a 0.27 μm nylon mesh twice and placed on a shaker in 50 ml falcon tubes. Filtered worms were continually passed through a Union Biometrica Complex Object Parametric Analyzer and Sorter (COPAS) platform with Biosort Device which has a 488 nm excitation filter and is capable of sorting green larvae from non-green. A total of 50,000 GFP positive and 50,000 GFP negative worms were isolated for JM147 and JM199 per sort. In the case of N2 and *elt-7(tm840)*, 50,000 worms were collected from the sorter and an unknown number (>100 000) of worms were collected from filtered but not sorted populations. All worms were stored in 50 ml falcon tubes in 1X M9 buffer on the shaker until the entire sample was collected. The worms were centrifuged five minutes at ≤ 1000 rpm 20^{o}C, pipetted into an RNase free tube and centrifuged at 3000 rpm for three minutes at room temperature. The supernatant was removed, 1 ml of TRizol and 10 μl of β-mercaptoethanol were added, the sample was mixed by inversion, flash frozen in liquid nitrogen and stored at -80^{o}C. Total RNA extractions were performed using the Ambion PureLink RNA Mini Kit. Quality and concentration of RNA samples were assayed using an Agilent Tapestation. Images of sorted worms were captured using a Zeiss Axioplan2i microscope with a Hamamatsu Orca digital camera and AxioVision (version 4.8.1) software. Length calibration was achieved using a Pyser-SGI micrometer slide. Body lengths of larvae were measured in ImageJ using segmented lines (Collins, 2007); t-tests were performed in Sigma Plot version 12.5.

4.3.3 Library Preparation and RNAseq (U of Calgary Core DNA Services)

Poly(A) mRNA isolation was performed as per the standard method for Life Technologies Dynabeads mRNA Direct Kit (part #61021). Whole transcriptome libraries were prepared as per Life Technologies Solid Total RNA-Seq kit standard input method (part #4445374) using chemical fragmentation and Array Script reverse transcriptase. Sequencing was performed on a Life Technologies 5500xl genome analyzer on two

separate runs. The first run was of 24 samples across three lanes and the second run was of eight samples on one lane.

4.3.4 Bioinformatic Analysis of RNAseq Results

Raw Sequence reads were mapped by Dr. Paul Gordon to the WS220 ce10.20130317.gtf version of the *C. elegans* genome using Lifescope ver. 2.5, all other bioinformatics were performed by myself. Quality control of sequence reads was assessed using FastQC version 0.52. Mapped reads were counted using HTSeqCount version 0.4.1 (Anders et al., 2014) under union mode with minimal alignment quality set to ten. Differential expression analysis including MA plot creation was performed in R Studio version 3.1.1 using the R Bioconductor package DESeq2 version 1.6.2 (Love et al., 2014). FPKM values for transcripts were generated using Cufflinks ver. 0.0.6 (Trapnell et al., 2010). Data was visualized using the BAM to BigWig converter (Kent et al., 2010) and subsequent uploading to the UCSC genome browser (Kent et al., 2002). Promoter sequences for genes of interest were obtained using the RSAT Retrieve EnsEMBL seq tool (Thomas-Chollier et al., 2011). Statistical analysis of promoter features (Mann-Whitney rank sum test) was performed in Sigma Plot version 12.5. Euler diagrams were created using Euler Ape version 3.0 (Micallef and Rodgers, 2014).

4.3.5 Western Blot for ELT-2

5000 starved L1 larvae used in each sample were obtained using the COPAS Biosort as described above. Samples were boiled 10 minutes in 1x protein loading buffer, run on a 6% stacking gel and 8% separating gel for one hour at 100 V. Gel contents were transferred to a PVDF membrane for 80 minutes, blocked with 5% milk for 2 hours, probed with the 455-2A4 ELT-2 specific monoclonal antibody and MH16 UNC-15 specific antibody overnight at 4^{o}C. The membrane was washed 3x in TBST, 2x in TBS, then incubated in HRP secondary antibody for 2 hours at room temperature. The membrane was washed as before and stained with GE ECL chemiluminescent western blotting detection kit and detected using a LAS4000 Imaging Station (GE Healthcare). Quantitation was performed in ImageJ.

4.4 Results

4.4.1 Isolation of Genetically Distinct *C. elegans* Populations

RNAseq was performed on polyA+-mRNA extracted from synchronized *C. elegans* larvae to better understand how transcription in the *C. elegans* intestine is mediated by ELT-2 and ELT-7. These larvae were hatched in the absence of food and thus were arrested at the first larval stage. The four strains sequenced in this study were: wild-type N2, *elt-7(tm840)* null mutants, *elt-2(ca15)* null mutants and *elt-7(tm840); elt-2(ca15)* double mutants. The *elt-7(tm840)* null mutants harbour a 616 bp deletion spanning the second to third exons that removes the first 22 amino acids (out of 25) of the DNA binding zinc finger (Sommermann et al., 2010) (Figure 4.1A). The *elt-2(ca15)* null mutants have a 2.231 kb deletion that removes the entire *elt-2* coding region (Figure 4.1A) (Fukushige et al., 1998). Both the single *elt-2* mutants and the double mutants are rescued by an extra-chromosomal array containing ~5 kb of the *elt-2* upstream region with the endogenous *elt-2* genomic locus, the dominant *rol-6(su1006)* transgenic marker and a *sur-5p::gfp* marker (McGhee et al., 2009). The rescuing transgene is present (presumably hundreds of copies) as an extra-chromosomal array in these strains and leads to segregation of both GFP-positive rescued larvae that develop to produce offpsring and L1 arrested mutant worms that are GFP-negative (Table 4.1).

Two methods were used to isolate six genetically distinct populations from these four strains, subsequently resulting in eight distinct populations of arrested L1 larvae whose mRNA was sequenced. In the first method, a 27 µm nylon mesh was used to twice filter the newly hatched wild-type and *elt-7(tm840)* worms to remove adult carcasses, approximately three hours after hatching. The COPAS Biosort was used in the second method to isolate rescued GFP-positive worms and GFP-negative worms from the *elt-2* single and double mutant strains (Figure 4.1B, see materials and methods). 50, 000 L1 larvae were isolated in each sort based on time of flight (tof, size) and GFP fluorescence, two parameters optimized for each strain through previous sorts (Figure 4.2) (McGhee et al., 2009). The sorting resulted in pure populations with an approximate 1% false positive rate (Figure 4.2). Wild-type and *elt-7* mutant populations were collected in the same manner, even though they already existed as a homogenous population. This was used as

Figure 4.1. A. Diagram of the *elt-7* and *elt-2* gene loci including location of the respective null deletion mutations *tm840* and *ca15*. B. Schematic depicting the experimental procedure using the COPAS Biosort for obtaining pure populations of starved L1 larvae for each genetic background. C. Flow chart depicting the bioinformatic analysis pipeline to identify differentially expressed genes.

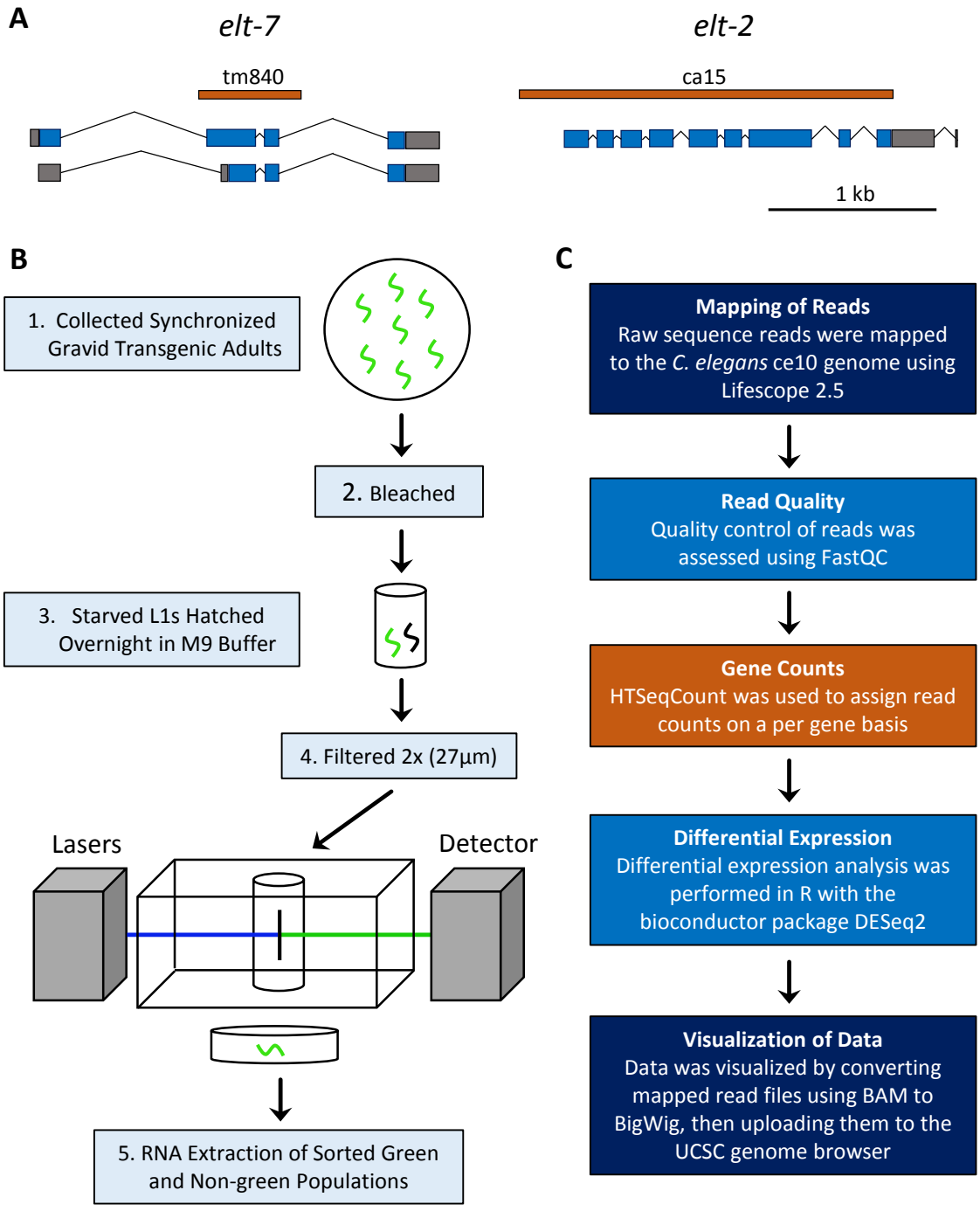
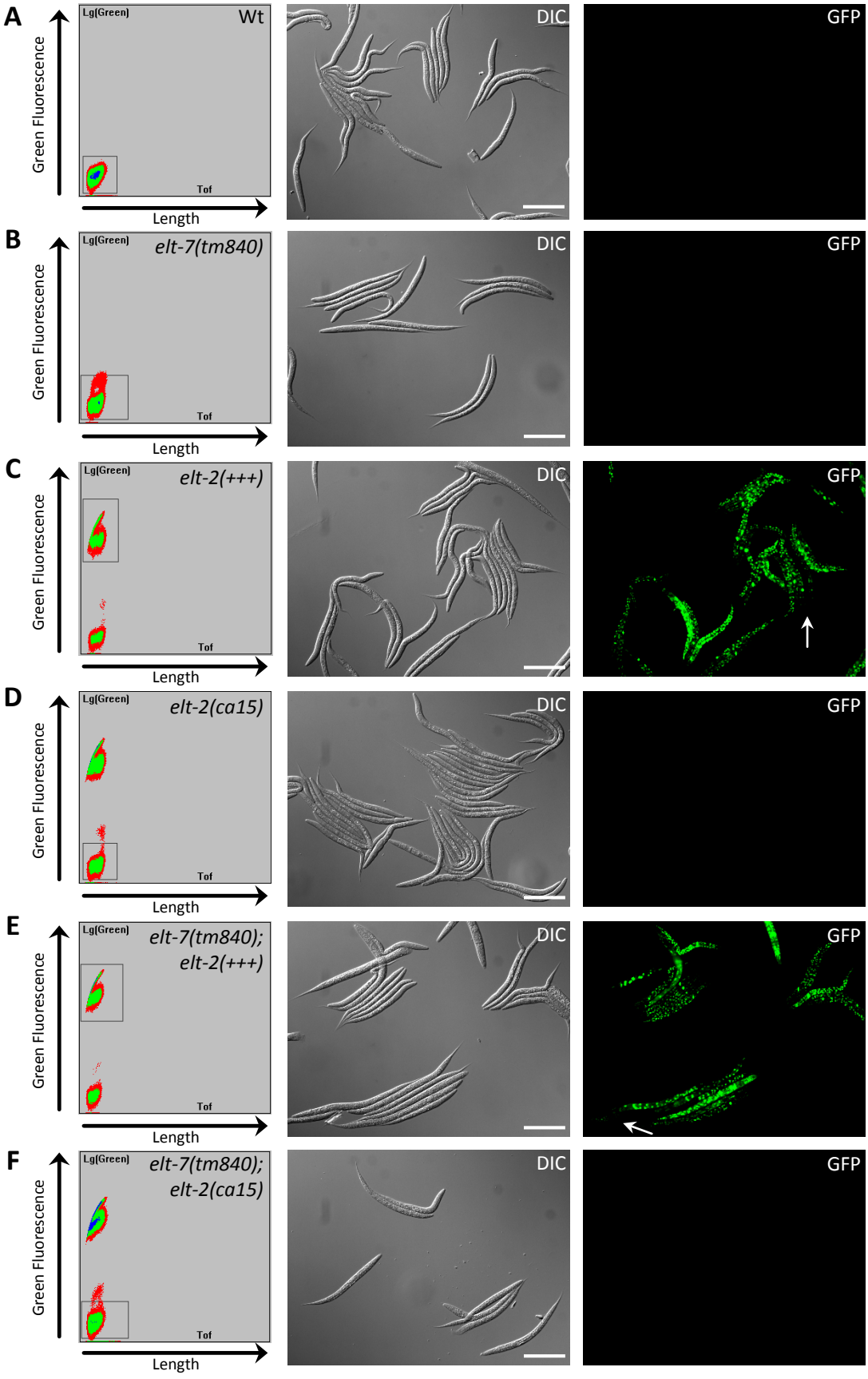

A
elt-7
tm840
elt-2
ca15
1 kb
B
1. Collected Synchronized Gravid Transgenic Adults
2. Bleached
3. Starved L1s Hatched Overnight in M9 Buffer
4. Filtered 2x (27µm)
Lasers
Detector
5. RNA Extraction of Sorted Green and Non-green Populations
C
Mapping of Reads
Raw sequence reads were mapped to the C. elegans ce10 genome using Lifescope 2.5
Read Quality
Quality control of reads was assessed using FastQC
Gene Counts
HTSeqCount was used to assign read counts on a per gene basis
Differential Expression
Differential expression analysis was performed in R with the bioconductor package DESeq2
Visualization of Data
Data was visualized by converting mapped read files using BAM to BigWig, then uploading them to the UCSC genome browser

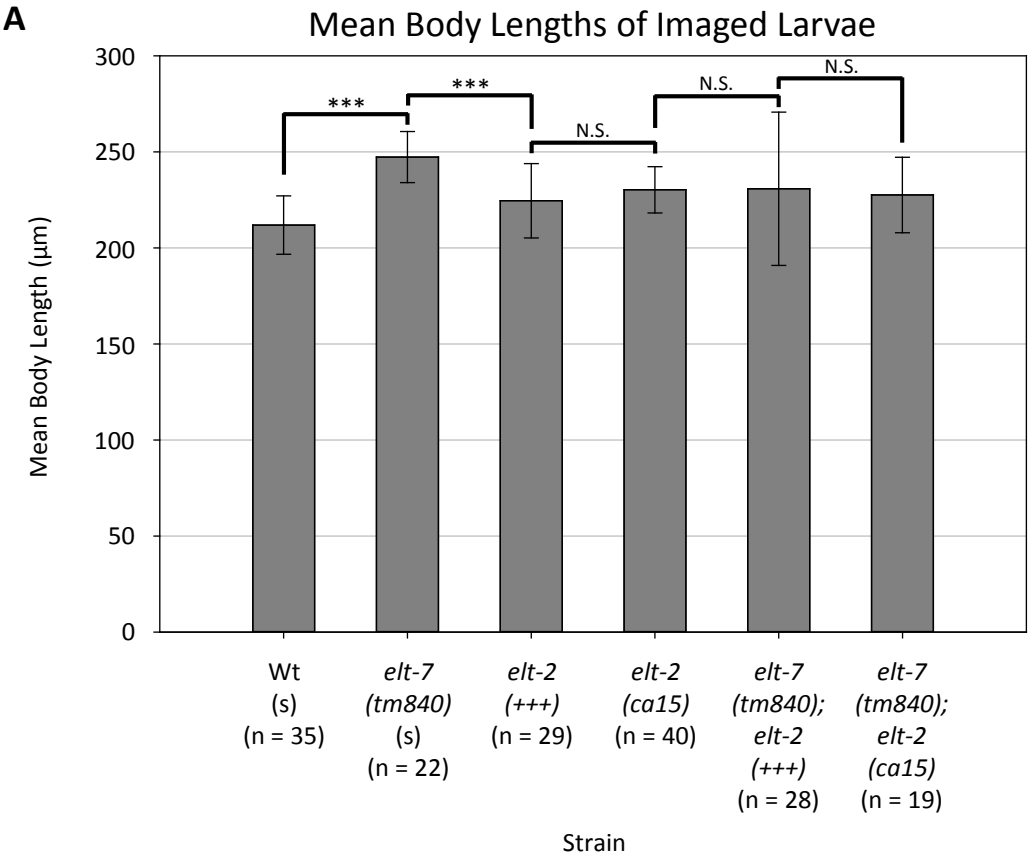
Figure 4.2. A-F. Left: COPAS Biosort graph depicting size and green fluorescence of *C. elegans* L1 larvae passed through the machine. Box indicates parameters for sorting; population being sorted indicated in the top right hand corner of COPAS graph. Right: DIC and GFP channel images of sorted larvae from corresponding genetic backgrounds. Arrows in C and E indicate larvae that are not GFP-positive. GFP exposures were 450 ms for each image. Scale bar is 100 µm.

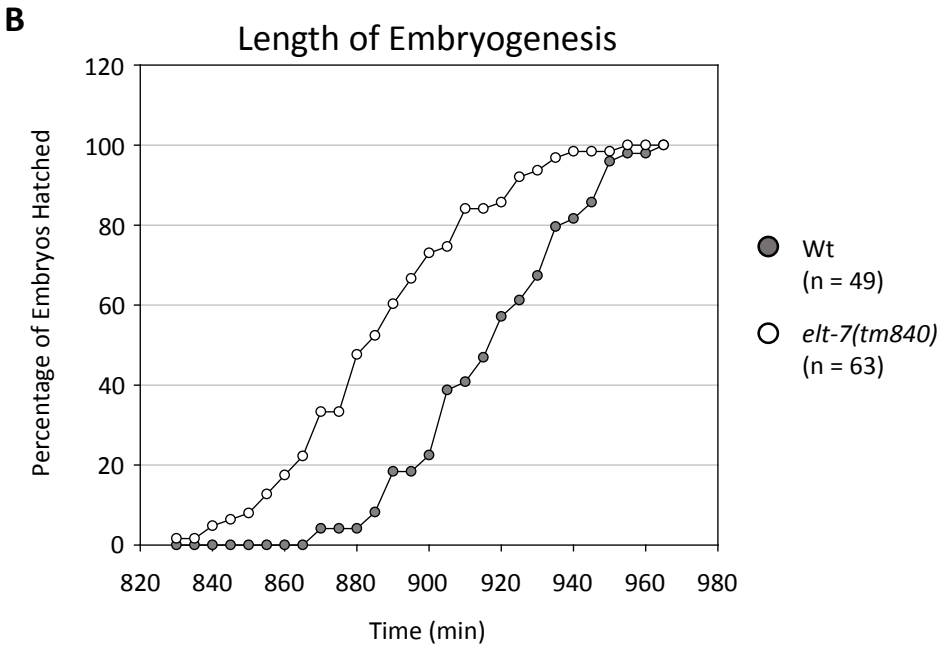

A
Lg(Green)
Wt
Tof
Green Fluorescence
Length
DIC
GFP
B
Lg(Green)
elt-7(tm840)
Tof
Green Fluorescence
Length
DIC
GFP
C
Lg(Green)
elt-2(+++)
Tof
Green Fluorescence
Length
DIC
GFP
D
Lg(Green)
elt-2(ca15)
Tof
Green Fluorescence
Length
DIC
GFP
E
Lg(Green)
elt-7(tm840);
elt-2(+++)
Tof
Green Fluorescence
Length
DIC
GFP
F
Lg(Green)
elt-7(tm840);
elt-2(ca15)
Tof
Green Fluorescence
Length
DIC
GFP

a control for any effects that sorting could have on gene expression in these worms. Two sets of 50, 000 sorted larvae were pooled together to create a single biological replicate of 100, 000 L1 larvae for each sorted population. Approximately 1000 *elt-2(+++)* worms harbouring multiple copies of the rescuing transgene were present in the GFP-negative populations for both the *elt-2(ca15)* single mutant and the *elt-7(tm840); elt-2(ca15)* double mutant populations. The same is true for the GFP-positive strains where GFP-negative worms were present at an estimated frequency of 1% as seen in Figure 4.2C, E. This could not be avoided; however, these larvae only represent a small fraction of the worms used for RNA extraction and was not predicted to greatly affect the results.

The wild-type, *elt-7(tm840)* and GFP-positive sorted populations were verified to be arrested at the L1 stage by comparing the mean larval lengths to that of the arrested GFP-negative mutants. This was initially assayed in real-time by comparing the tof parameter on the COPAS Biosort, an estimation of worm length as larger worms will take more time to travel through the scanning apparatus. Comparing the GFP-positive and GFP-negative populations revealed that there was no difference in tof for these populations (Figure 4.2). Additionally, sorted worms were imaged under the microscope and measured for body length. The results in Table 4.2 and Figure 4.3A show that there are differences in mean length between some populations, but no population was larger than the 250 µm reported for wild-type L1 length after hatching (Mörck and Pilon, 2006). Any difference in L1 size between the isolated populations was considered minor despite being statistically significant. Similar variation has been observed with different feeding defective mutant strains that only have a biologically relevant body length phenotype at later larval stages (Mörck and Pilon, 2006). Four out of the six populations were found to be not significantly different from each other (Table 4.2, Figure 4.3A). Interestingly, the *elt-7(tm840)* mutants were the longest and the most different from wild-type (Figure 4.3A). This may be related to the finding that the *elt-7(tm840)* mutants complete embryogenesis and hatch approximately 30 min faster than wild-type ($p < 0.001$) (Figure 4.3B). This unpredicted phenotype of a strain for which none had previously been described suggests that ELT-7's primary function could be during embryogenesis and the early stages of the E lineage.

Figure 4.3. A. Mean body length (± SD) measurements for sorted populations. Statistically significant differences in mean body lengths between populations are apparent. * denotes $p < 0.001$, N.S. indicates no difference. Corresponding data can be found in Table 4.2. B. Time in minutes for wild-type and *elt-7(tm840)* mutant embryos isolated at the one to four cell stage to hatch at 20ºC. *elt-7(tm840)* mutant embryos completed embryogenesis on average about 3.3% faster than wild-type embryos ($p < 0.001$).**


A
Mean Body Lengths of Imaged Larvae
Mean Body Length (µm)
300
250
200
150
100
50
0
***
***
N.S.
N.S.
N.S.
Wt (s) (n = 35)
elt-7 (tm840) (s) (n = 22)
elt-2 (+++) (n = 29)
elt-2 (ca15) (n = 40)
elt-7 (tm840); elt-2 (+++) (n = 28)
elt-7 (tm840); elt-2 (ca15) (n = 19)
Strain


B
Length of Embryogenesis
Percentage of Embryos Hatched
120
100
80
60
40
20
0
820
840
860
880
900
920
940
960
980
Time (min)
Wt (n = 49)
elt-7(tm840) (n = 63)

It has previously been reported that between a third and a half of protein coding gene expression is affected by nutritional status of the worm (Baugh et al., 2009). Starvation inducing L1 arrest was found to have a complete gene expression response within three to six hours after hatching without available nutrients, an expression pattern that is mostly initiated within an hour and maintained from six to 24 hours after hatching (Baugh et al., 2009). In the present study, the filtered wild-type and filtered *elt-7(tm840)* samples were obtained from larvae that had been starved for approximately three hours after hatching, after the majority of gene expression changes caused by starvation would have already occurred. Furthermore, RNA was extracted from the filtered wild-type and *elt-7(tm840)* samples at the same time point during this transition and even if starvation induced changes in gene expression were occurring, they should be similar between these two sets of samples. RNA extraction from the sorted population samples was not performed until the completion of each sort, which began approximately three hours after the larvae were hatched in the absence of food. Wild-type or *elt-7(tm840)* strains took approximately four hours from sorting start to RNA extraction, for a total time at RNA extraction of about seven hours post-hatching. GFP-negative populations were sorted first for a total time at RNA extraction of about seven to eight hours post-hatching, while GFP-positive populations were sorted second for a total time of approximately 11 hours. All of the sorted samples were obtained between seven and 11 hours post-hatching in the absence of food, at which point all gene expression changes due to starvation arrest should be complete and stable (Baugh et al., 2009). That is, the differentially expressed genes between the sorted populations are expected to be due to the differences in the genetic backgrounds, not their duration of starvation.

4.4.2 General Findings of the RNAseq

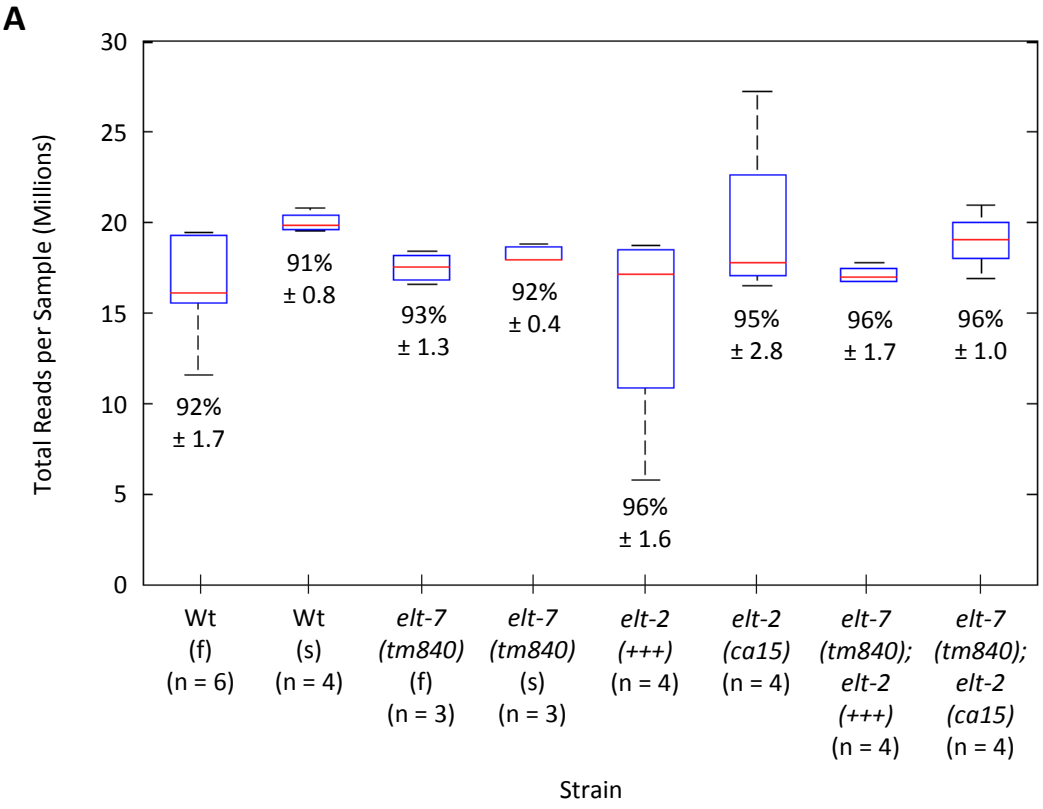
A total of 32 samples were sequenced from the eight different populations, the number of biological replicates for each population ranged from three to six. Life Technologies' 5500xl SOLiD system was used to generate paired-end 50 bp sequence reads in two separate runs, one with 24 samples across three lanes and the second with eight samples on one lane. The mean total number of sequence reads per population varied from 14.7 million to 20 million, slightly higher on average than the target of 15

million reads per sample (Figure 4.4A, Table 4.3). The option to sequence to a greater depth at the cost of biological replicates in order to identify low expressed transcripts within each dataset was considered. However, it has been reported that increasing the number of biological replicates significantly increases the statistical power to identify differences in gene expression, much more than increasing sequencing depth, particularly once sequencing depth reached 20 million reads (Travers et al., 2014). The goal of this study was to identify differentially expressed genes known to be primarily expressed in the intestine for the purpose of comparing ELT-2 and ELT-7 functions. Numerous previous studies have identified many intestinally expressed genes by a variety of methods (Blazie et al., 2015; McGhee et al., 2007; McGhee et al., 2009; Pauli et al., 2006). As such, more biological replicates were performed at a lower sequencing depth with the aim of minimizing sample to sample variation, thereby increasing the power to detect differences between ELT-2 and ELT-7 regulation of intestinal targets.

An overview of the data analysis pipeline that was used to map, assess the quality and determine differentially expressed genes can be found in Figure 4.1C. The *C. elegans* genome version ce10 was used to map sequenced reads for each of the 32 samples. Mean mapping percentage per population was between 91% and 96% (Figure 4.4A, Table 4.3). FastQC was used to check sequence reads for quality control such as sequence duplication, per base sequence quality and per N content. This analysis did not result in any concerns about the quality of the sequence reads from the samples.

The number of transcripts that were expressed (mean Fragments Per Kilobase of exon per Million fragments mapped (FPKM) value greater than 0.5) was examined to determine if there were any gross differences in gene expression between the different populations. As seen in Figure 4.4B, out of the approximately 48 000 currently annotated transcripts in the *C. elegans* genome, about 20 000 transcripts were expressed in each of these starved L1 populations. Only the double mutant populations stand out with slightly more detected transcripts on average than the other populations, but many of these transcripts were expressed at a very low level as evidenced by the increased error size. Power law plots illustrate that there is no difference in the number of transcripts expressed at a given FPKM value between biological replicates within a population or between different populations (Figure 4.5).

Figure 4.4. A. Box plots of the total number of sequence reads (millions) obtained from each population. Numbers under boxes represent mean percentage of reads ± SD mapped to the *C. elegans* genome per population. Number of biological replicates per sample indicated under the genetic background. (f) indicates samples obtained via filtering alone, (s) by sorting (see methods), all other samples not specifically indicated were obtained by sorting. Corresponding data can be found in Table 4.3. B. Bar graph displaying the mean number of transcripts (thousands) detected as expressed in each genetic background. Error bars were calculated as the number of transcripts no longer expressed after subtracting one SD from their mean expression value.


A
Total Reads per Sample (Millions)
30
25
20
15
10
5
0
92%
± 1.7
91%
± 0.8
93%
± 1.3
92%
± 0.4
96%
± 1.6
95%
± 2.8
96%
± 1.7
96%
± 1.0
Wt (f) (n = 6)
Wt (s) (n = 4)
elt-7 (tm840) (f) (n = 3)
elt-7 (tm840) (s) (n = 3)
elt-2 (+++) (n = 4)
elt-2 (ca15) (n = 4)
elt-7 (tm840); elt-2 (+++) (n = 4)
elt-7 (tm840); elt-2 (ca15) (n = 4)
Strain

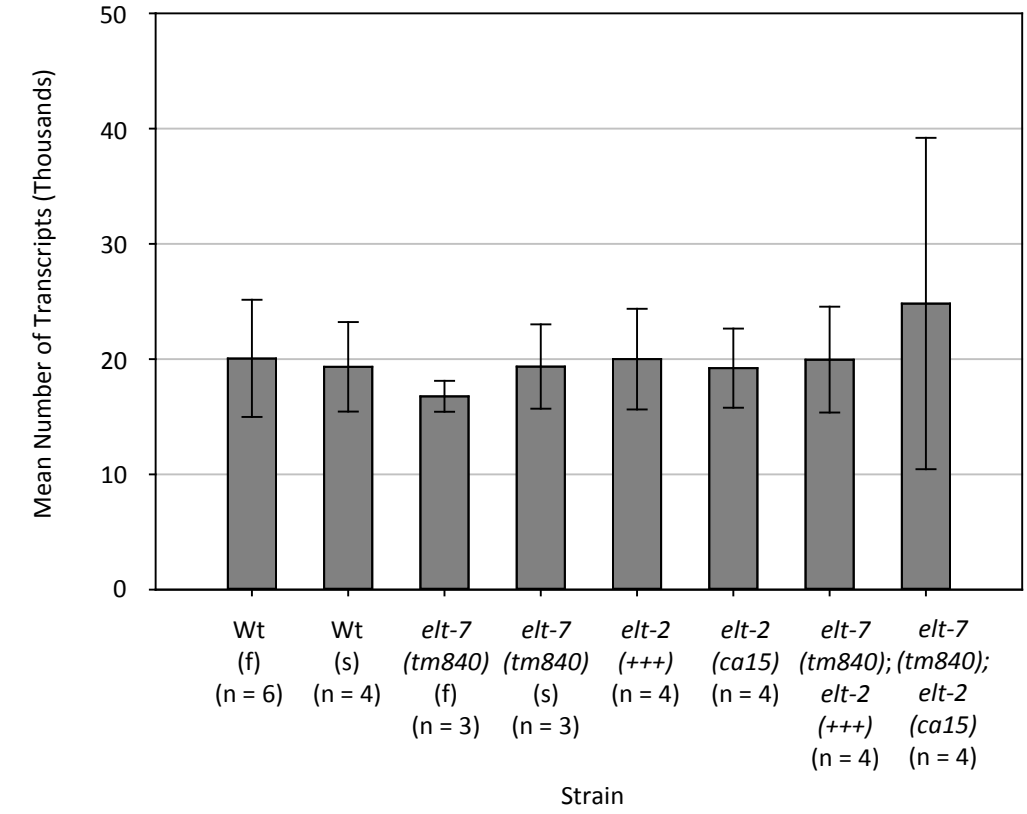
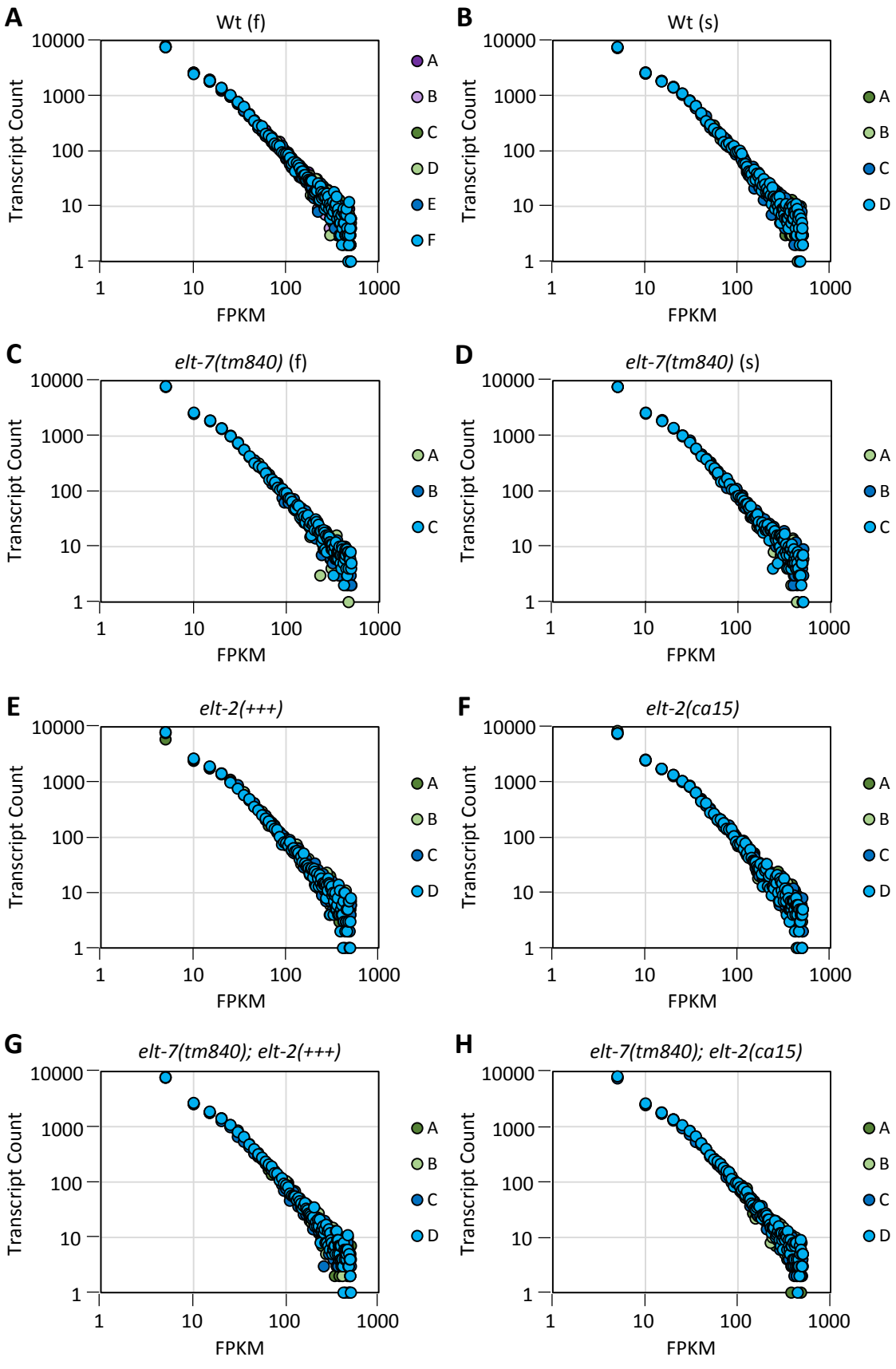

B
Mean Number of Transcripts (Thousands)
50
40
30
20
10
0
Wt (f) (n = 6)
Wt (s) (n = 4)
elt-7 (tm840) (f) (n = 3)
elt-7 (tm840) (s) (n = 3)
elt-2 (+++) (n = 4)
elt-2 (ca15) (n = 4)
elt-7 (tm840); elt-2 (+++) (n = 4)
elt-7 (tm840); elt-2 (ca15) (n = 4)
Strain

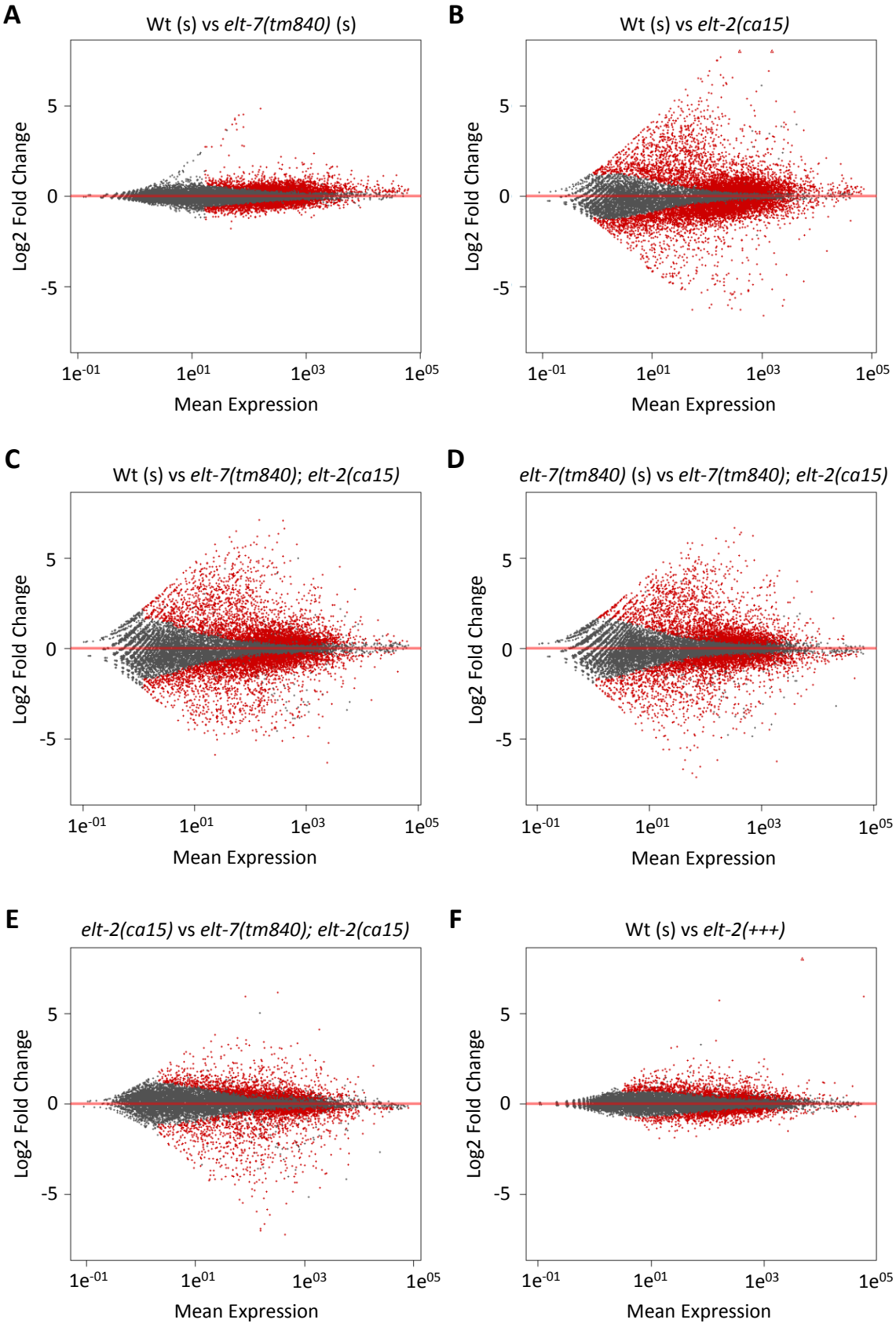
Figure 4.5. Power law plots of the log number of transcripts vs the log FPKM value for that transcript per replicate (A, B, C etc.) for each genetic background. There is no difference within or between populations for the number of transcripts expressed at a given FPKM value.


A
Wt (f)
B
Wt (s)
C
elt-7(tm840) (f)
D
elt-7(tm840) (s)
E
elt-2(+++)
F
elt-2(ca15)
G
elt-7(tm840); elt-2(+++)
H
elt-7(tm840); elt-2(ca15)
Transcript Count
FPKM
10000
1000
100
10
1
A
B
C
D
E
F

A list of genes known to be expressed in the intestine was assembled for the purpose of identifying genes potentially under direct regulation by ELT-2 and ELT-7. This included previous Serial Analysis of Gene Expression (SAGE) of enriched intestinal transcripts (McGhee et al., 2007; McGhee et al., 2009), genes already analyzed for regulation by both ELT-2 and ELT-7 (Sommermann et al., 2010) and any gene from WormBase (Yook et al., 2012) annotated with intestinal expression. A total of 2208 unique intestinal genes were used in the subsequent analyses of differential gene expression in the populations. The R Bioconductor package DESeq2 (Love et al., 2014) was used for differential expression analysis in a number of comparisons between the populations, with the default settings of false discovery rate (FDR) at 10% and adjusted p value threshold at 0.1. All results were filtered for genes that had a base mean expression in DESeq2 of greater than 100 to exclude any low expressed transcripts.

Biological replicates for the sorted wild-type population were compared to those of the single mutants, double mutants and *elt-2(+++)* overexpressed populations to identify all differentially expressed genes from each comparison (Figure 4.6). As an additional analysis, each single mutant was compared with the double mutants. Of all the populations, the *elt-2(+++)* overexpressed worms were the most similar to wild-type with only 1585 genes differentially expressed between the two groups (Table 4.4). Even at first glance, there appear to be more genes differentially expressed in the *elt-2(ca15)* vs wild-type comparison than in the *elt-7(tm840)* vs wild-type comparison (Figure 4.6, B vs A). There were 5236 genes differentially expressed in the *elt-2(ca15)* mutants and 2923 in the *elt-7(tm840)* mutants relative to the sorted wild-type replicates. The log2 fold changes of the differentially expressed genes in the *elt-2(ca15)* mutants were much larger than those observed in the *elt-7(tm840)* mutants, indicating that ELT-2 has a stronger effect on transcription levels than ELT-7. Comparing the *elt-7(tm840); elt-2(ca15)* double mutant worms to the sorted wild-type worms resulted in 4295 differentially expressed genes, 941 fewer genes than *elt-2(ca15)* alone. This finding could be due to genes expressed at a low level in the wild-type subsequently not transcribed in the double mutants. Analyzing gene expression in the double mutants vs the *elt-7(tm840)* single mutants identified 3644 differentially expressed genes, whereas the double mutants vs e*lt-2(ca15)* single mutants had fewer, only 2472. The basic conclusion of this analysis is

Figure 4.6. MA plots showing genes called as differentially expressed (red dots) in the indicated comparison as calculated by DESeq2. Up-regulated genes are above the horizontal line and down-regulated genes are below the line. All values are means of the biological replicates for each genetic background. Corresponding data can be found in Table 4.4.


A
Wt (s) vs elt-7(tm840) (s)
B
Wt (s) vs elt-2(ca15)
C
Wt (s) vs elt-7(tm840); elt-2(ca15)
D
elt-7(tm840) (s) vs elt-7(tm840); elt-2(ca15)
E
elt-2(ca15) vs elt-7(tm840); elt-2(ca15)
F
Wt (s) vs elt-2(+++)
Log2 Fold Change
Mean Expression
5
0
-5
1e-01
1e01
1e03
1e05

that the gene expression profile of the worms harbouring the *ca15* deletion of *elt-2* was more similar to that of the double mutants. Conversely, the gene expression profile of the *elt-7(tm840)* mutants appears more similar to wild-type worms than it is to the double mutants. Taken together, these preliminary results suggest that ELT-2 is involved in the regulation of many more genes than ELT-7.

In general, genes that have large changes in expression between the different genetic backgrounds are more likely to be direct targets of these two GATA transcription factors than genes with smaller changes in expression (assuming equal statistical significance). Many of the differentially expressed genes identified here are actually up-regulated or down-regulated by a relatively minor amount. For this reason, the differentially expressed genes were filtered by their log2 fold change values to identify those with larger changes in gene expression. Cut-offs of greater than log2 fold change of 0.5 or less than -0.5 were applied to the data, and genes that were not on the list of intestinally expressed genes were excluded.

After further filtering of the data, compared to wild-type there were 38 known intestinal genes down-regulated in the *elt-7(tm840)* mutants and 174 down-regulated in the *elt-2(ca15)* mutants (Table 4.4). On average, intestinal genes were more down-regulated in the *elt-2(ca15)* populations (mean log2 fold change of -1.31) than in the *elt-7(tm840)* populations (mean log2 fold change of -0.61). Unexpectedly, it was found that there were far more intestinal genes that were up-regulated than down-regulated for both single mutant strains vs wild-type (Table 4.4). There were about the same number of intestinal genes up-regulated (238) and down-regulated (223) in the double mutants compared to wild-type. The mean log2 fold change in down-regulated intestinal genes of the double mutants was -1.41, similar to the average value of the *elt-2(ca15)* single mutants.

The double mutants had 200 up-regulated and 209 down-regulated intestinal genes compared to the *elt-7(tm840)* worms, a result similar to the general analysis outlined above where all genes were considered. About three times as many intestinal genes (171) were found to be down-regulated than up-regulated (60) in the double mutant compared to the *elt-2(ca15)* single mutant. Worms carrying the *elt-2(+++)* array had 90 intestinal genes that were up-regulated relative to the sorted wild-type worms and an

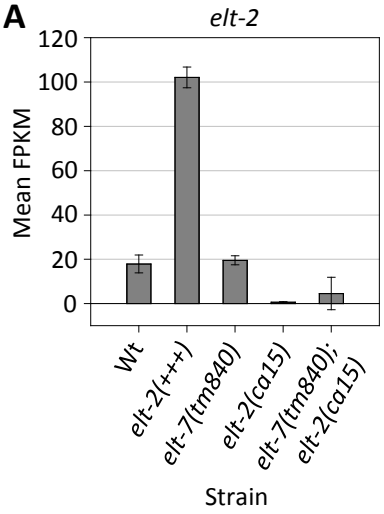
additional 24 that were down-regulated. The fact that intestinal genes were actually decreased in expression level in worms with high levels of ELT-2 imply that this protein may also function as a direct or indirect repressor of transcription, a feature of ELT-2 transcriptional regulation that has been suggested (McGhee et al., 2009) but not shown experimentally.

The mRNA expression levels of the two GATA transcription factors in each population must be taken into consideration in order to fully understand how they are involved in intestinal gene regulation. Expectedly, the extra copies of wild-type *elt-2* locus resulted in dramatic up-regulation of *elt-2* mRNA expression relative to wild-type levels (log2 fold change of 2.28, p adj = $2.06E^{-10}$) (Figure 4.7A). The *elt-2* mRNA expression levels in wild-type and the *elt-7(tm840)* mutant were essentially identical and not significantly different (p adj = 0.57). *elt-2* expression in the *ca15* deletion mutant was practically abolished, but not completely absent due to the aforementioned presence of ~1000 *elt-2(+++)* worms per biological replicate. The log2 fold change in these *elt-2* null mutants compared to the wild-type replicates was quite large at -4.69 (p adj = $3.58E^{-62}$). In contrast, the double mutants displayed slightly more *elt-2* expression than the single mutant, with much more variation between replicates. This was due to one outlier replicate skewing the mean, resulting in an automatic exclusion of the gene by DESeq2 and no adjusted p value (Figure 4.7A). These results can be visualized in Figure 4.7B where the per base sequence count (single replicate from each population) at the *elt-2* locus was plotted using the UCSC genome browser (Kent et al., 2002). These graphs reveal three key points about the RNAseq results. First, they show that the RNAseq reads were clean since the sequence reads map extensively to the exons of the *elt-2* gene and not to the non-coding introns or to flanking genomic regions. Second, it is evident that the *elt-2(+++)* replicate does indeed have many more reads mapping to the *elt-2* gene than either the wild-type or the *elt-7(tm840)* mutant. Third, very little sequence is obtained from the single or double mutants carrying the *ca15* deletion that spans the entire *elt-2* locus, which further reinforces that these populations have very few contaminating worms from the sorting protocol.

This RNAseq data confirm that *elt-2* transcription in the *elt-2(+++)* populations is greatly upregulated and in the *elt-2(ca15)* populations it is basically absent. It does not

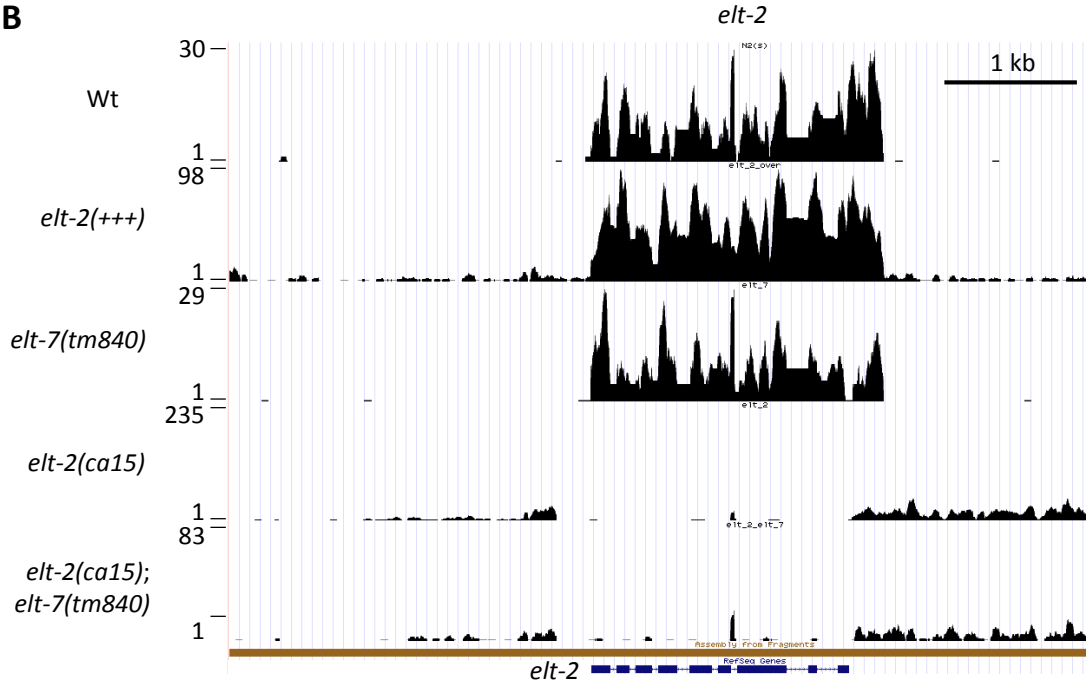
Figure 4.7. A. RNAseq results for *elt-2*. Left: Graph of the mean FPKM values for each indicated genetic background. Right: Table with the log2 fold change difference between genetic backgrounds and associated adjusted p values. NA indicates that DESeq2 did not assign an adjusted p-value because of an outlier replicate in the comparison. B. UCSC genome browser showing raw per base read counts for the *elt-2* gene (coding region diagrammed in blue below) and surrounding region of the genome in different genetic backgrounds. Note the varying scales and genetic background labels of each track on the left. C. Western blot of ELT-2 protein levels in indicated genetic backgrounds with UNC-15 (paramyosin) as loading control. D. Quantitation of ELT-2 protein levels in the western blot relative to wild-type levels.

A

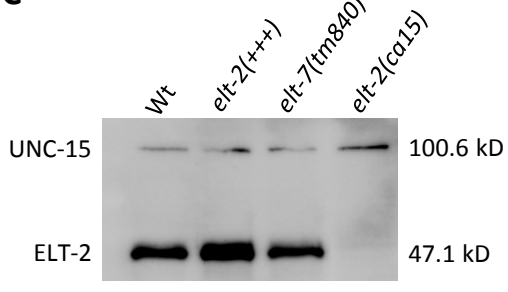


Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-2(+++)*	2.28	2.06E-10
Wt	*elt-7(tm840)*	0.13	0.57
Wt	*elt-2(ca15)*	-4.69	3.58E-62
Wt	*elt-7(tm840); elt-2(ca15)*	-1.46	NA
elt-2(+++)	*elt-2(ca15)*	-7.11	2.51E-232
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	-1.41	NA
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	1.35	NA

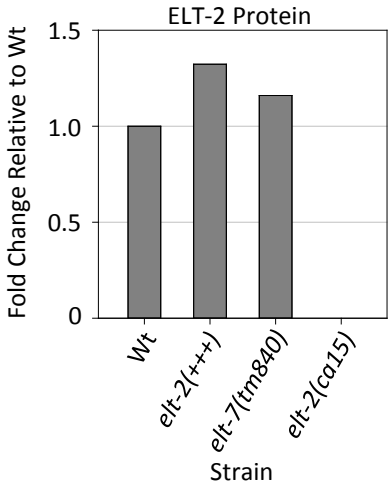
B



C



D



provide any information about the ELT-2 protein levels in these worms. A western blot probed with the ELT-2 specific 455-2A4 monoclonal antibody was performed to see if ELT-2 protein levels match the observed mRNA levels. ELT-2 protein levels were found to be mostly constant between wild-type and *elt-7(tm840)* mutants, while expression of ELT-2 in the *elt-2(+++)* worms appeared higher (Figure 4.7C). Quantitating the expression level relative to wild-type revealed a 1.3-fold increase in ELT-2 in the overexpressing strain (Figure 4.7D), compared to the five-fold increase observed in mRNA levels. Expression was not detected in the *elt-2(ca15)* null mutants. These results indicate that the presence of the extrachromosomal array carrying multiple copies of the endogenous *elt-2* locus results in greatly increased levels of *elt-2* transcripts but only a slight increase in the level of ELT-2 protein.

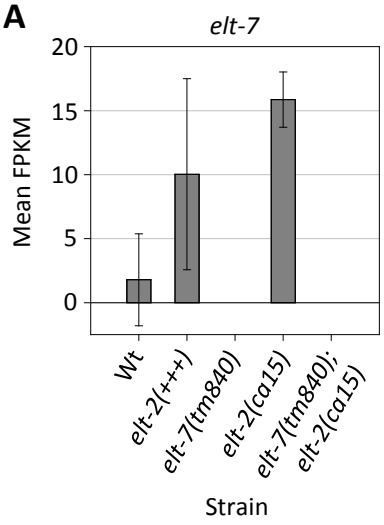
The mRNA expression levels of *elt-7* in wild-type were very low and variable, so much so that DESeq2 excluded *elt-7* from every comparison (Figure 4.8A). It appears that expression of *elt-7* was close to the threshold of detectability at this sequencing depth as *elt-7* is known to be expressed from the 2E cell stage of embryogenesis throughout the life of the worm (Sommermann et al., 2010) and has been previously detected in starved L1 larvae (McGhee et al., 2009). Furthermore, reads were mapped to the *elt-7* locus (Figure 4.8B), suggesting that *elt-7* mRNA may have been present in the libraries but at a very low abundance. Expression of *elt-7* in the *elt-2(+++)* populations was higher than in wild-type (but variable too), which fits with the fact that heat shock driven ELT-2 activates *elt-7* reporter expression (Sommermann et al., 2010). Expression of this transcription factor was surprisingly highest and most consistent in the *elt-2(ca15)* null mutant populations, suggesting that ELT-2 or downstream targets of ELT-2 repress *elt-7* at the L1 larval stage. This apparent paradox could be because of both direct and indirect effects of ELT-2 action on the *elt-7* promoter.

4.4.3 ELT-2 is Necessary to Activate Expression of Some Intestinal Genes

As outlined in the introduction, ELT-2 has been shown to directly activate expression of numerous intestinal genes and has been implicated in the regulation of many more. Therefore, there should be genes that positively respond to the ectopic levels of ELT-2 found in the *elt-2(+++)* populations.

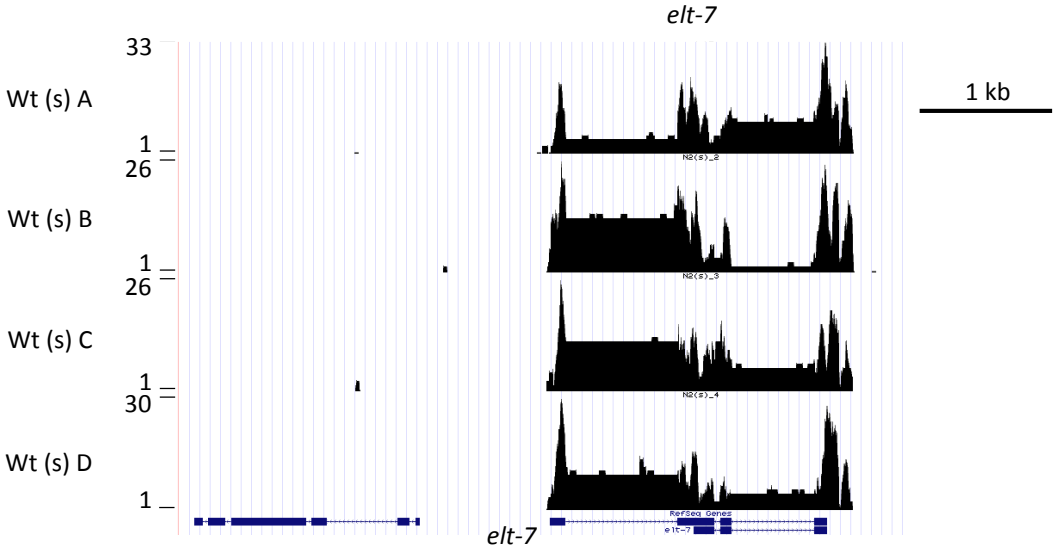
Figure 4.8. RNAseq results for *elt-7*. Left: Graph of the mean FPKM values for each indicated genetic background. Right: Table with the log2 fold change difference between genetic backgrounds and associated adjusted p values. NA indicates that DESeq2 called *elt-7* as not expressed. B. UCSC genome browser showing raw per base read counts for the *elt-7* gene (coding region diagrammed in blue below) and surrounding region of the genome in the wild-type replicates. Note the varying scales and labels of each track on the left.

A



Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-2(+++)*	NA	NA
Wt	*elt-7(tm840)*	NA	NA
Wt	*elt-2(ca15)*	NA	NA
Wt	*elt-7(tm840); elt-2(ca15)*	NA	NA
elt-2(+++)	*elt-2(ca15)*	NA	NA
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	NA	NA
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	NA	NA

B

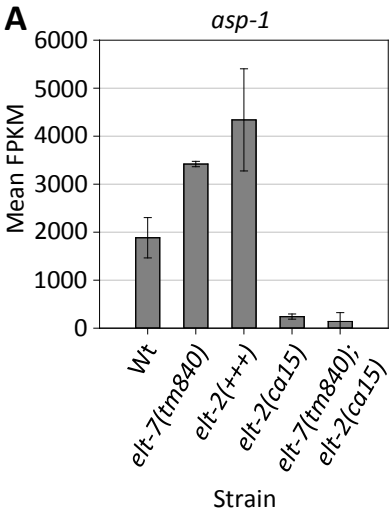


There were 90 identified intestinal genes up-regulated above a log2 fold change of 0.5 in the *elt-2(+++)* replicates compared to the wild-type replicates. In order to increase the chances of identifying direct targets of ELT-2 and ELT-7, analysis of targets was limited to those expressed predominantly or exclusively in the intestine. Forty-seven of these genes were identified as being expressed either exclusively or primarily in the intestine (Table 4.5). A gene expression pattern was annotated as primarily in the intestine based on a subjective assessment of what other tissue(s), cell number, and temporal aspects it had. For example, *lec-6* has been reported to be expressed in the intestine as well as the grinder cells located in the back of the pharynx (Maduzia et al., 2011) and so was annotated with the designation of primarily intestine. Any gene that was also expressed in the hypodermis was immediately excluded due to the number of cells in that tissue and the fact that hypodermal GATA transcription factors likely also regulate expression of the gene (Gilleard and McGhee, 2001).

Three examples of these 47 ELT-2 positively responding genes can be found in Figure 4.9. The *asp-1* gene encodes an aspartic protease that is exclusively expressed in the *C. elegans* intestine and is homologous to the human enzyme Cathepsin D Aspartic Protease (Tcherepanova et al., 2000). Expression of *asp-1* is reported to be strongest toward the end of embryogenesis and in the early larval stages, after which expression disappears by the adult stage (Tcherepanova et al., 2000). The *asp-1* gene was one of the most highly expressed intestinal genes found in these RNAseq datasets and expression was further up-regulated in the *elt-2(+++)* overexpressed worms (Figure 4.9A). Consistent with this result was the finding that loss of ELT-2 resulted in a drastic reduction in *asp-1* mRNA. Thus, transcription of the *asp-1* gene positively responds to ELT-2 levels in the intestine and appears to greatly depend upon ELT-2 for activation. This is clearly observable in the *elt-2(ca15)* null mutants, which display a mean FPKM ~13% of the mean wild-type FPKM value. Analysis of the proximal promoter (2 kb upstream of the first ATG) in *asp-1* revealed three TGATAA sites as well as six other GATA sites. Knockout of either of the two 3' most TGATAA sites results in greatly diminished expression of an *asp-1* transcriptional reporter (B. Lancaster, personal communication), providing further evidence that ELT-2 is directly activating expression of this gene.

Figure 4.9. RNAseq results for representative genes that positively respond to excess ELT-2. A. *asp-1*. B. *clec-85*. C. *mtl-2*. Left: Graph of the mean FPKM values in each genetic background. Right: Table with the log2 fold change difference between genetic backgrounds and associated adjusted p values. Corresponding data can be found in Table 4.5.

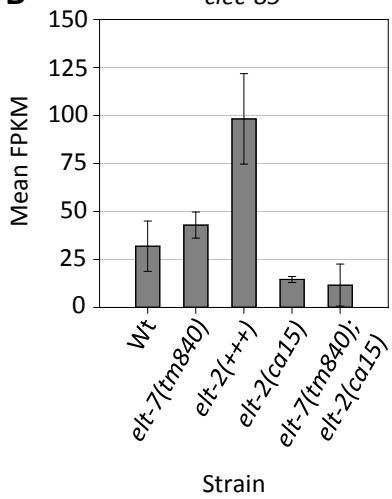
A



ELT-2 Positive Responder

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	0.84	1.20E-07
Wt	*elt-2(+++)*	0.99	9.24E-06
Wt	*elt-2(ca15)*	-3.03	3.58E-42
Wt	*elt-7(tm840); elt-2(ca15)*	-2.85	3.66E-04
elt-2(+++)	*elt-2(ca15)*	-4.15	4.31E-101
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	-0.41	NA

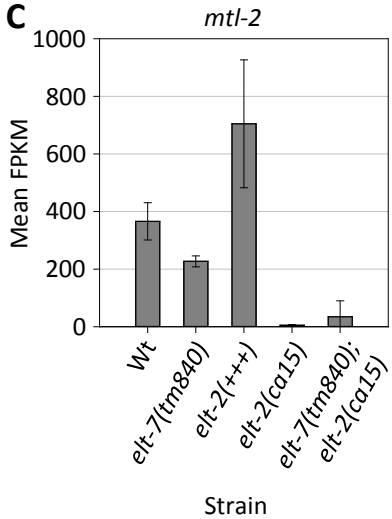
B



ELT-2 Positive Responder

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	0.31	0.39
Wt	*elt-2(+++)*	1.24	8.56E-07
Wt	*elt-2(ca15)*	-1.26	1.23E-07
Wt	*elt-7(tm840); elt-2(ca15)*	-1.42	1.78E-02
elt-2(+++)	*elt-2(ca15)*	-2.72	1.29E-48
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	-0.25	NA

C



ELT-2 Positive Responder

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	-0.82	3.68E-06
Wt	*elt-2(+++)*	0.52	6.05E-02
Wt	*elt-2(ca15)*	-6.16	2.01E-53
Wt	*elt-7(tm840); elt-2(ca15)*	-2.56	3.01E-03
elt-2(+++)	*elt-2(ca15)*	-6.41	7.82E-56
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	1.57	NA

The protein product of the *clec-85* gene is a C-type lectin predicted to have carbohydrate binding activity and is expressed exclusively in the intestine (Alper et al., 2007). The expression of this gene in the various populations of worms is typical of a gene that is positively regulated by ELT-2 (Figure 4.9B). The mean FPKM value in the *elt-2(ca15)* null mutants was only down to about 46% that of the mean wild-type value, suggesting that while this gene strongly responded to ELT-2 overexpression, there must be other factors that can activate transcription of this gene.

The *mtl-2* gene encodes one partner of a pair of metallothioneins (metal binding proteins) in the *C. elegans* genome and it is expressed exclusively in the intestine in response to the presence of metals (Freedman et al., 1993). Analysis of the *mtl-2* promoter by (Moilanen et al., 1999) found two TGATAA sites that ELT-2 is capable of binding and that are necessary for expression. Furthermore, *mtl-2* reporter expression was observed to be induced outside of the intestine by ectopic ELT-2 expression and *elt-2(ca15)* null mutants did not express the reporter at all (Moilanen et al., 1999). These findings are reproduced in this study where expression of *mtl-2* was up-regulated upon overexpression of ELT-2 and practically abolished in *elt-2(ca15)* mutants (Figure 4.9C). The mean FPKM value for *mtl-2* RNA expression in an *elt-2(ca15)* null mutant background was a meager 2% of the wild-type mean FPKM value. Given that some worms in these replicates were known to be overexpressing ELT-2, it is safe to conclude that little to no transcription occurred in the *elt-2* mutants. The results presented here provide more evidence for the critical function of ELT-2 in activating expression of a set of terminal intestinal genes.

It should be noted, that despite seemingly exclusive regulation by ELT-2, these genes did respond modestly to ELT-7 levels (Figure 4.9). For example, *asp-1* mRNA expression was significantly up-regulated in the *elt-7(tm840)* mutants signifying possible constitutive partial repression by ELT-7. Oppositely, *mtl-2* mRNA levels were significantly down-regulated upon ELT-7 loss, evidence that ELT-7 may also activate expression of this gene. Expression of *clec-85* was not significantly different from wild-type in the *elt-7(tm840)* mutants.

Nevertheless, *asp-1* and *mtl-2* expression in the *elt-2(ca15)* mutants was ~13% and ~2% respectively of mean wild-type expression. This suggests that ELT-7 is

incapable of activating transcription of these genes in the absence of ELT-2. In contrast, of the 38 intestinal genes down-regulated in the *elt-7(tm840)* mutant samples, none were reduced to levels lower than about half of their wild-type levels (Figure 4.10). This result implies that ELT-7 is involved in the activation of transcription of these genes, but it is not the only *trans*-acting factor that can do so. In other words, there is no intestinal gene that appears to be activated predominantly by ELT-7, unlike the above examples that show this is the case for a set of ELT-2 targets.

4.4.4 ELT-2 is a Major Activator of Intestinal Genes While ELT-7 is a Minor Activator

The initial analysis of differentially expressed genes in section 4.4.2 introduced the idea that ELT-2 is the primary GATA factor regulating intestinal gene expression and ELT-7 is partially redundant. Comparison of the genes that were down-regulated relative to wild-type in both single mutants and the double mutants was employed to further elucidate the different functions of these GATA factors.

About half (16 out of 38) of the intestinal genes down-regulated in the *elt-7(tm840)* (vs wild-type) samples were also down-regulated in the *elt-2(ca15)* samples (vs wild-type) (Figure 4.11A). Exactly two thirds of the down-regulated intestinal genes in the *elt-2(ca15)* vs wild-type comparison (116 out of 174) were also down-regulated in the double mutants vs wild-type. Most of these genes (102) were not down-regulated in the *elt-7* single mutants, providing more evidence in favour of ELT-2 acting as the primary GATA regulator in the intestine. There were 14 intestinal genes found to be down-regulated in both of the single mutants and the double mutant (each compared to wild-type), which are good candidates for regulation by ELT-2 and ELT-7 jointly in some manner (Figure 4.11A). Ten genes overlapped between the *elt-7(tm840)* mutants and the double mutants, but were not found in *elt-2(ca15),* suggesting that these few genes are regulated only by ELT-7.

A slightly different Euler diagram displays the same overlap between the *elt-2(ca15)* single mutant and the double mutant strains vs wild-type, and also includes the 171 genes that were down-regulated in the double mutant relative to the *elt-2(ca15)* single mutant (Figure 4.11B). The overlaps between the ellipses in this diagram were

Figure 4.10. Scatter plot depicting the *elt-7(tm840)* mutant population mean FPKM expression value of ELT-7 activated genes as a percentage of the wild-type mean FPKM value. All genes fall between 50-80% of mean wild-type FPKM value.

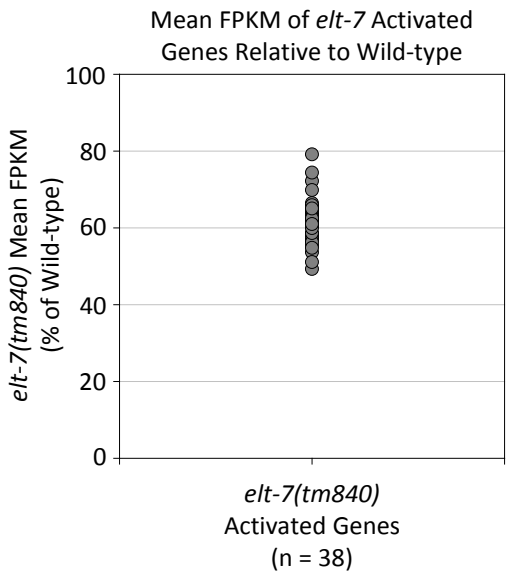
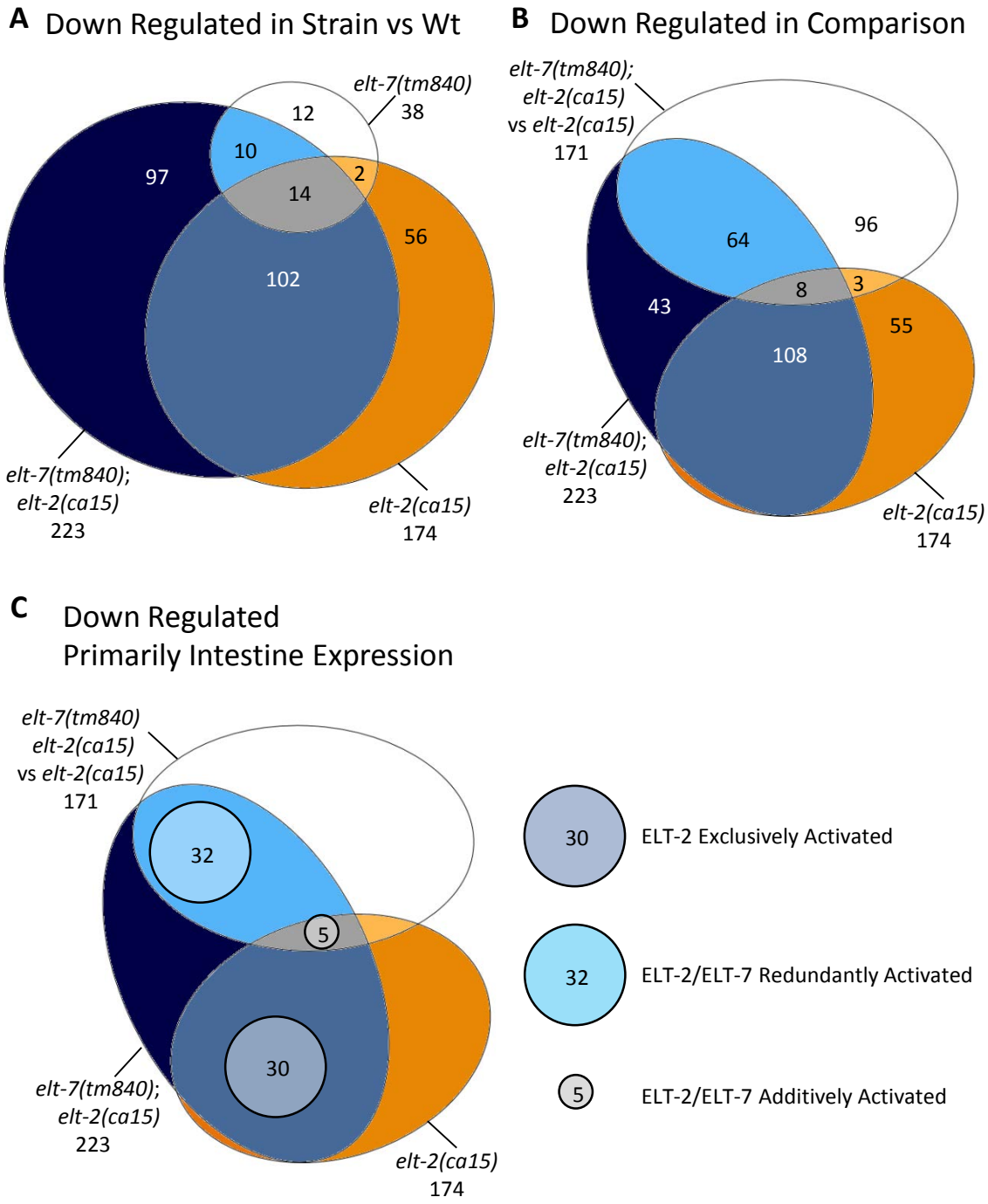

Mean FPKM of *elt-7* Activated Genes Relative to Wild-type
100
80
60
40
20
0
elt-7(tm840) Mean FPKM (% of Wild-type)
elt-7(tm840) Activated Genes (n = 38)

Figure 4.11. Euler diagrams that show the number and overlap of down regulated intestinal genes in A. *elt-7(tm840)* single mutants, *elt-2(ca15)* single mutants, and *elt-7(tm840); elt-2(ca15)* double mutants compared to wild-type. B. *elt-2(ca15)* single mutants and *elt-7(tm840); elt-2(ca15)* double mutants compared to wild-type; *elt-7(tm840); elt-2(ca15)* double mutants compared to *elt-2(ca15)* single mutants. C. Same Euler diagram as in B with inner circles highlighting the number of genes in each region that are expressed exclusively or primarily in the intestine and prediction of how expression of these genes are regulated in the intestine.


A Down Regulated in Strain vs Wt
elt-7(tm840)
38
12
10
2
97
14
56
102
elt-7(tm840);
elt-2(ca15)
223
elt-2(ca15)
174
B Down Regulated in Comparison
elt-7(tm840);
elt-2(ca15)
vs elt-2(ca15)
171
96
64
8
3
43
55
108
elt-7(tm840);
elt-2(ca15)
223
elt-2(ca15)
174
C
Down Regulated
Primarily Intestine Expression
elt-7(tm840)
elt-2(ca15)
vs elt-2(ca15)
171
32
5
30
elt-7(tm840);
elt-2(ca15)
223
elt-2(ca15)
174
30
ELT-2 Exclusively Activated
32
ELT-2/ELT-7 Redundantly Activated
5
ELT-2/ELT-7 Additively Activated

used to classify these intestinal genes based on how they are regulated by ELT-2 and ELT-7 (Figure 4.11C). Genes expressed primarily or exclusively in the intestine can be predicted to fall into five different classes based on how they are activated by ELT-2 and/or ELT-7, summarized as follows.

i) There were 30 intestinal genes exclusively activated by ELT-2, found to be down-regulated in all populations that carry the *elt-2* null mutation, but not down in the *elt-7* mutants (Table 4.6). As seen in Figure 4.12A, the *asp-5* gene encoding an aspartic protease is an example of such a gene. Expression of *asp-5* in the *elt-7(tm840)* mutant background was not significantly different from wild-type. In contrast, there was a significant down-regulation in expression in both the *elt-2(ca15)* mutants and the *elt-7(tm840); elt-2(ca15)* double mutants compared to wild-type.

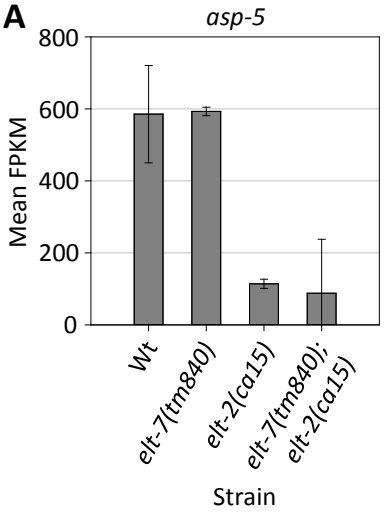
ii) Eight of the ten genes that appeared to be exclusively activated by ELT-7, were confirmed to be unaffected by levels of ELT-2 (not further down-regulated in the double mutant vs *elt-7(tm840)*). Only two of these genes were expressed primarily or exclusively in the intestine. *F56C9.7* encodes a nematode conserved protein involved in intestinal dipeptide transport as well as fat storage (Benner et al., 2011) and *Y4C6B.5* is uncharacterized. Both of these genes were expressed ~60% of wild-type FPKM levels in the *elt-7(tm840)* null mutants indicating other factors contribute to their transcriptional activation.

iii) The 32 primarily expressed intestinal targets that are activated redundantly in an either/or manner by ELT-2 and ELT-7 are only down-regulated in the double mutants (Table 4.7). An example of a redundantly activated gene would be the C-type lectin *clec-65* seen in Figure 4.12B. There was no observable effect on transcription in either single mutant compared to wild-type, but the double mutant had significantly reduced *clec-85* mRNA levels relative to wild-type. Intestinal genes of this class are therefore observed to be sufficiently activated by either ELT-2 or ELT-7 in the absence of the other, but insufficiently expressed when both factors are missing.

iv) Genes that require additive activation by both GATA factors for normal transcription levels should be down-regulated relative to wild-type in both single mutants, and further down-regulated in the double mutants (Table 4.8). There are eight genes identified that fit this profile, five of which are expressed primarily or exclusively in the

Figure 4.12. RNAseq results for representative genes from the ELT-2 exclusively activated class A. *asp-5*; the ELT-2/ELT-7 redundantly activated class B. *clec-65*; and the ELT-2/ELT-7 additively activated class C. *pcp-3*. Left: Graph of the mean FPKM values in each genetic background. Right: Table with the log2 fold change difference between genetic backgrounds and associated adjusted p value. Corresponding data can be found in Tables 4.6, 4.7 and 4.8.

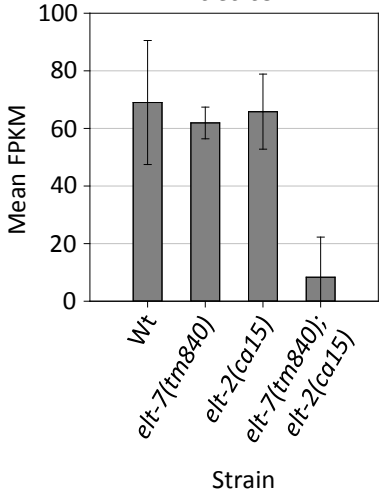
A



ELT-2 Exclusively Activated

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	0.09	0.70
Wt	*elt-2(ca15)*	-2.38	8.44E-52
Wt	*elt-7(tm840); elt-2(ca15)*	-1.93	3.40E-02
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	-1.78	NA
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	-0.18	NA

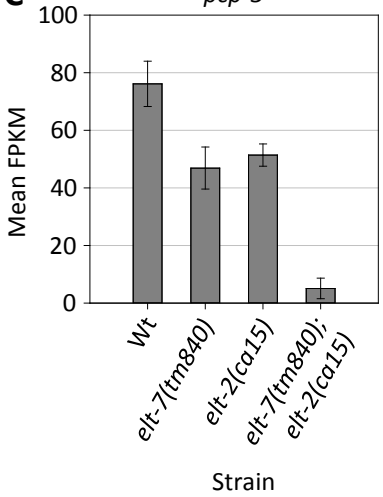
B



ELT-2/ELT-7 Redundantly Activated

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	-0.07	0.79
Wt	*elt-2(ca15)*	-0.21	0.38
Wt	*elt-7(tm840); elt-2(ca15)*	-2.21	1.20E-02
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	-1.94	NA
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	-1.36	4.40E-02

C



ELT-2/ELT-7 Additively Activated

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	-0.62	3.09E-06
Wt	*elt-2(ca15)*	-0.52	3.42E-06
Wt	*elt-7(tm840); elt-2(ca15)*	-3.65	1.35E-17
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	-2.94	4.92E-09
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	-2.87	3.09E-13

intestine (Figure 4.11C). However, only a single gene, the serine-type peptidase *pcp-3*, actually fulfilled all criteria of this class (Figure 4.12C). Expression of *pcp-3* mRNA was significantly lower in both the *elt-7(tm840)* mutants and the *elt-2(ca15)* mutants compared to wild-type. The double mutants also had significantly reduced expression compared to wild-type, but more importantly, they were down-regulated compared to both single mutants. It appears from this RNAseq data that *pcp-3* requires roughly equal contributions from ELT-2 and ELT-7 for transcriptional activation in the intestine.

In contrast, the other four genes predicted to be in this class were uncharacteristic (Table 4.8). As predicted, all four are significantly down-regulated in both the *elt-2(ca15)* mutant populations and the double mutant populations when compared to wild-type. They are further down-regulated in the double mutants relative to the *elt-2(ca15)* single mutants, but the *elt-7(tm840)* mutants were not significantly different from wild-type. Based on these results it appears that ELT-7 is activating transcription from these genes, albeit this effect is masked by wild-type ELT-2 levels suggesting that ELT-2 has a stronger effect and the primary activator of transcription of these four genes.

A second gene, *ndg-4*, does appear to be clearly activated in an additive manner by ELT-2 and ELT-7. It was not initially identified for two reasons. First, it fell just below the base mean expression cut-off of 100 in the *elt-2(ca15)* vs double mutant DESeq2 comparison and second, the gene is expressed in parts of the hypodermis in addition to the intestine. Similar to *pcp-3*, *ndg-4* was down-regulated in both single mutants and further down-regulated in the double mutant relative to the single mutants (Table 4.8).

v) There are 14 intestinal genes down-regulated in both of the single mutants and the double mutants, two of which (*pcp-3* and *ndg-4*) are additively regulated by ELT-2 and ELT-7. Six of the remaining twelve genes are expressed primarily or exclusively in the intestine, but none are conclusively regulated synergistically by the two GATA factors (Table 4.9). Five of the six were down-regulated more in the *elt-2(ca15)* mutants than the *elt-7(tm840)* mutants. Furthermore, four of the six had base mean expression below the minimum cut-off of 100 in various comparisons between the single and double mutant populations. It appears that these genes are primarily regulated by ELT-2 with

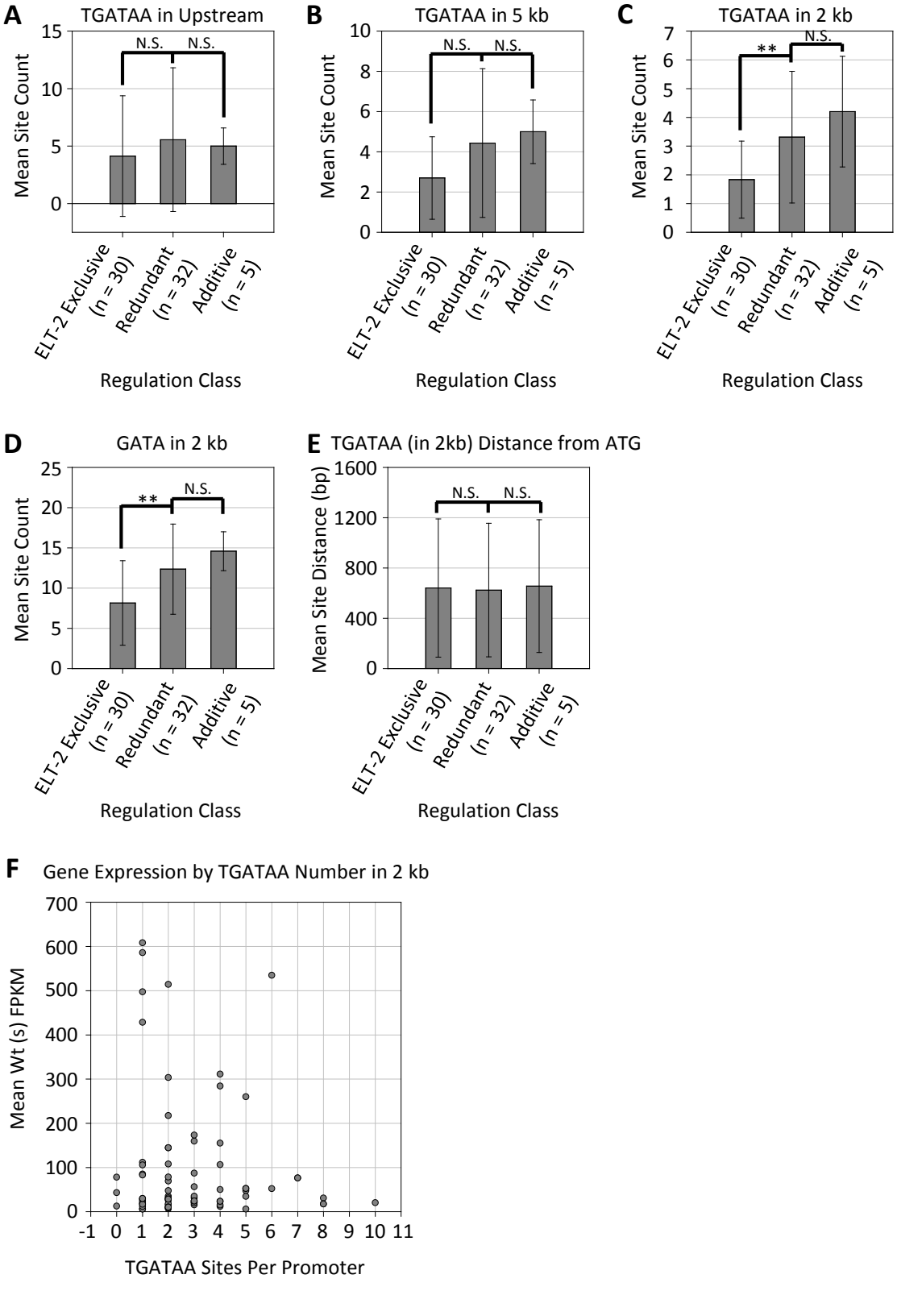
some role for ELT-7, but the low expression levels of these genes make it difficult to determine how these two factors are functioning.

4.4.5 Joint Targets Have More *Cis*-Regulatory Elements in their Promoters

The genes identified in these classes were selected based on the fact that they were primarily expressed in the intestine. Thus, there should be an enrichment for TGATAA sites in their promoters that function to mediate activation of transcription by ELT-2 and ELT-7 (McGhee et al., 2009). The entire upstream region stretching from the first ATG in each gene to the next upstream gene was retrieved using Regulatory Sequence Analysis Tools (RSAT) (Thomas-Chollier et al., 2011) to search for potential *cis*-regulatory elements. Analysis of the number of TGATAA sites present in this region for each class of genes revealed no significant differences in the mean number of sites between the three activation classes (Figure 4.13A). This is likely due to the large variation in distance to the next upstream gene. For example, the size of the upstream regions in the exclusive class of genes ranged from 209 bp to 30.9 kb and spanned 355 bp to 15.5 kb in the redundant class of genes. Limiting the analysis to 5 kb upstream of the first ATG resulted in a noticeable trend of increasing number of TGATAA sites with an increasing role of ELT-7. As seen in Figure 4.13B, there were slightly more TGATAA sites on average in the redundant and additive classes of genes than in the ELT-2 exclusive class. The difference between the exclusive and redundant classes was close to significant ($p = 0.062$), suggesting that there may be a clue in this region to how these two classes are regulated differently. Indeed, reducing the analysis even further to the proximal 2 kb upstream of the first ATG revealed that there are significantly more TGATAA sites found in genes that are regulated by both ELT-2 and ELT-7 than genes that are controlled solely by ELT-2 (Figure 4.13C).

The number of general GATA sites in the 2 kb region upstream of each class of genes was analyzed to ensure that the difference observed for TGATAA sites was not balanced out by the overall number of GATA sites. Similar to the result for TGATAA sites, there were significantly more GATA sites found in the 2 kb region upstream of the first ATG in the redundant and additive classes compared to the exclusive class (Figure 4.13D). Taken together, these results indicate that closer to the first ATG of an intestinal

Figure 4.13. Analysis of the regions upstream in each class of activated intestinal genes. A. No difference in the number of TGATAA sites between classes is observed when all sequence upstream to the next gene is analyzed. N.S. indicates no significant difference. B. In the proximal 5 kb upstream of the first ATG there is a trend of more TGATAA sites in genes activated by both ELT-2 and ELT-7 (redundantly or additively) compared to those activated exclusively by ELT-2. C. In the proximal 2 kb upstream of the first ATG the increased number of TGATAA sites is statistically significant. ** denotes significance of $p < 0.01$. D. The number of GATA sites in the proximal 2 kb is also significantly increased in genes activated by ELT-2 and ELT-7 compared to those activated by ELT-2 alone. E. There is no significant difference between the classes in the mean distance (bp) of TGATAA sites in the proximal 2 kb promoter relative to the first ATG. F. Plotting mean wild-type (s) FPKM values vs number of TGATAA sites in the proximal 2 kb promoter for genes in all classes reveals no correlation between expression level and TGATAA number.


A
TGATAA in Upstream
Mean Site Count
N.S.
N.S.
ELT-2 Exclusive (n = 30)
Redundant (n = 32)
Additive (n = 5)
Regulation Class
B
TGATAA in 5 kb
C
TGATAA in 2 kb
**
D
GATA in 2 kb
E
TGATAA (in 2kb) Distance from ATG
Mean Site Distance (bp)
F
Gene Expression by TGATAA Number in 2 kb
Mean Wt (s) FPKM
TGATAA Sites Per Promoter

gene, there are significantly more intestinal *cis*-regulatory elements found in genes that have ELT-2 and ELT-7 input compared to those that only respond to ELT-2. Therefore, the mean distance from the first ATG to the TGATAA site was scored to see if any organizational differences existed in the proximal 2 kb promoters of these genes. There was no difference in the location of TGATAA sites relative to the first ATG between any of the classes, again suggesting that ELT-2 and ELT-7 do not function synergistically (Figure 4.13E).

It is possible that the number of TGATAA and GATA sites has nothing to do with the way in which these genes are activated and simply reflects a difference in expression level, with more sites resulting in increased expression. Sixty-six of the 68 identified genes from the three classes were plotted by their mean sorted wild-type expression (FPKM value) against the number of TGATAA sites found in the proximal 2 kb region of their promoter to test this hypothesis. The *asp-1* gene was not plotted as it is expressed considerably higher than any of the other genes and *ndg-4* was excluded (as it was for the above analysis) due to its hypodermal expression. There was no observed trend of increasing expression with increasing number of TGATAA sites for these genes, indicating that the differences in site numbers between these classes may reflect the different inputs by ELT-2 and ELT-7 for transcriptional activation.

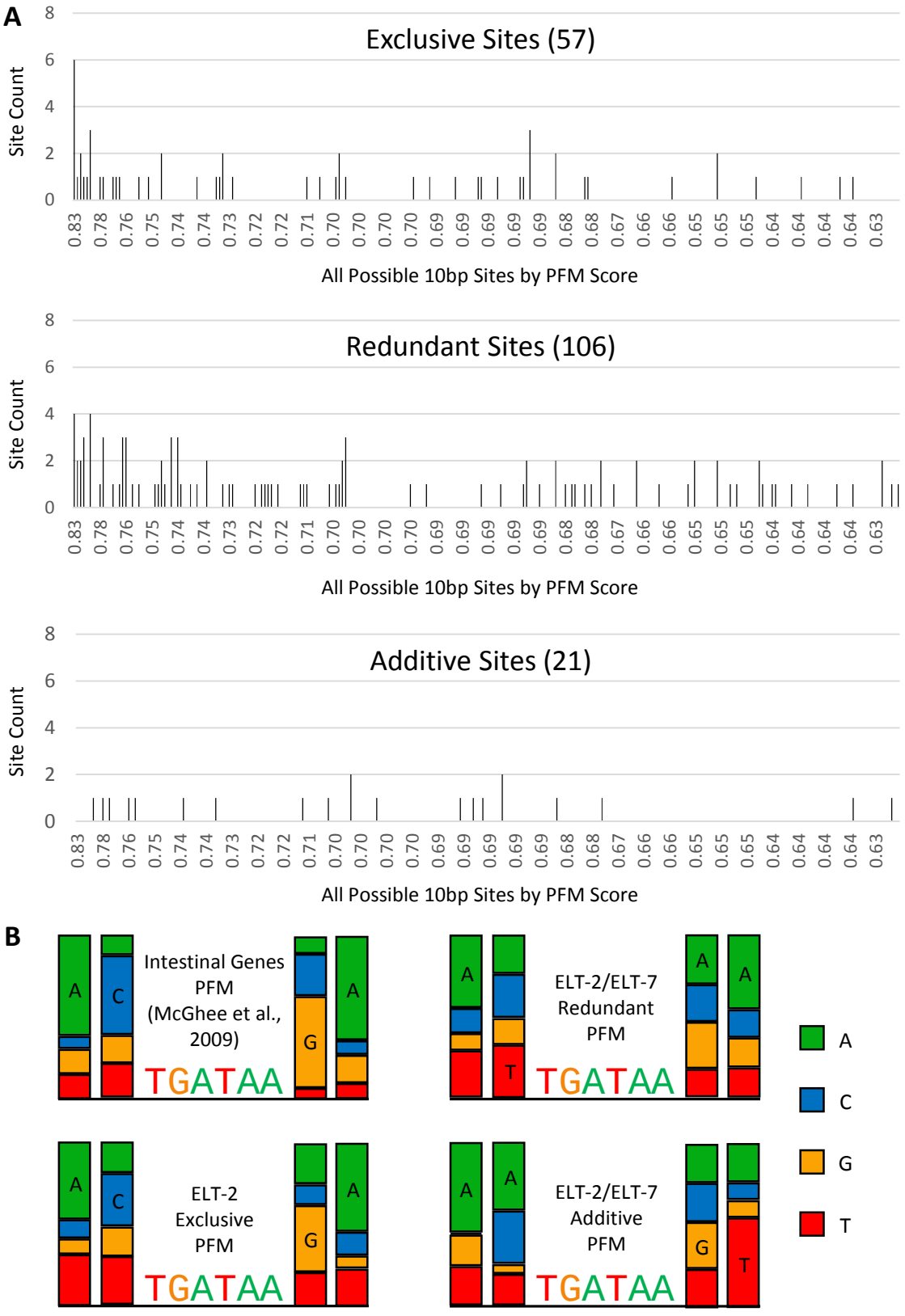
The TGATAA sites in these classes themselves may also provide information on how ELT-2 and ELT-7 mediate activation of transcriptional targets in the intestine. A position frequency matrix (PFM) of overrepresented TGATAA sequences found in promoters of intestinally expressed genes encompasses a ten base pair sequence with two base pairs flanking at both the 5' and 3' ends of the TGATAA core sequence (McGhee et al., 2009) (Table 4.10). The frequency of each base found at each position along the 10 bp sequence can be summed to provide a score relative to all 256 possible sequences. For example, (McGhee et al., 2009) found that the most frequent base pair sequence was 5'-AC-TGATAA-GA-3' (Table 4.10), which based on the frequency of each base at those positions would yield a score of 0.83. This is the highest possible scoring sequence, where scores can range from 0.63 to 0.83. The underlying hypothesis of this analysis is that higher scoring sites will have a higher affinity for ELT-2, and possibly ELT-7, but this has yet to be experimentally proven. The TGATAA sites for 67 of the 68 genes

(excluding *ndg-4* due to hypodermal expression) were counted and scored to determine if there were any sequences favoured by a particular class. There was no obvious bias for high vs low scoring sites in any of the classes – with the exception that the highest scoring site 5'-AC-TGATAA-GA-3' was found more frequently in the exclusive class (6/57, ~10% of all exclusive sites compared to 4/106, ~4% of all redundant sites) (Figure 4.14A). The additively activated class of genes did not have enough sites to extract any information, other than the fact that sites were spread across the entire spectrum from high scoring to low, as was the case for the other classes (Figure 4.14A).

PFMs were created using the sites within 2 kb of the first ATG from each class of activated intestinal genes and then compared to the previously identified PFM. As seen in Figure 4.14B and Table 4.10, the frequency of the 5' flanking (1st and 2nd positions) AC and 3' flanking (9th and 10th positions) GA was quite high in the original PFM. This was also the most frequent sequence found in the sites of the ELT-2 exclusively activated class of intestinal genes; however, there was less of a bias over other bases at these positions compared to the original PFM. Interestingly, in the ELT-2/ELT-7 redundantly activated intestinal genes, there was a preference for T (over C) at the 2nd position and A (over G) at the 9th position, in contrast to the exclusive and original PFMs. The ELT-2/ELT-7 additively activated genes also displayed a different PFM than the original, favouring A (over C) at the 2nd position and strongly favouring T (over A) at the 10th position. This finding comes with the caveat that there were only five genes and 21 sites analyzed in this additive class making it difficult to reach any conclusions about the data. Considering the PFM results for all three classes, they raise the possibility that ELT-2 and ELT-7 may have different binding preferences for the flanking nucleotides of TGATAA sites.

The results of this section and the previous section indicate that terminal intestinal genes can be distinguished into four distinct classes based on how the GATA transcription factors of this organ activate them. The most abundant class of genes are those that appear to be activated exclusively by ELT-2 (108). Then there are genes that are activated by either ELT-2 or ELT-7 and have significantly more TGATAA sites in their proximal 2 kb upstream region than those in the first class (64). There are a small

Figure 4.14. A. Bar graphs depicting all 256 possible 10 bp TGATAA sites and the number of times they appear in the promoters of genes from each class of intestinally activated genes. Site score refers to the score of each sequence based on the position frequency matrix of TGATAA sites from intestinal gene promoters (McGhee et al, 2009). B. Diagrams depicting the frequency of each base at positions flanking TGATAA sites show variation for nucleotide preference between the different gene classes at positions 1, 2, 9 and 10. Corresponding data for PFMs can be found in Table 4.9.


A
Exclusive Sites (57)
Site Count
All Possible 10bp Sites by PFM Score
Redundant Sites (106)
Site Count
All Possible 10bp Sites by PFM Score
Additive Sites (21)
Site Count
All Possible 10bp Sites by PFM Score
B
Intestinal Genes PFM (McGhee et al., 2009)
TGATAA
ELT-2/ELT-7 Redundant PFM
TGATAA
ELT-2 Exclusive PFM
TGATAA
ELT-2/ELT-7 Additive PFM
TGATAA
A
C
G
T

number of genes that seem to require both ELT-2 and ELT-7 for activation (8), most of which seem to depend more on ELT-2 than ELT-7 for activation. There is no evidence of intestinal genes regulated synergistically by these factors. Finally, two intestinal genes appear to be regulated by ELT-7 independently of ELT-2. The conclusion from these results is that ELT-2 is a major transcriptional activator of gene expression in the intestine while ELT-7 is a minor transcriptional activator and partially redundant with ELT-2.

4.4.6 ELT-2 and ELT-7 are Required for Repression of Some Intestinal Genes

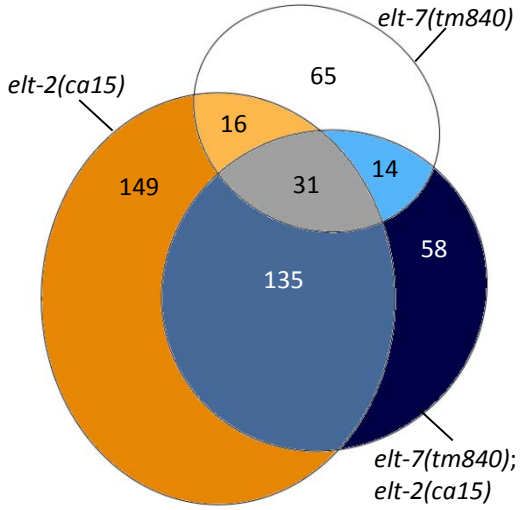
Thus far, the results presented here have focused on the transcriptional activation role of the GATA transcription factors ELT-2 and ELT-7. All terminal intestinal genes shown to be under direct ELT-2 control to date require ELT-2 as an activator of gene expression. However, as previously reported above, there were a number of intestinally expressed genes that were up-regulated in the *elt-7(tm840), elt-2(ca15)* and double mutant populations. The same analysis that was performed on the down-regulated genes was also used to identify regulation classes for the up-regulated genes expressed primarily or exclusively in the intestine.

There were a total of 126 intestinal genes up-regulated in the *elt-7(tm840)* mutants relative to wild-type, 331 genes up-regulated in the *elt-2(ca15)* mutants and 238 in the double mutants. When comparing the up-regulated genes in the single mutants and double mutants, it was found that many more of the up-regulated genes were shared between the *elt-2(ca15)* mutants and the double mutants than with the *elt-7(tm840)* mutants (Figure 4.15A). There were 31 genes found to be up-regulated in all three comparisons to wild-type, implying that they could be regulated by both ELT-2 and ELT-7. Comparing the *elt-2* single mutants and the double mutants resulted in 60 intestinal genes found to be up-regulated in the double compared to the single (Figure 4.15B). As before, genes from different repressed classes were selected for further analysis if they were expressed exclusively or primarily in the intestine.

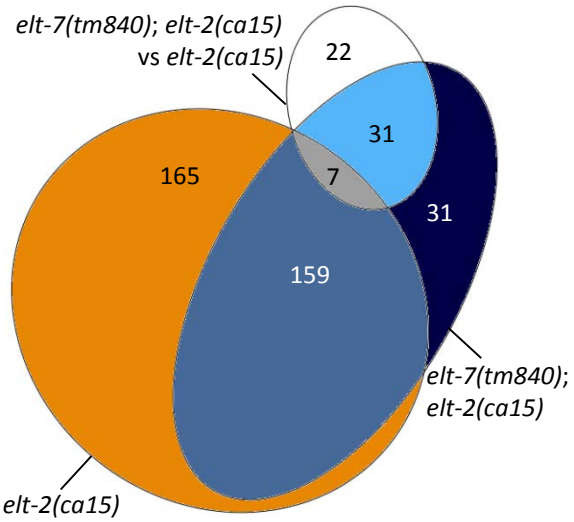
i) An example of one of the 28 ELT-2 exclusively repressed genes (Table 4.11) is *pqn-25*, a gene that encodes a prion-like domain containing protein (Figure 4.16A). Expression of this gene was not observed to be significantly different from wild-type in

Figure 4.15. Euler diagrams that show the number and overlap of up-regulated intestinal genes in A. *elt-7(tm840)* single mutants, *elt-2(ca15)* single mutants, and *elt-7(tm840); elt-2(ca15)* double mutants compared to wild-type. B. *elt-2(ca15)* single mutants and *elt-7(tm840); elt-2(ca15)* double mutants compared to wild-type; *elt-7(tm840); elt-2(ca15)* double mutants compared to *elt-2(ca15)* single mutants. C. Same Euler diagram as in B with inner circles highlighting the number of genes in each region that are expressed exclusively or primarily in the intestine and prediction of how expression of these genes are regulated in the intestine.

A Up Regulated in Strain vs Wt



B Up Regulated in Comparison



C Up Regulated
Primarily Intestine Expression

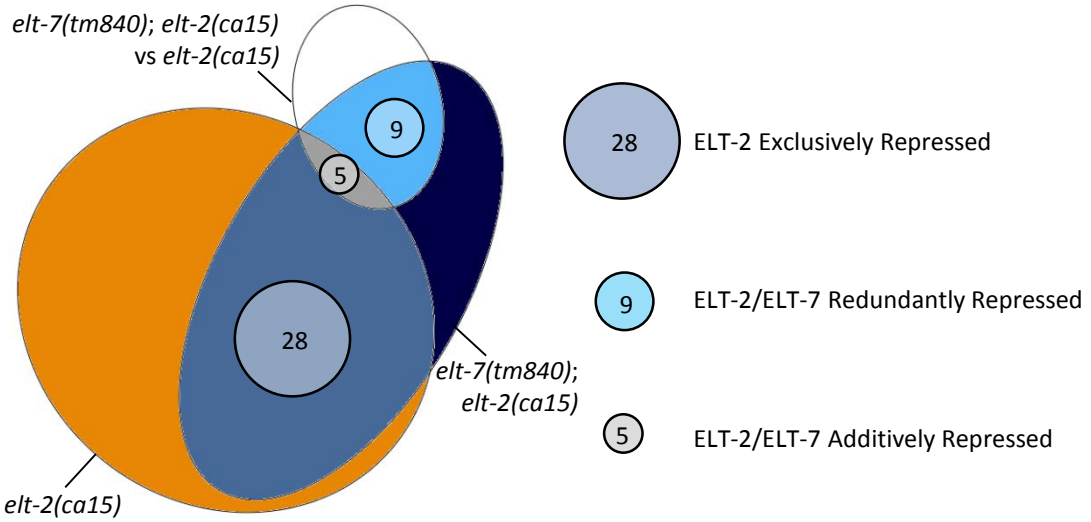
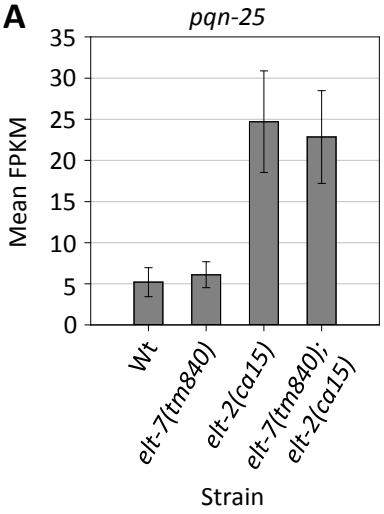


Figure 4.16. RNAseq results for representative genes from the ELT-2 exclusively repressed class A. *pqn-25*; the ELT-2/ELT-7 redundantly repressed class B. *odd-2*; and the ELT-2/ELT-7 additively repressed class C. *cebp-1*. Left: Graph of the mean FPKM values in each genetic background. Right: Table with the log2 fold change difference between genetic backgrounds and associated adjusted p values. Corresponding data can be found in Tables 4.10, 4.11 and 4.12.

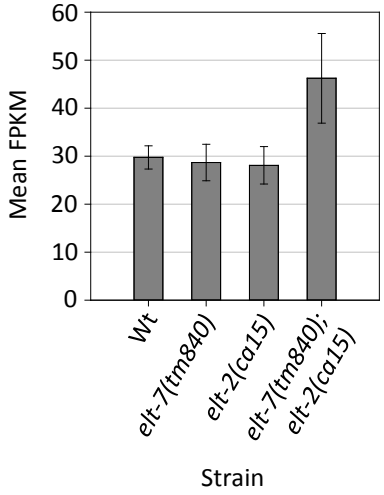
A



ELT-2 Exclusively Repressed

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	0.28	0.32
Wt	*elt-2(ca15)*	2.16	4.89E-25
Wt	*elt-7(tm840); elt-2(ca15)*	2.00	4.49E-16
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	1.69	2.38E-11
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	-0.16	0.64

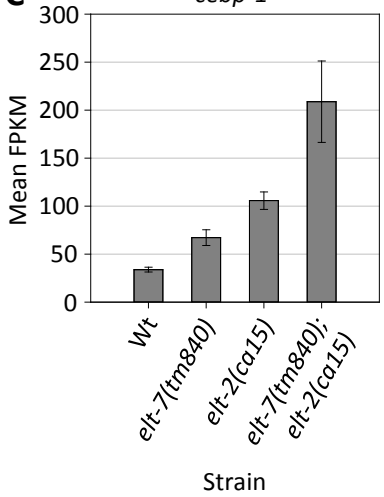
B



ELT-2/ELT-7 Redundantly Repressed

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	0.02	0.95
Wt	*elt-2(ca15)*	0.00	1.00
Wt	*elt-7(tm840); elt-2(ca15)*	0.58	1.06E-02
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	0.58	2.73E-02
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	0.56	1.36E-02

C



ELT-2/ELT-7 Independently Repressed

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	1.02	7.73E-17
Wt	*elt-2(ca15)*	1.66	1.47E-76
Wt	*elt-7(tm840); elt-2(ca15)*	2.63	4.76E-45
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	1.58	4.86E-13
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	0.96	3.31E-07

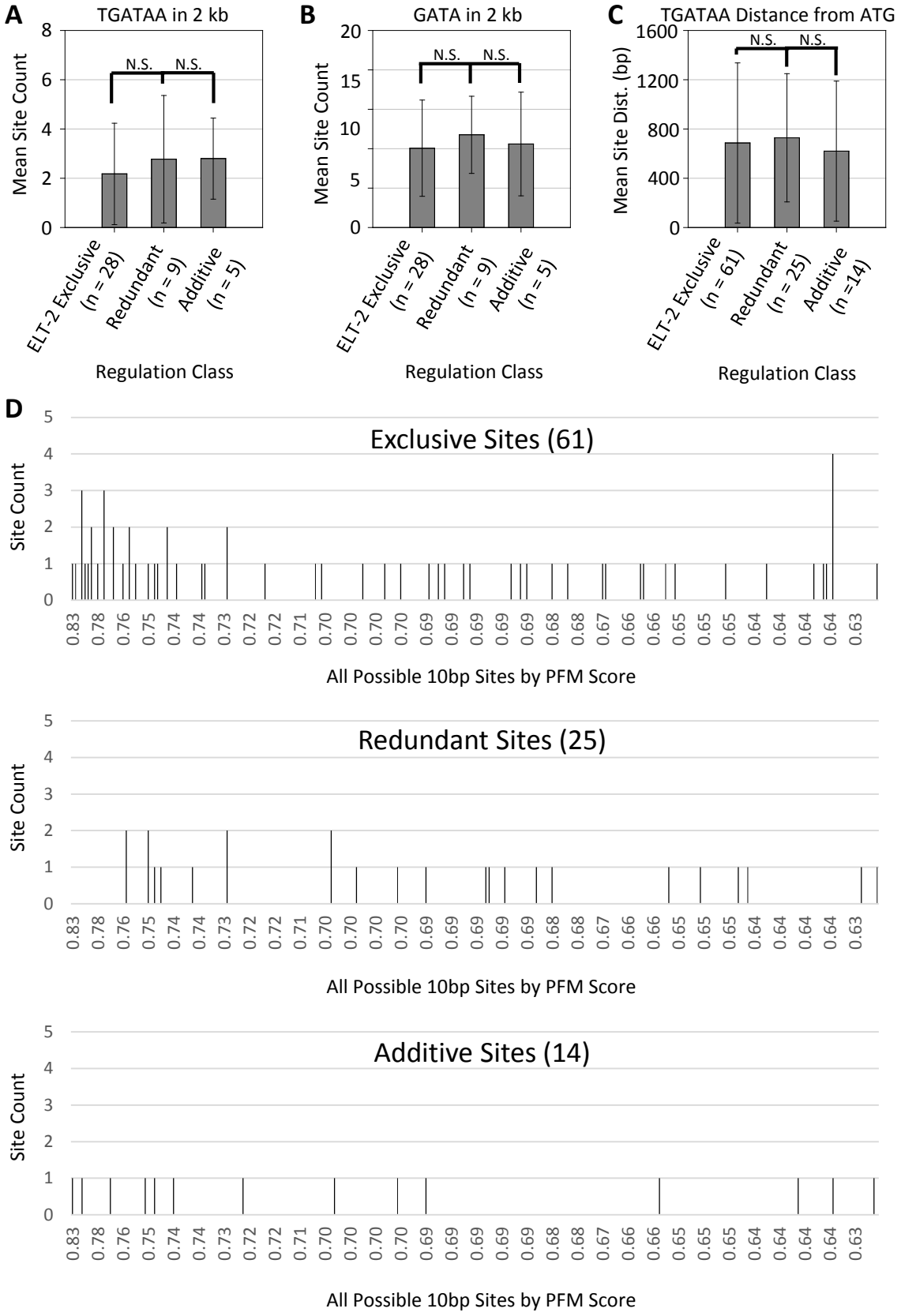
the *elt-7(tm840)* mutants, but was up-regulated in both the *elt-2(ca15)* populations and the double mutant populations.

ii) The nine intestinal genes that are redundantly repressed by ELT-2 or ELT-7 (Table 4.12) are only up-regulated in the double mutant compared to wild-type. This is what is observed for *odd-2* (Figure 4.16B), which is the *C. elegans* homolog of the Drosophila *Odd-skipped* like gene. There was no significant difference in mRNA levels between the wild-type and either of the single mutants. In contrast, the double mutants had significantly higher levels of *odd-2* mRNA compared to wild-type.

iii) The third class of up-regulated intestinal genes are those that can be repressed by both ELT-2 and ELT-7, presumably in an independent manner. One representative of these five genes, *cebp-1*, encodes a putative bZip transcription factor that appears to be negatively regulated by both ELT-2 and ELT-7 (Table 4.13). As seen in Figure 4.16C, transcription of this gene was moderately higher in the *elt-7(tm840)* mutants compared to wild-type and even more so in the *elt-2(ca15)* populations. The strongest effect was clearly observed in the double mutant population, which was up-regulated compared to both wild-type and the single mutants (Figure 4.16C).

The proximal 2 kb promoters of these three classes of repressed intestinal genes were examined to see if the same pattern of increased number of TGATAA sites was observed for the repressed genes. There was no difference in TGATAA site number between the ELT-2 exclusively repressed, the ELT-2 or ELT-7 redundantly repressed and the ELT-2/ELT-7 independently repressed classes of intestinal genes (Figure 4.17A). Additionally, there was no difference in the number of GATA sites in this region between any of the classes (Figure 4.17B), nor the distance of TGATAA sites from the first ATG (Figure 4.17C). Similar to the activated genes, there was no observed preference for site score between the classes (Figure 4.17D). These results are based on the analysis of a small number of genes so it is difficult to draw any real conclusions from this data. Furthermore, neither ELT-2 nor ELT-7 have been shown to directly repress an intestinal target gene so it is possible that these two factors are indirectly repressing transcription of these targets.

Figure 4.17. Analysis of the promoter regions (ATG up to -2 kb or next upstream gene) in each class of repressed intestinal genes. No significant difference between ELT-2 exclusively repressed, ELT-2/ELT-7 redundantly repressed and ELT-2/ELT-7 additively repressed for A. TGATAA sites per promoter; B. total GATA sites per promoter; C. mean distance of TGATAA site from ATG. N.S. denotes no significant difference. D. Bar graphs depicting all 256 possible 10 bp TGATAA sites and the number of times they appear in the promoters of genes from each class of intestinally repressed genes. Site score refers to the score of each sequence based on the position frequency matrix of TGATAA sites from intestinal gene promoters (McGhee et al, 2009).


A
TGATAA in 2 kb
Mean Site Count
N.S.
N.S.
ELT-2 Exclusive (n = 28)
Redundant (n = 9)
Additive (n = 5)
Regulation Class
B
GATA in 2 kb
Mean Site Count
N.S.
N.S.
ELT-2 Exclusive (n = 28)
Redundant (n = 9)
Additive (n = 5)
Regulation Class
C
TGATAA Distance from ATG
Mean Site Dist. (bp)
N.S.
N.S.
ELT-2 Exclusive (n = 61)
Redundant (n = 25)
Additive (n =14)
Regulation Class
D
Exclusive Sites (61)
Site Count
All Possible 10bp Sites by PFM Score
Redundant Sites (25)
Site Count
All Possible 10bp Sites by PFM Score
Additive Sites (14)
Site Count
All Possible 10bp Sites by PFM Score

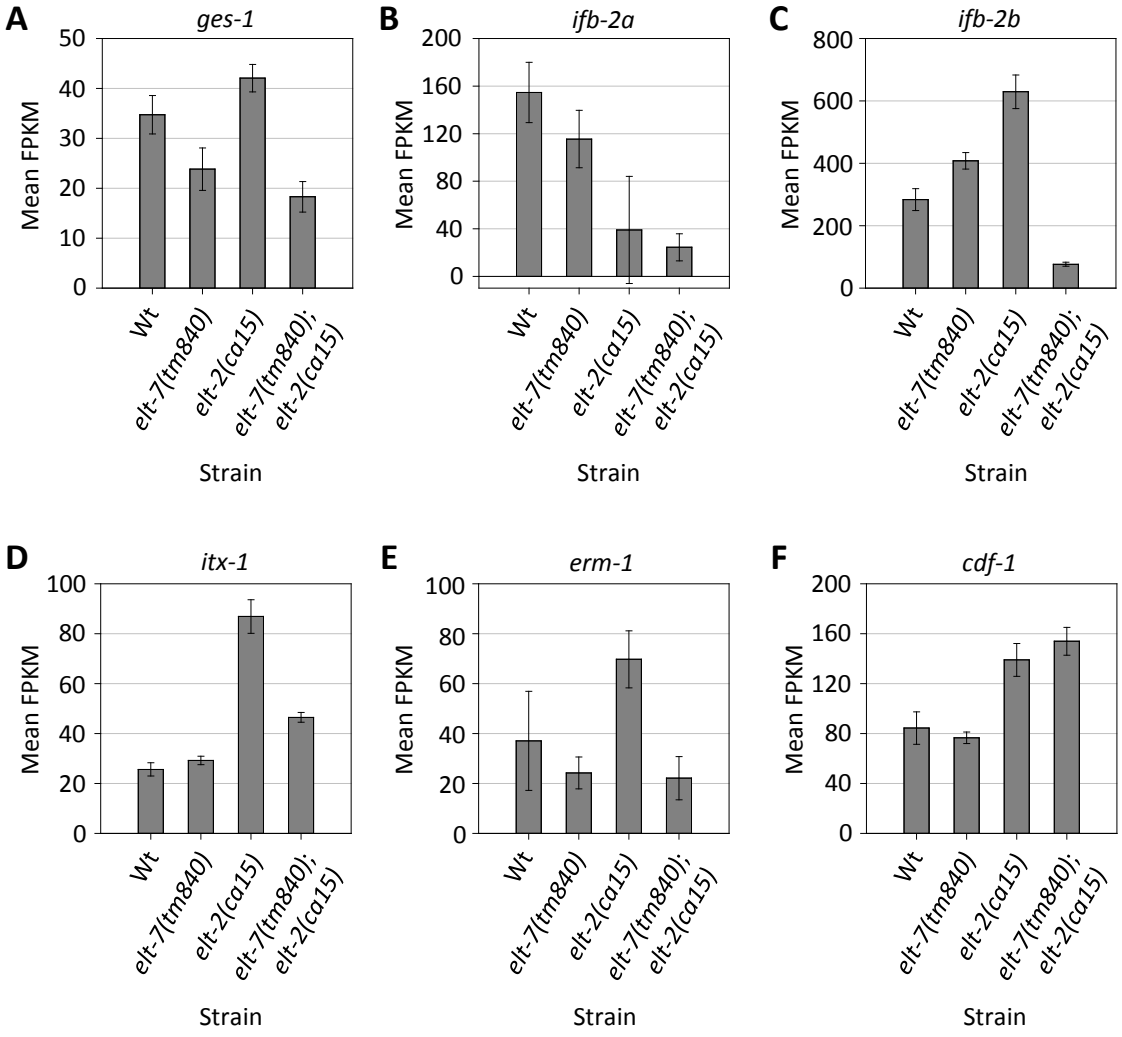
4.4.7 Known Targets of ELT-2 and ELT-7

Previous study of the role of ELT-7 in intestinal gene regulation concluded that loss of function mutations in *elt-2* and *elt-7* synergize and that both factors regulate expression of some target genes during embryogenesis (Sommermann et al., 2010). Their evidence and the results here suggest partial redundancy between ELT-2 and ELT-7, with a much more important role for ELT-2. To further investigate this possibility, the intestinal targets examined in (Sommermann et al., 2010) were analyzed. The authors of the previous study examined eight genes from three distinct categories; redundantly activated by ELT-2 or ELT-7 (*ajm-1, ges-1, ifb-2, itr-1* and *itx-1*) (see i-v below), exclusively activated by ELT-2 (*let-767* and *erm-1*) (see vi-vii below) and independently repressed by both factors (*cdf-1*) (see viii below).

i) *ajm-1* codes for an apical junction molecule expressed on the apical side of all epithelia in *C. elegans,* of which the intestine is just one tissue (Köppen et al., 2001). Expression of *ajm-1* mRNA was not detected in any of the eight different populations sequenced in this study (Table 4.14). This was perhaps due to an insufficient sequencing depth in the samples and suggests that *ajm-1* transcripts are expressed in low abundance in these starved L1 worms.

ii) The *ges-1* gene is expressed exclusively in the intestine and produces a type B carboxylesterase (Edgar and McGhee, 1986; Kennedy et al., 1993). Expression of *ges-1* has been shown to respond positively to ELT-2 levels but not require ELT-2 for expression (Fukushige et al., 1998). *ges-1* expression was significantly down-regulated in the *elt-7(tm840)* mutants and in the double mutants compared to wild-type (Table 4.14, Figure 4.18A). Unexpectedly, expression was significantly up-regulated in the *elt-2(ca15)* mutant populations relative to wild-type. This result does not fit with previous findings and is interpreted as due to the up-regulated levels of *elt-7* mRNA detected in the *elt-2(ca15)* mutants, as opposed to a repressive action by ELT-2 on expression of this gene. This conclusion is supported by the finding that the double mutants had significantly lower *ges-1* expression relative to the *elt-7(tm840)* mutants alone. These results suggest that both ELT-2 and ELT-7 activate *ges-1* expression, with ELT-7 appearing to have a stronger activation effect than ELT-2.

Figure 4.18. RNAseq results for intestinal genes previously examined in Sommermann et al. (2010). Data are presented as bar graphs of the mean FPKM values in each genetic background. *ifb-2a* and *ifb-2b* refer to different transcript isoforms of the *ifb-2* gene. Corresponding data can be found in Table 4.13.


A
ges-1
Mean FPKM
50
40
30
20
10
0
Wt
elt-7(tm840)
elt-2(ca15)
elt-7(tm840); elt-2(ca15)
Strain
B
ifb-2a
Mean FPKM
200
160
120
80
40
0
Wt
elt-7(tm840)
elt-2(ca15)
elt-7(tm840); elt-2(ca15)
Strain
C
ifb-2b
Mean FPKM
800
600
400
200
0
Wt
elt-7(tm840)
elt-2(ca15)
elt-7(tm840); elt-2(ca15)
Strain
D
itx-1
Mean FPKM
100
80
60
40
20
0
Wt
elt-7(tm840)
elt-2(ca15)
elt-7(tm840); elt-2(ca15)
Strain
E
erm-1
Mean FPKM
100
80
60
40
20
0
Wt
elt-7(tm840)
elt-2(ca15)
elt-7(tm840); elt-2(ca15)
Strain
F
cdf-1
Mean FPKM
200
160
120
80
40
0
Wt
elt-7(tm840)
elt-2(ca15)
elt-7(tm840); elt-2(ca15)
Strain

iii) The *ifb-2* gene encodes an intermediate filament protein expressed exclusively in the intestine (Hüsken et al., 2008). mRNA levels of *ifb-2* were found to be significantly up-regulated in both the *elt-7(tm840)* and *elt-2(ca15)* backgrounds but strongly down-regulated in the double mutants (Table 4.14). These contradictory results stem from the fact that DESeq2 calls differential expression on a per gene basis, even when there are multiple transcripts for a gene. As seen in Figure 4.18B and C, the Cufflinks generated FPKM values show two annotated transcripts for the *ifb-2* gene that are expressed at different levels in the different genetic backgrounds. The first transcript was observed to be down-regulated in both of the single mutants and in the double mutants (Figure 4.18B). On the other hand, the second transcript was expressed at a higher level than the first and appeared to be up-regulated in both single mutants but down-regulated in the double mutants (Figure 4.18C). This is interpreted as both ELT-2 and ELT-7 having activating effects on transcription of this gene, albeit with complex and completely unknown effects at the level of individual transcript isoforms.

iv) The *itr-1* gene was excluded from this analysis as it is expressed in multiple tissues, there are six independent transcripts for this gene, five of which were detected in the RNAseq samples, but all of which were expressed at low and variable FPKM levels. However, it should be noted that DESeq2 did call this gene as significantly down-regulated in the double mutants compared to wild-type, as noted in the (Sommermann et al., 2010) findings for isoform C (Table 4.14).

v) In the initial (Sommermann et al., 2010) paper, the authors found expression of the *C. elegans* Caspr ortholog (Yook et al., 2012) *itx-1* was down-regulated only when both ELT-2 and ELT-7 function was disrupted. The RNAseq results for the *itx-1* gene were opposite to what was reported, indicating up-regulation in the *elt-2(ca15)* mutants and the double mutants relative to wild-type.

vi) *let-767* mRNA was not detected in any of the samples but given its critical role in cholesterol metabolism (Desnoyers et al., 2007; Entchev et al., 2008; Kuervers et al., 2003) and expression in multiple tissues (Entchev et al., 2008; Kuervers et al., 2003) it is expected that this gene was expressed in the sampled worms.

vii) The gene *erm-1* was found to be only significantly different in the *elt-2(ca15)* null mutants, where it was up-regulated when compared to wild-type (Figure 4.18E,

Table 4.14). This gene is expressed in many tissues (Göbel et al., 2004; Haag et al., 2014) and therefore sequencing mRNA from whole worms may not be sensitive enough to detect all of the transcriptional differences occurring within a single tissue in these populations.

viii) The eighth gene to be examined was *cdf-1*, a cation diffusion facilitator expressed in the intestine and vulva (Bruinsma et al., 2002). Reporter expression of this gene was actually found to be negatively regulated by both ELT-2 and ELT-7 (Sommermann et al., 2010) study. The RNAseq results show that there was no difference in expression levels in the *elt-7(tm840)* mutants compared to wild-type, but *elt-2(ca15)* mutants were significantly up-regulated for *cdf-1* mRNA (Figure 4.18F, Table 4.14). The double mutants were not significantly different from the single *elt-2(ca15)* nulls suggesting that only ELT-2 is repressing transcription of this gene.

In summary, there were some differences in the results presented here for the eight genes compared to those reported by Sommermann et al., (2010). Of the five genes that are exclusively or primarily expressed in the intestine, four were detected as expressed in these samples and the RNAseq results for three of these genes are mostly in agreement with the previous study. These results provide further evidence of ELT-2/ELT-7 redundancy and/or additivity.

4.4.8 Intestinal Transcription Regulation More Complex than ELT-2 and ELT-7

A list of transcription factors expressed in the intestine was compiled by comparing a list of all transcripts known to be expressed in the adult intestine (E. Osborne-Nishimura, B. Goszczynski, J. McGhee, personal communication) with the previously published annotation of all genes encoding transcription factors in *C. elegans* (Reece-Hoyes et al., 2005). A total of 379 other transcription factors (381 with ELT-2 and ELT-7) were found to be expressed in the adult intestine and analyzed for differential expression between the various genetic backgrounds of these L1 larvae.

There were 95 transcription factors found to be differentially expressed (compared to wild-type, any log2 fold change, p adj < 0.1) in the single *elt-2(ca15)* mutants, single *elt-7(tm840)* mutants and the double mutants (Table 4.15). Almost half of all differentially expressed transcription factors (41) are nuclear hormone receptors, but

this is not surprising given the vast expansion of nuclear hormone genes in *C. elegans* (284) (Antebi, 2006) and the number that were found to be expressed in the adult intestine (140). The rest constituted a selection from a variety of transcription factor families (Table 4.15).

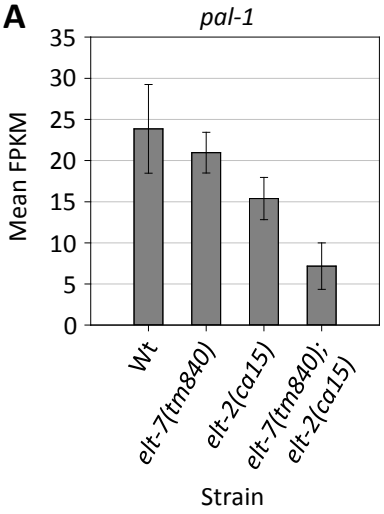
Some *trans*-acting factors are activated by both ELT-2 and ELT-7, such as the homeodomain *caudal* ortholog *pal-1* (Figure 4.19A), which is known to have important roles in many embryonic lineages in the formation of midline structures (Edgar et al., 2001). Levels of *pal-1* mRNA were not observed to be significantly reduced in the *elt-7(tm840)* worms. However, they were down-regulated in both the *elt-2(ca15)* mutants and the double mutants, with the double mutants also significantly lower than *elt-2(ca15)* mutants alone. This suggests a role for both GATA factors in activating *pal-1* expression in the E lineage.

The *mxl-1* gene encodes a *C. elegans* homolog of vertebrate myc/mad/max family of transcription factors and is expressed in the posterior intestine and some neurons (Yuan et al., 1998). Expression of this gene was significantly decreased in the populations harbouring the *elt-2(ca15)* deletion indicating that ELT-2 is required in part for normal levels of transcription from this gene (Figure 4.19B). Interestingly, *mxl-1* was previously identified as one of the top 15 transcription factors up-regulated in starved L1 larvae, along with the forkhead transcription factor *fkh-7* (Baugh et al., 2009). In contrast to the results for *mxl-1*, transcription from the *fkh-7* gene was found to be up-regulated relative to wild-type in both single mutants and the double mutant, suggesting independent repression by both ELT-2 and ELT-7 (Figure 4.19C). A second forkhead transcription factor, *fkh-9*, is known to be expressed in the intestine (among other tissues) (Hope et al., 2003), but was not detected in any of the populations analyzed here.

The sterol regulatory element binding protein (*sbp-1*) is a critical activator of lipid metabolism required for growth and development of *C. elegans* (Ashrafi et al., 2003; Kniazeva et al., 2004; McKay et al., 2003) and is exclusively expressed in the intestine (McKay et al., 2003). Other than ELT-2, it is the only identified intestinal transcription factor for which loss of intestinal function mutations lead to larval arrest (McGhee et al., 2009). Expression of this gene may depend upon direct activation by ELT-2 (McGhee et al., 2009), but mRNA levels were not found to vary at all between any of the populations

Figure 4.19. RNAseq results for representative genes encoding transcription factors that are differentially expressed in various genetic backgrounds. A. *pal-1*. B. *mxl-1*. C. *fkh-7*. Left: Graph of the mean FPKM values in each genetic background. Right: Table with the log2 fold change difference between genetic backgrounds and associated adjusted p values. Corresponding data can be found in Table 4.14.

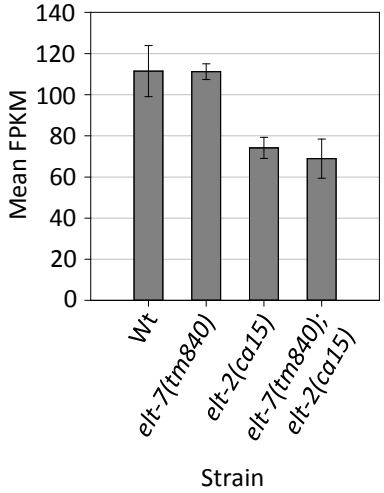
A



ELT-2/ELT-7 Activated

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	-0.17	0.35
Wt	*elt-2(ca15)*	-0.60	2.00E-05
Wt	*elt-7(tm840); elt-2(ca15)*	-1.31	2.72E-18
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	-1.11	1.23E-09
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	-0.71	2.00E-04

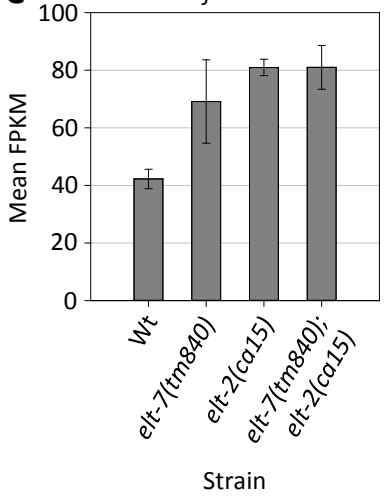
B



ELT-2 Activated

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	-0.11	0.48
Wt	*elt-2(ca15)*	-0.71	6.97E-13
Wt	*elt-7(tm840); elt-2(ca15)*	-0.79	2.17E-09
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	-0.66	6.40E-05
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	-0.08	0.71

C



ELT-2/ELT-7 Independently Repressed

Control Strain	Experimental Strain	Log2 Fold Change	p adj
Wt	*elt-7(tm840)*	0.67	7.05E-07
Wt	*elt-2(ca15)*	0.65	2.07E-17
Wt	*elt-7(tm840); elt-2(ca15)*	0.70	1.98E-06
elt-7(tm840)	*elt-7(tm840); elt-2(ca15)*	0.02	0.94
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	0.05	NA

analyzed. The conclusion from this analysis of intestinal transcription factor expression is that regulation of intestinal gene expression appears to be complex, likely involving many different factors that potentially work in combination with ELT-2 and ELT-7. It is also likely that many of these other transcription factors are under the control of ELT-2 and ELT-7.

4.5 Discussion

4.5.1 ELT-2 is the Major Effector of Intestinal Transcriptional Regulation

Our lab's current model of intestinal differentiation holds that ELT-2 is the master regulator of intestinal gene expression and the only necessary GATA transcription factor required post-specification in this lineage (Fukushige et al., 1998; Kormish et al., 2010; McGhee, 2013; McGhee et al., 2007; McGhee et al., 2009). Of the other two GATA transcription factors expressed exclusively in the intestine from embryogenesis to adult, neither ELT-7 nor ELT-4 are necessary to coordinate and maintain a differentiated intestine (Fukushige et al., 2003; McGhee et al., 2007). Strikingly, a quadruple mutant lacking the E cell specifying GATA factors *end-1/end-3* as well as *elt-7* and *elt-4* can be rescued by over-expressing ELT-2 two cell cycles early in the E cell (T. Wiesenfahrt, personal communication). ELT-7 has been proposed to function with ELT-2 as evidenced by loss of some intestinal gene reporter expression in single *elt-7* mutants or RNAi knockdown (Murray et al., 2008) and in double mutants with *elt-2* (Sommermann et al., 2010). Here, the functional relationship between ELT-2 and ELT-7 has been examined to fully understand how each GATA factor contributes to the regulation of transcription of intestinal targets. These results provide further proof that ELT-2 is the major GATA transcription factor in the intestine while ELT-7 is partially redundant, as supported by three independent lines of evidence.

The first line of evidence is that there were more intestinal genes differentially expressed in the *elt-2(ca15)* null mutants than there were in the *elt-7(tm840)* mutants when compared to wild-type. The basic conclusion from this result is that ELT-2 contributes to the transcriptional regulation of many more genes in the intestine than ELT-7. This conclusion is supported by the fact that there were more intestinal genes

identified as regulated exclusively by ELT-2 than those regulated redundantly by either ELT-2 or ELT-7. There was only a very small number of genes identified as involving activation from both ELT-2 and ELT-7 or ELT-7 alone for normal expression levels. Furthermore, there was no intestinal gene identified as regulated synergistically by ELT-2 and ELT-7. Together, these finding indicate that the majority of intestinal genes are expressed at correct levels by the action of ELT-2, even in the absence of ELT-7. It is possible that a large number of intestinal genes that are regulated by ELT-7 were not identified from these datasets due to the incomplete nature of the list of intestinally expressed genes used in this analysis. The consequence of this would therefore be that ELT-7's role has been underestimated. However, this is not likely the case since there were approximately three times as many genes in general that were differentially expressed in the *elt-2(ca15)* mutants relative to wild-type than in the *elt-7(tm840)* mutants, suggesting that if all genes with intestinal expression were included in the analysis, the results would have still reflected more genes being regulated by ELT-2.

These conclusions fit with other findings that suggest ELT-2 has widespread involvement in diverse aspects of intestinal gene expression. For example, recovery from acute bacterial infection has been reported to encompass down-regulation of innate immunity genes and up-regulation of detoxification and homeostasis genes, all of which depend on ELT-2 activity (Head and Aballay, 2014; Kimura et al., 2012). Similarly, the Gram-negative soil bacterium *Burkholderia pseudomallei* has been shown to bypass *C. elegans* intrinsic defense mechanisms by specifically targeting ELT-2 for degradation, resulting in significant down-regulation of ELT-2 targets during infection (Lee et al., 2013). ELT-7 also appears to be important in mediating response to infection (Elliott et al., 2011), consistent with a partially redundant role in intestinal gene regulation. However, ELT-2 has been reported as the major GATA factor regulating response to infection (Kerry et al., 2006; Shapira et al., 2006). Additionally, ELT-2 has also been shown to regulate transcription of iron, zinc and heme responsive genes (Moilanen et al., 1999; Roh et al., 2014; Romney et al., 2008; Sinclair and Hamza, 2010) and is implicated in the regulation of metabolism, aging and oxidative stress genes (Bansal et al., 2014; Burmeister et al., 2008; Cypser et al., 2013; Kenyon, 2010; Son et al., 2011). This

abundance of evidence all points to the conclusion that ELT-2 is the major intestinal regulator of gene expression.

The second line of evidence that ELT-2 is a major activator and ELT-7 is a minor partially redundant activator is the effect on transcription each factor appears to have. mRNA expression for some intestinal transcripts in the *elt-2(ca15)* null mutants was a small fraction of the wild-type levels, suggesting that in these worms ELT-2 is not just the major GATA factor activating transcription of these genes, but is the predominant transcription factor activating these genes. It also implies that ELT-7 is incapable of significantly activating transcription of some intestinal genes in the absence of ELT-2. In contrast, not a single intestinal transcript that was observed to be activated by ELT-7 was more than half down in the *elt-7(tm840)* mutants compared to wild-type levels. Even if some of these transcripts are regulated completely independently of ELT-2, there must be other intestinal transcription factors contributing to their regulation.

With respect to ELT-2 function, there is evidence that this GATA factor can work in combination with other activators in the intestine. Such examples include teaming up with the CSL/notch effector LAG-1 to activate transcription of the bHLH factor *ref-1* (Neves et al., 2007), the mediator subunit MDT-15, which is proposed to help ELT-2 interact with an as yet unidentified zinc response factor to activate *mtl-1/mtl-2* expression (Moilanen et al., 1999; Roh et al., 2014), and the unidentified iron response factor ACT to activate the iron response genes *ftn-1*, *ftn-2* and *smf-3* (Romney et al., 2011). ELT-2 directly activates the *vit-2* vitellogenin gene in opposition to repression by MAB-3 (Inoue and Nishida, 2010; Yi and Zarkower, 1999). Thus, it is not hard to imagine a situation where ELT-7 also works in combination with other *trans*-acting factors to regulate intestinal gene expression.

The third line of evidence of ELT-2's dominant role is the phenotypes of the individual null mutants themselves. The *elt-2(ca15)* mutant arrests after hatching due to a non-functioning intestine whereas the *elt-7(tm840)* mutant hatches about 30 minutes before wild-type embryos do, but is otherwise not obviously different from wild-type. If these two GATA factors were equally important for activating expression of intestinal genes for differentiation and were completely redundant, then one would expect no phenotype in the single loss of function mutants. This raises the question of what is the

function of ELT-7 in the E lineage if it is not regulating expression of many terminal intestine genes.

4.5.2 Role of ELT-7 in the Intestinal Transcriptional Hierarchy

The *C. elegans* embryonic endoderm transcriptional network exhibits both a hierarchy and redundancy of GATA transcription factors that specify and differentiate the E lineage. Many other organisms appear to employ significant redundancy in the GATA factors that control the development of endoderm structures as outlined above. Comparatively, it appears that ELT-7 is only partially redundant to ELT-2 for the function of driving differentiation of the intestine. The results presented here for known joint targets of ELT-2 and ELT-7 differ from those presented in (Sommermann et al., 2010). The data presented here is likely a more accurate measure of how some of these genes are regulated at the level of transcription because it measured levels of the endogenous mRNA transcript as opposed to subjective scoring of extrachromosomal GFP reporter transgenes and antibody staining. It is possible that the transgenes used in (Sommermann et al., 2010) lacked some of the necessary *cis*-regulatory elements involved in transcriptional regulation, which may account for minor differences observed for these genes. Moreover, of the eight genes examined by Sommermann et al. (2010), expression of two were not detected here and another two were expressed in multiple tissues making it difficult to detect any differences in intestinal expression with the RNAseq approach. Even so, many of results presented here for these genes were in agreement. The ELT-2/ELT-7 dual target genes identified in that study become activated early in the E lineage such as *ges-1* (Edgar and McGhee, 1986; Kennedy et al., 1993) and *ifb-2* (Hüsken et al., 2008). Transcriptional activation of *ges-1* occurs at the 4E cell stage, approximately when *elt-2* is also transcriptionally activated and therefore it is not surprising to find that *ges-1* responds more strongly to ELT-7 activation than ELT-2. The fact that there was residual expression in the double mutants indicates the transiently expressed END-1 and END-3 GATA factors may also be activating *ges-1* during embryogenesis. Both studies support the conclusion that ELT-2 and ELT-7 have restricted overlapping functions during embryogenesis and E lineage differentiation.

However, neither study detected any experimental evidence of synergy between these two factors, only redundancy and limited additivity.

Ectopic expression of ELT-7 in the embryo is capable of activating terminal differentiation in pharyngeal cells after they are committed to their normal cell fate (Riddle et al., 2013). Interestingly, ELT-7 is capable of mediating this transdifferentiation for a wider period during embryogenesis than ELT-2 (Riddle et al., 2013). This suggests that there is something distinctive about the function of ELT-7 during embryogenesis, which is in agreement with the embryonic phenotype described here. Considering all the evidence, it is proposed that ELT-7 is partially redundant with ELT-2 and has a primary function as a supporting regulator of intestinal gene expression during early embryogenesis. However, ELT-2 is the major activator of intestinal gene expression and is capable of regulating differentiation of E cells without ELT-7 activity (Figure 4.20A).

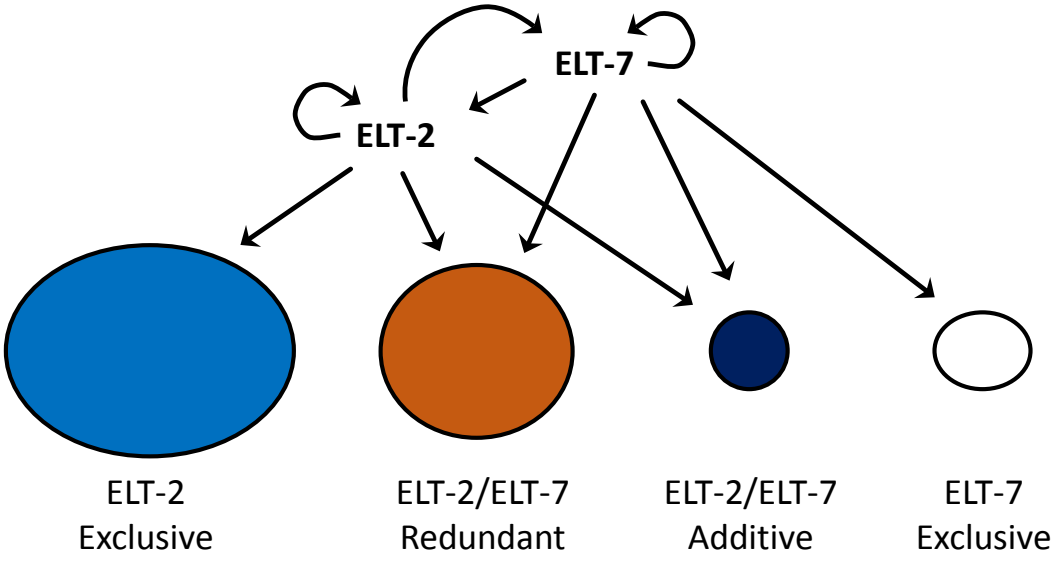
It is possible that ELT-7 has other important functions for intestinal gene regulation not experimentally tested yet. This GATA factor may play a specific role in regulating appropriate gene expression responses during environmental stresses such as low or high temperatures or pH. By the same token, the fact that ELT-7 is partially redundant to ELT-2 could be a general feature of the intestinal transcriptional network that enables the worm to better deal with intestinal stress situations and to maintain stable gene expression. ELT-7 has conserved homologs in other *Caenorhabditae* species including *briggsae, remanei, japonica* and *brenneri*, implying that there is a functional importance for this gene in nematode intestinal biology.

4.5.3 More TGATAA Sites in Promoters of Genes Regulated by ELT-2 and ELT-7

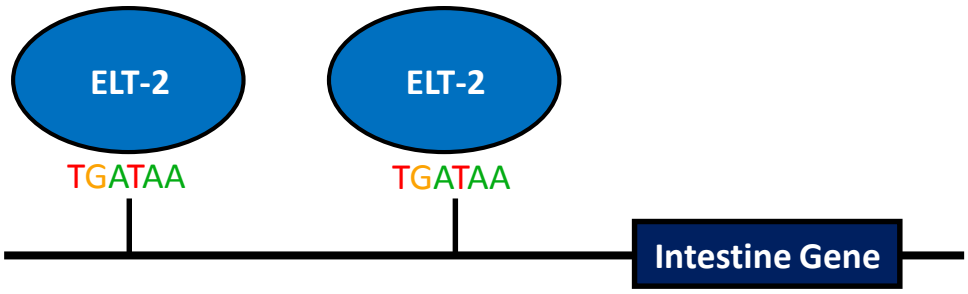
The observation of more TGATAA (and GATA) sites in the proximal 2 kb promoter region of intestinal genes that have regulatory input from both ELT-2 and ELT-7 is a clue that may help understand how these factors function. There was no significant difference in the number of TGATAA sites within 5 kb of the first ATG, but there was an apparent trend. The simple interpretation of this result is that there are GATA regulatory elements important for transcriptional activation beyond those found in the 2 kb upstream region, and that intestinal genes that respond to ELT-7 input also have more sites beyond 2 kb than those that are exclusively regulated by ELT-2. There was no obvious

Figure 4.20. Model of how activation of intestinal genes occurs for A. Intestinal genes activated exclusively by ELT-2. B. Intestinal genes activated redundantly either by ELT-2 or ELT-7.

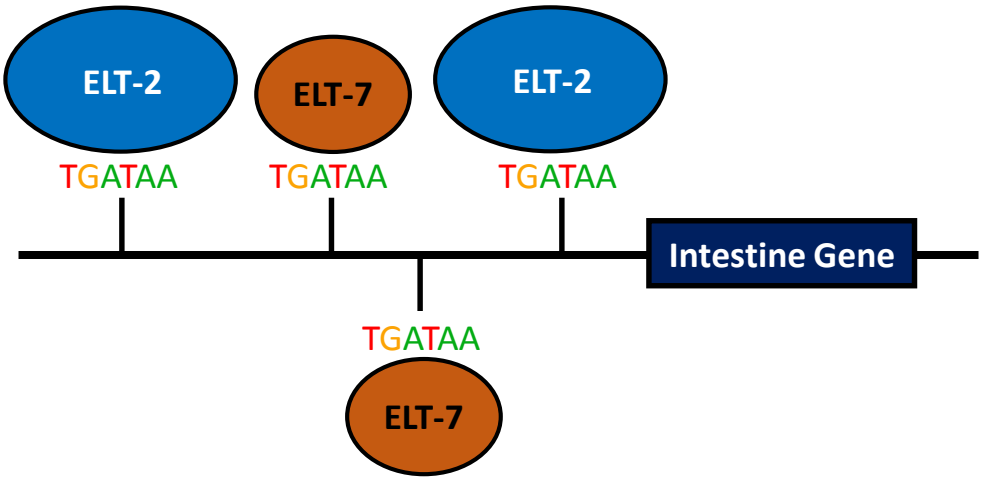
A



B ELT-2 Exclusively Activated



C ELT-2/ELT-7 Redundantly Activated



distribution or clustering of TGATAA sites in the proximal 2 kb promoter indicating that it is unlikely that ELT-2 and ELT-7 are binding cooperatively. Such co-operative action between transcription factors has been observed for zinc responsive intestinal genes where the zinc responsive *cis*-regulatory element (HZA) was reported as frequently found adjacent to a functional TGATAA binding site (Roh et al., 2014). Instead, it appears the promoters studied here exhibit a general accumulation of potential intestinal GATA factor binding sites.

The finding that either ELT-2 or ELT-7 can redundantly activate expression of many intestinal genes begs the question of whether it is direct or indirect regulation by each factor. ELT-2 has been predicted to directly regulate over 80% of intestinal transcripts (McGhee et al., 2009). Moreover, of the redundantly activated intestinal genes expressed primarily or exclusively in the intestine, the vast majority (almost 90%) had two or more TGATAA sites within their promoter. It is therefore reasonable to predict that many of these genes are directly activated by both of these two factors.

Slight differences in the TGATAA flanking nucleotides were observed between the different classes of activated intestinal genes, which could underlie a functional divergence in binding preferences of ELT-2 and ELT-7. In the event that this is true, it would also be expected that some TGATAA sites in redundant promoters would be preferentially bound by ELT-2 and others by ELT-7. To date, there is no evidence that this is how these two GATA factors mediate their activity on targets gene. For example, both ELT-2 and ELT-7 are capable of binding TGATAA sites with different flanking nucleotides in the ELT-2 promoter (T. Wiesenfahrt, B. Goszczynski, M. McGhee, personal communication). Still, the observation that the redundant class of intestinal genes most frequently had A (over C) at position two of the ten base pair sequence and A (over G) at position nine suggests that there may be some differences in affinities for TGATAA sites between these two transcription factors. The finding that ELT-2 regulates more intestinal genes might imply that the PFM generated from all intestinal genes (McGhee et al., 2009) is the preferred binding site of ELT-2, and that ELT-7 has a different ten base pair TGATAA site that it binds with the strongest affinity. Divergence of DNA binding preferences may indeed be the case as it has been shown that evolutionarily, the *elt-7* gene is almost as closely related to the hypodermal GATA factor

gene *elt-3* as it is to *elt-2* (Gillis et al., 2008). At the protein level, ELT-7 is approximately half the length and half the molecular weight of ELT-2 (Hawkins and McGhee, 1995; McGhee et al., 2009; Sommermann et al., 2010), so it is reasonable to predict that these two proteins may use different mechanisms for activating transcription of target genes.

Given all the results to date, including those presented here, it is hypothesized that both ELT-2 and ELT-7 can bind to most if not all TGATAA sites in the promoters of redundantly activated intestinal genes to directly activate them (Figure 4.20B, C). Thus, the difference in regulation mechanisms of the ELT-2 exclusive class and the ELT-2/ELT-7 redundant class would be the number of TGATAA sites in the promoter. One prediction of this model would be that ELT-2 would outcompete ELT-7 for binding to exclusive gene promoters. Alternatively, it may be that ELT-2 is better able to mediate transcriptional activation of these genes than ELT-7 via interaction with other *trans*-acting factors. This model does not exclude the potential for some preferences for each factor at different sites based on flanking nucleotides but it is predicted that this is a minor feature of intestinal gene activation. It will be interesting to see if this model can be applied to redundant endodermal GATA transcription factor networks in other organisms. Furthermore, it is possible that the hypodermal ELT-1 and ELT-3 GATA transcription factors that work together to activate hypodermal genes function by a similar mechanism.

4.5.4 ELT-2 and ELT-7 also Function as Repressors in the Intestine

Previous analysis has identified that most intestinal genes are activated by ELT-2, but some do appear to be repressed (McGhee et al., 2009). To date, there has been no study that identified an intestinal gene that is repressed by direct action on the promoter. There were a large number of genes found to be up-regulated in both of the single mutants and the double mutants when compared to wild-type. This is in contrast with the results presented here where more intestinal genes are up-regulated than down-regulated in the *elt-2(ca15)* mutants.

It is important to highlight that while more genes were up-regulated than down-regulated in the *elt-2(ca15)* null mutants, this could be due to other factors including

varying levels of ELT-7 expression. Genes activated by both ELT-2 and ELT-7, including *ges-1, erm-1* and *ifb-2* were significantly up-regulated in the *ca15* mutant background relative to wild-type but down-regulated in the double mutant, a confusing dichotomy. However, expression of *elt-7* mRNA was highest and most stable in the *elt-2(ca15)* populations, indicating that any genes redundantly activated by ELT-2 or ELT-7 likely were up-regulated in these worms due to this increase. This fits with the data presented by Sommermann et al. (2010), which indicated that *ges-1, erm-1* and *ifb-2* were not observed to be up-regulated in an *elt-2* loss of function background. Consequently, this variability in *elt-7* levels could account for the unexpectedly higher number of up-regulated genes in the *elt-2(ca15)* mutants. Additionally, more genes could have been observed to be up-regulated due to indirect effects. ELT-2 is generally thought of as a global activator of intestinal gene expression but there are many other transcription factors expressed in the intestine that could be mediating repression of these genes downstream of ELT-2. Undoubtedly some of these repressed genes are indirect targets of ELT-2 and ELT-7, as many of these genes are expressed in multiple tissues. It is even possible these results reflect changes in gene expression beyond the intestine as total RNA was extracted from whole worms.

However, there are descriptions of intestinal genes only becoming activated in the correct environmental conditions, such as the zinc responsive genes (Freedman et al., 1993; Moilanen et al., 1999; Roh et al., 2014), iron responsive genes (Romney et al., 2008; Romney et al., 2011) and heme responsive genes (Severance et al., 2010; Sinclair and Hamza, 2010). This suggests that there could very well be genes that are directly repressed by ELT-2 until the appropriate environmental, temporal or developmental stimuli occurs.

Certainly a number of these repressed genes have multiple TGATAA sites within their promoters suggesting that it is possible that ELT-2 and ELT-7 are directly repressing transcription of these genes. For example, *cpr-3* encodes a cysteine protease (Maeda et al., 2001) that was strongly up-regulated in the *elt-2(ca15)* mutants relative to wild-type and has five TGATAA sites (16 other GATA sites) within the proximal 2 kb of its first ATG. This gene is primarily expressed in the intestine and would be a good

candidate to test if regulation of transcription is directly repressed by ELT-2 binding to this promoter.

There are examples of genes for which ELT-2 is insufficient to activate without the combination of another activating factor, for instance the heme responsive gene *hrg-1*. The *hrg-1* gene is exclusively expressed in the intestine and activated only upon binding of a co-activator (and possible release of binding of a repressor) to the heme response element (HERE) (Sinclair and Hamza, 2010). Despite the presence of five conserved GATA sites, ELT-2 can only activate this gene in response to low heme levels and this requires the HERE sequence. It is possible that another model is true, one where ELT-2 can function as both a repressor and an activator depending on the different co-factors with which it is interacting. In this model, transcription of *hrg-1* does not occur when a repressor co-factor is bound to the HERE and ELT-2, but only upon release of repressor binding and subsequent co-activator binding. There is precedent for GATA transcription factors acting in this manner. ELT-2's closest related vertebrate homolog GATA-4 has been shown to both activate and repress target endothelial genes in an artificial way, when fused to either the VP16 transcriptional activator or engrailed repressor (Kamei et al., 2011). Also, the more distantly related GATA-3 transcription factor is purported to regulate CD4 T cell differentiation by directly activating and repressing targets genes (Yagi et al., 2011). If this hypothesis about ELT-2 dual functionality in transcriptional regulation dynamics of the intestine is correct, it might also apply to intestinal genes that are spatially expressed, such as *pho-1*.

4.5.5 Regulation of Intestinal Genes Involves Many Transcription Factors

The intestine is an organ with a simple tube structure and yet it carries out many fundamental functions for the worm. This organ must be capable of responding to a variety of conditions including developmental progression or environmental stress such as nutrient availability and temperature. Therefore, it is likely there are multiple inputs that can modify expression of intestinal genes that must at least in part be mediated by other transcription factors. Similarly, not all intestinal genes are expressed in every intestinal cell or at all stages of the worm's life cycle. ELT-2 and ELT-7 on the other hand are expressed at uniform levels in all intestinal cells from early embryogenesis until

the death of the animal. Therefore, spatial and temporal patterns of gene expression in the intestine must be mediated by other transcription factors. Within the intestine, *pal-1*, is expressed only in the two int-V cells (Edgar et al., 2001). Hence *pal-1* may be involved in spatial patterning of intestinal gene expression. This would indicate a somewhat conserved role in gut development as caudal is known to be necessary for development of the *Drosophila* hindgut (Lengyel and Iwaki, 2002).

There were a number of transcription factors observed as differentially expressed in both of the single mutants and the double mutants compared to wild-type. These included POP-1, a factor involved in specification of the E lineage and other factors known to interact with it (BAR-1, PEB-1). Expression of *pop-1* mRNA was up-regulated in the *elt-7(tm840)* mutants, which provides more evidence that ELT-7 is key (but not essential) during embryogenesis. One caveat to this proposal is that *pop-1* is expressed in many different lineages (Lin et al., 1998), so these results are not proof that ELT-7 regulates expression of this gene during embryogenesis. Many of the transcription factors were inconsistently differentially expressed, but a number were either always up or always down-regulated in each single mutants and the double mutants relative to wild-type. Those factors consistently up-regulated upon loss of ELT-2 and ELT-7 comprised a wide range of types of transcription factors including Y-box domain, bHLH, forkhead, NHRs, zinc fingers and bZIPs. Up-regulated transcription factors could be activators of genes found to be repressed by the action of ELT-2 and ELT-7 and thus may be the mechanism by which the two GATA factors indirectly regulate repression of some target genes. Those consistently down-regulated were less diverse, mainly NHRs and zinc fingers. Of all the transcription factors expressed in the intestine only ELT-2 is necessary, which suggests that it is acting as an ON/OFF switch in the intestine, partially backed up by ELT-7. The hundreds of other *trans*-acting factors expressed in the intestine must play a less significant role in gene expression, such as mediating responses to the environment and development to fine tune levels of gene expression.

Three transcription factors stand out as differentially expressed in these strains; a forkhead transcription factor *fkh-7* and two bHLH factors similar to vertebrate MAX proteins, *mxl-1* and *mxl-2*. These genes were identified as three of the 15 most significantly up-regulated transcription factors during starvation in L1 worms and are

likely involved in regulating the gene expression in response to starvation (Baugh et al., 2009). The changes in gene expression levels reported here for these factors indicate that there was no global difference in gene expression between the strains due to the effects of starvation on transcription. Gene expression changes that take place during starvation are stable from six to 24 hours post-hatching without food (Baugh et al., 2003; Maxwell et al., 2012). If there were differences in transcription levels due to the transition of gene expression states from hatching to a starved state, then this would be reflected by a similar pattern of expression of *fkh-7* and *mxl-1* between each strain. However, *fkh-7* was consistently up-regulated in the mutants relative to wild-type while *mxl-1* transcription was down-regulated when in *elt-2(ca15)* mutants and the double mutants. Another line of evidence that supports this conclusion is the expression of a bHLH transcription factor *sbp-1*, which is a critical regulator of growth (Kniazeva et al., 2004). This intestinal transcription factor activates expression of the fatty acid elongases *elo-5* and *elo-6*, which function to produced monomethyl branched-chain fatty acids (mmBCFAs) that are essential for *C. elegans* larval development and viability (Kniazeva et al., 2004). Expression of *sbp-1* was not observed to be significantly different in any of the populations sequenced confirming that the larvae in each population were all in a stable starvation state. TGATAA sites in the *sbp-1* promoter are necessary for expression of this gene (J. Kalb, J. McGhee personal communication), suggesting that there is GATA transcription factor regulation of this gene. It is possible that the starved nutritional status prevented any differences in *sbp-1* expression between the different genetic backgrounds. *elo-6* was found to be in the ELT-2 exclusively activated class while *elo-5* was identified as an additively activated target of ELT-2 and ELT-7, signifying that SBP-1 acts in combination with these GATA factors to regulate target intestinal genes.

An intriguing finding of starvation induced gene expression was that RNA polymerase II was found paused on many growth and development genes (Baugh et al., 2009). Thus, one could predict that some intestinal genes were greatly reduced in expression in all populations and hindered the identification of ELT-2 and ELT-7 intestinal targets. In retrospect, it would have been advantageous to also perform RNAseq on starved populations that had been re-fed for an hour. Once accounting for feeding induced changes in gene expression as reported by (Baugh et al., 2009), this hypothetical

data could have been mined for changes in intestinal gene expression due to the underlying genetic backgrounds.

4.5.6 Negative Feedback on ELT-2 and ELT-7 Expression

The two key features of the *C. elegans* endoderm GATA transcription factor hierarchy are one, that each level of GATA factors directly activates the succeeding level and two, that within each level the factors are capable of positive auto-regulation and positive feedback on each other (Maduro and Rothman, 2002; McGhee, 2013). This is what activates the expression of ELT-7 and then ELT-2 early in the E lineage and how expression is maintained throughout the life of the worm. However, expression of *elt-2* mRNA was observed to be very stable in both the wild-type populations and when the gene was present presumably in greatly increased copy numbers. This implies that there must be a negative/repressive input on *elt-2* transcription to prevent runaway accumulation of this protein. There are a few possible mechanisms by which this could occur.

First, repressive input on *elt-2* transcription could be mediated by one or more downstream transcription factor targets in a negative feedback loop. Evidence for this mechanism working in *C. elegans* has been previously reported. The stress response pathway is mediated by the bZIP transcription factor SKN-1, expression of which is repressed by its own activation target WDR-23 to balance stress resistance and growth/development (Leung et al., 2013). As further evidence of the plausibility of this mechanism, the male specific MAB-3 transcription factor has already been reported as a negative regulator of *elt-2* transcription in males (Inoue and Nishida, 2010). Therefore, it is realistic to hypothesize there are *trans*-acting factors in the hermaphrodite that function in this manner as well.

An alternative negative feedback mode of action that functions in *C. elegans* to regulate transcription factor levels is miRNAs. In the heterochronic pathway, developmentally regulated miRNAs are produced to regulate the temporal changes in mRNA abundance of critical transcription factors in order to mediate development (Resnick et al., 2010). It is possible that miRNAs negatively feedback to regulate *elt-2* mRNA levels and thus protein levels to ensure stable expression of this gene.

The third mechanism of action could be by directly regulating ELT-2 protein levels, either by degradation or repression of translation by an mRNA binding protein. This hypothesis is supported by the abundance of *elt-2* mRNA relative to ELT-2 protein in the worms with extra copies of the *elt-2* gene. Despite a five-fold increase in mRNA, ELT-2 protein levels in these worms were only slightly up-regulated compared to wild-type, suggesting that there is some unidentified system preventing accumulation of the protein. Ultimately, there could be more than one means in place to maintain stable expression of ELT-2 in the intestine. Future work should focus on elucidating which of these proposed mechanisms may actually be at work.

It is unclear if *elt-7* expression is under similar control. mRNA expression of this gene was quite low and variable in most of the *elt-7(+)* populations, hovering just on the edge of detection at the level of sequencing depth used. These findings are consistent with previous SAGE analysis of dissected intestines that reported *elt-2* vs *elt-7* transcripts in an abundance ratio of 25:1 (McGhee et al., 2007). Expression of *elt-7* could be down-regulated in these populations due to starvation, yet this is unlikely as *elt-7* transcripts were detected in SAGE analysis of starved L1 worms (McGhee et al., 2009). Furthermore, an integrated *elt-7* reporter transgene indicated that this gene was still expressed in starved L1 larvae (data not shown), suggesting that these transcripts were present at too low a concentration to be detected in the samples.

Functionally, low levels of transcripts do not necessarily mean that the ELT-7 protein is not present. One reason for this could be the size of the *elt-7* mRNA transcripts themselves, at 351 bp and 597 bp each, it is possible that they are rapidly degraded and do not accumulate significantly. mRNAs for transcription factors are known for being unstable and rapidly turned over (Schwanhäusser et al., 2011). Another cause of low reads for the *elt-7* gene that is associated with transcript length could be inherent bias of RNAseq itself. RNAseq has been shown to have substantial under representation for short transcripts (Rehrauer et al., 2013). Furthermore, despite accounting for transcript length, calling of differentially expressed genes still shows some bias based on this feature (Oshlack and Wakefield, 2009). Based on this supporting evidence it appears that *elt-7* is expressed in the L1 larvae sequenced here, albeit at a much lower amount than *elt-2*. Interestingly, *elt-7* mRNA levels were highest in the *elt-2* mutants, which implies

that some aspect of ELT-2 function may result in repression of *elt-7* transcription despite evidence that ELT-2 activates this gene. This raises the possibility that downstream targets of ELT-2 negatively feed-back on *elt-7* expression. In the absence of ELT-2, this inhibition is released and ELT-7 can positively auto-activate itself.

4.6 Conclusions

The work presented here illustrates the distinct functions of the ELT-2 and ELT-7 intestinal GATA transcription factors. All of the evidences indicates that ELT-2 is clearly the major regulator of transcription in the intestine. The action of ELT-2 in the absence of ELT-7 is entirely sufficient to mediate the regulation of most intestinal genes and indeed produces a fully functional, differentiated organ. ELT-7 does appear to have an important role during embryogenesis and is partially redundant with ELT-2.

Future directions should combine quantitative RT-PCR with promoter analysis and bandshift assays to confirm how these factors regulate different classes of intestinal genes. Furthermore, it will be of high interest to identify if intestinal genes are directly repressed by ELT-2 and ELT-7 to ascertain if these factors have dual regulation capabilities. Prospective studies should test the model that ELT-2 and ELT-7 bind to any of the TGATAA sites in the promoters of redundantly activated class as well as if ELT-2 does outcompete ELT-7 for binding to sites in the exclusively regulated class. In addition, Systematic Evolution of Ligands by EXponential enrichment (SELEX) experiments (Djordjevic, 2007; Tuerk and Gold, 1990) will be performed to identify the sequences to which ELT-2 and ELT-7 bind, including the different preferences that each factor has for different sequences. The SELEX results can be combined with data from an ongoing collaborative ChIPseq project to identify ELT-2 binding sites in the *C. elegans* genome. Another key experiment would be to create an ELT-7 specific antibody to perform ChIPseq for this GATA factor as well. The data from all of these proposed experiments in conjunction with the data presented here should provide a very clear understanding of how these two factors regulate intestinal gene expression in *C. elegans*.

Chapter V: Overall Thesis Conclusions

5.1 Three Independent Studies with Connecting Threads

The work here describes in detail three projects that investigated seemingly diverse aspects of nematode biology. There were three key findings of the study of body size regulation by the TGF-β Sma/Mab signaling pathway. First, this signaling pathway regulates pharynx size, second pharynx signaling contributes partially to body size regulation in a non cell-autonomous manner and third, this pathway may be involved in coordinating growth of different organs in the worm. The study of intestinal gene spatial patterning found that LIN-14 is regulating the *pho-1* transcriptional expression pattern via an indirect mechanism. This implies that LIN-14 may be regulating other intestinal transcription factors that ultimately mediate *pho-1* spatial patterning. It is possible that these unknown *trans*-acting factors are regulated by ELT-2 and/or ELT-7 given that about 25% of all intestinal transcription factors are differentially expressed in *elt-2/elt-7* mutant strains. The RNAseq study of intestinal gene expression provided further evidence for ELT-2's critical role in this organ.

TGF-β signaling is a key regulator of various liver diseases (Zhang et al., 2014) and so a better understanding of how GATA factors regulate gene expression in endodermal organs such as the liver will hopefully contribute to better treatments for these diseases. Interestingly, the ELT-2 mammalian homolog GATA-4 has been shown to work synergistically with the TGF-β downstream Smad effectors. GATA-4 was found to complex with Smad2, Smad3, and Smad4 to directly activate expression of the intestinal epithelial genes intestinal alkaline phosphatase (IAP) and intestinal fatty acid binding protein (IFABP) (Belaguli et al., 2007). Future work could aim to see if this is a conserved feature of intestinal gene regulation that occurs in *C. elegans* as well. Hints that this may be the case come from the hypodermis where the GATA factors ELT-1 and ELT-3 have been reported to directly regulate expression of a cuticle collagen gene that is a target of the TGF-β Sma/Mab signaling pathway (Yin et al., 2015). Furthermore, *C. elegans* innate immunity genes respond to TGF-β Sma/Mab signaling (Julien-Gau et al., 2014; Roberts et al., 2010), suggesting that ELT-2 may indeed act at least in combination with the downstream Smads to regulate these intestinal genes.

My work highlights the importance of understanding basic concepts in biology in order to better understand and treat human disease. A number of genes that were differentially expressed in the RNAseq data set such as *asah-1* and *F08A8.2* are orthologs and homologs of human intestinal genes that cause disease when mutated. Additionally, the GATA factor specification regulatory transcriptional hierarchy that controls expression of these genes is conserved in many organisms including humans. Similarly, the TGF-β Sma/Mab signaling pathway has multiple essential roles in human biology. It is critical to amass knowledge of how these proteins function in simple organisms that are cheap and quick to work with in order to better understand and treat aspects of human disease.

Bibliography

Alper, S., McBride, S. J., Lackford, B., Freedman, J. H. and Schwartz, D. A. (2007). Specificity and complexity of the Caenorhabditis elegans innate immune response. *Mol. Cell. Biol.* **27**, 5544–53.

Ambros, V. (2000). Control of developmental timing in Caenorhabditis elegans. *Curr. Opin. Genet. Dev.* **10**, 428–433.

Ambros, V. and Horvitz, H. R. (1984). Heterochronic mutants of the nematode Caenorhabditis elegans. *Science* **226**, 409–416.

Ambros, V. and Horvitz, H. R. (1987). The lin-14 locus of Caenorhabditis elegans controls the time of expression of specific postembryonic developmental events. *Genes Dev.* **1**, 398–414.

Anders, S., Pyl, P. T. and Huber, W. (2014). HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–9.

Anokye-Danso, F., Anyanful, A., Sakube, Y. and Kagawa, H. (2008). Transcription Factors GATA/ELT-2 and Forkhead/HNF-3/PHA-4 Regulate the Tropomyosin Gene Expression in the Pharynx and Intestine of Caenorhabditis elegans. *J. Mol. Biol.* **379**, 201–211.

Antebi, A. (2006). Nuclear hormone receptors in C. elegans. *WormBook* 1–13.

Aronson, B. E., Stapleton, K. A. and Krasinski, S. D. (2014). Role of GATA factors in development, differentiation, and homeostasis of the small intestinal epithelium. *Am. J. Physiol. Gastrointest. Liver Physiol.* **306**, G474–90.

Ashrafi, K., Chang, F. Y., Watts, J. L., Fraser, A. G., Kamath, R. S., Ahringer, J. and Ruvkun, G. (2003). Genome-wide RNAi analysis of Caenorhabditis elegans fat regulatory genes. *Nature* **421**, 268–72.

Axäng, C., Rauthan, M., Hall, D. H. and Pilon, M. (2007). The twisted pharynx phenotype in C. elegans. *BMC Dev. Biol.* **7**, 61.

Bansal, A., Kwon, E.-S., Conte, D., Liu, H., Gilchrist, M. J., MacNeil, L. T. and Tissenbaum, H. A. (2014). Transcriptional regulation of Caenorhabditis elegans FOXO/DAF-16 modulates lifespan. *Longev. Heal.* **3**, 5.

Bates, D. L., Chen, Y., Kim, G., Guo, L. and Chen, L. (2008). Crystal Structures of Multiple GATA Zinc Fingers Bound to DNA Reveal New Insights into DNA Recognition and Self-Association by GATA. *J. Mol. Biol.* **381**, 1292–1306.

Baugh, L. R., Hill, A. a, Slonim, D. K., Brown, E. L. and Hunter, C. P. (2003). Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. *Development* **130**, 889–900.

Baugh, L. R., Demodena, J. and Sternberg, P. W. (2009). RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science* **324**, 92–94.

Beh, C. T., Ferrari, D. C., Chung, M. A. and McGhee, J. D. (1991). An acid phosphatase as a biochemical marker for intestinal development in the nematode Caenorhabditis elegans. *Dev. Biol.* **147**, 133–43.

Belaguli, N. S., Zhang, M., Rigi, M., Aftab, M. and Berger, D. H. (2007). Cooperation between GATA4 and TGF-beta signaling regulates intestinal epithelial gene expression. *Am. J. Physiol. Gastrointest. Liver Physiol.* **292**, G1520–G1533.

Benner, J., Daniel, H. and Spanier, B. (2011). A glutathione peroxidase, intracellular peptidases and the TOR complexes regulate peptide transporter PEPT-1 in C. elegans. *PLoS One* **6**, e25624.

Bertrand, V. and Hobert, O. (2010). Lineage programming: navigating through transient regulatory states via binary decisions. *Curr. Opin. Genet. Dev.* **20**, 362–8.

Blazie, S. M., Babb, C., Wilky, H., Rawls, A., Park, J. G. and Mangone, M. (2015). Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. *BMC Biol.* **13**,.

Borgonie, G., Claeys, M. and Waele, D. De (1995). Ultrastructure of the intestine of the bacteriophagous nematodes caenorhabditis-elegans, panagrolaimus superbus and acrobeloides maximus (nematoda, rhabditida). **18**, 123–133.

Bowerman, B., Draper, B. W., Mello, C. C. and Priess, J. R. (1993). The maternal gene skn-1 encodes a protein that is distributed unequally in early C. elegans embryos. *Cell* **74**, 443–52.

Boxem, M. and van den Heuvel, S. (2001). lin-35 Rb and cki-1 Cip/Kip cooperate in developmental regulation of G1 progression in C. elegans. *Development* **128**, 4349–4359.

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

Britton, C., McKerrow, J. H. and Johnstone, I. L. (1998). Regulation of the Caenorhabditis elegans gut cysteine protease gene cpr-1: requirement for GATA motifs. *J. Mol. Biol.* **283**, 15–27.

Bruinsma, J. J., Jirakulaporn, T., Muslin, A. J. and Kornfeld, K. (2002). Zinc ions and cation diffusion facilitator proteins regulate Ras-mediated signaling. *Dev. Cell* **2**, 567–78.

Burmeister, C., Lüersen, K., Heinick, A., Hussein, A., Domagalski, M., Walter, R. D. and Liebau, E. (2008). Oxidative stress in Caenorhabditis elegans: protective effects of the Omega class glutathione transferase (GSTO-1). *FASEB J.* **22**, 343–354.

Carrasco, M., Delgado, I., Soria, B., Martín, F. and Rojas, A. (2012). GATA4 and GATA6 control mouse pancreas organogenesis. *J. Clin. Invest.* **122**, 3504–3515.

Clarke, N. D. and Berg, J. M. (1998). Zinc fingers in Caenorhabditis elegans: finding families and probing pathways. *Science* **282**, 2018–22.

Collins, T. J. (2007). ImageJ for microscopy. *Biotechniques* **43**, 25–30.

Consortium, C. elegans S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* **282**, 2012–8.

Costa, R. H., Kalinichenko, V. V and Lim, L. (2001). Transcription factors in mouse lung development and function. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **280**, L823–38.

Cypser, J. R., Kitzenberg, D. and Park, S.-K. (2013). Dietary restriction in C. elegans: Recent advances. *Exp. Gerontol.* **48**, 1014–1017.

Desnoyers, S., Blanchard, P.-G., St-Laurent, J.-F., Gagnon, S. N., Baillie, D. L. and Luu-The, V. (2007). Caenorhabditis elegans LET-767 is able to metabolize androgens and estrogens and likely shares common ancestor with human types 3 and 12 17 -hydroxysteroid dehydrogenases. *J. Endocrinol.* **195**, 271–279.

Djordjevic, M. (2007). SELEX experiments: New prospects, applications and data analysis in inferring regulatory pathways. *Biomol. Eng.* **24**, 179–189.

Driever, W. and Nüsslein-Volhard, C. (1988). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell* **54**, 95–104.

Edgar, L. G. and McGhee, J. D. (1986). Embryonic expression of a gut-specific esterase in Caenorhabditis elegans. *Dev. Biol.* **114**, 109–18.

Edgar, L. G., Carr, S., Wang, H. and Wood, W. B. (2001). Zygotic expression of the caudal homolog pal-1 is required for posterior patterning in Caenorhabditis elegans embryogenesis. *Dev. Biol.* **229**, 71–88.

Elliott, S. L., Sturgeon, C. R., Travers, D. M. and Montgomery, M. C. (2011). Mode of bacterial pathogenesis determines phenotype in elt-2 and elt-7 RNAi Caenorhabditis elegans. *Dev. Comp. Immunol.* **35**, 521–4.

Entchev, E. V., Schwudke, D., Zagoriy, V., Matyash, V., Bogdanova, A., Habermann, B., Zhu, L., Shevchenko, A. and Kurzchalia, T. V. (2008). LET-767 Is Required for the Production of Branched Chain and Long Chain Fatty Acids in Caenorhabditis elegans. *J. Biol. Chem.* **283**, 17550–17560.

Flemming, A. J., Shen, Z. Z., Cunha, A., Emmons, S. W. and Leroi, A. M. (2000). Somatic polyploidization and cellular proliferation drive body size evolution in nematodes. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5285–90.

Freedman, J. H., Slice, L. W., Dixon, D., Fire, A. and Rubin, C. S. (1993). The novel metallothionein genes of Caenorhabditis elegans. Structural organization and inducible, cell-specific expression. *J. Biol. Chem.* **268**, 2554–64.

Frise, E., Knoblich, J. A., Younger-Shepherd, S., Jan, L. Y. and Jan, Y. N. (1996). The Drosophila Numb protein inhibits signaling of the Notch receptor during cell-cell interaction in sensory organ lineage. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11925–32.

Fukushige, T., Hawkins, M. G. and McGhee, J. D. (1998). The GATA-factor elt-2 is essential for formation of the Caenorhabditis elegans intestine. *Dev. Biol.* **198**, 286–302.

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. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11883–8.

Fukushige, T., Goszczynski, B., Tian, H. and McGhee, J. D. (2003). The evolutionary duplication and probable demise of an endodermal GATA factor in Caenorhabditis elegans. *Genetics* **165**, 575–88.

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. *Dev. Biol.* **279**, 446–461.

Gaudet, J. and Mango, S. E. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. *Science* **295**, 821–5.

Gaudet, J., VanderElst, I. and Spence, A. M. (1996). Post-transcriptional regulation of sex determination in Caenorhabditis elegans: widespread expression of the sex-determining gene fem-1 in both sexes. *Mol. Biol. Cell* **7**, 1107–21.

Gaudet, J., Muttumu, S., Horner, M. and Mango, S. E. (2004). Whole-genome analysis of temporal gene expression during foregut development. *PLoS Biol.* **2**, e352.

Gerhart, J. (1999). 1998 Warkany lecture: signaling pathways in development. *Teratology* **60**, 226–39.

Gho, M., Bellaïche, Y. and Schweisguth, F. (1999). Revisiting the Drosophila microchaete lineage: a novel intrinsically asymmetric cell division generates a glial cell. *Development* **126**, 3573–3584.

Gilleard, J. S. and McGhee, J. D. (2001). Activation of hypodermal differentiation in the Caenorhabditis elegans embryo by GATA transcription factors ELT-1 and ELT-3. *Mol. Cell. Biol.* **21**, 2533–44.

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. *Mol. Cell. Biol.* **17**, 2301–11.

Gillis, W. Q., Bowerman, B. A. and Schneider, S. Q. (2008). The evolution of protostome GATA factors: molecular phylogenetics, synteny, and intron/exon structure reveal orthologous relationships. *BMC Evol. Biol.* **8**, 112.

Göbel, V., Barrett, P. L., Hall, D. H. and Fleming, J. T. (2004). Lumen morphogenesis in C. elegans requires the membrane-cytoskeleton linker erm-1. *Dev. Cell* **6**, 865–73.

Goldstein, B. (1992). Induction of gut in Caenorhabditis elegans embryos. *Nature* **357**, 255–7.

Grewal, S. S. (2009). Insulin/TOR signaling in growth and homeostasis: a view from the fly world. *Int. J. Biochem. Cell Biol.* **41**, 1006–10.

Gumienny, T. L. and Savage-Dunn, C. (2013). TGF-β signaling in C. elegans. *WormBook* 1–34.

Gumienny, T. L., MacNeil, L. T., Wang, H., de Bono, M., Wrana, J. L. and Padgett, R. W. (2007). Glypican LON-2 is a conserved negative regulator of BMP-like signaling in Caenorhabditis elegans. *Curr. Biol.* **17**, 159–64.

Gumienny, T. L., Macneil, L., Zimmerman, C. M., Wang, H., Chin, L., Wrana, J. L. and Padgett, R. W. (2010). Caenorhabditis elegans SMA-10/LRIG is a conserved transmembrane protein that enhances bone morphogenetic protein signaling. *PLoS Genet.* **6**, e1000963.

Guo, M., Jan, L. Y. and Jan, Y. N. (1996). Control of daughter cell fates during asymmetric division: interaction of Numb and Notch. *Neuron* **17**, 27–41.

Haag, A., Gutierrez, P., Bühler, A., Walser, M., Yang, Q., Langouët, M., Kradolfer, D., Fröhli, E., Herrmann, C. J., Hajnal, A., et al. (2014). An in vivo EGF receptor localization screen in C. elegans Identifies the Ezrin homolog ERM-1 as a temporal regulator of signaling. *PLoS Genet.* **10**, e1004341.

Hartenstein, V. and Posakony, J. W. (1990). A dual function of the Notch gene in Drosophila sensillum development. *Dev. Biol.* **142**, 13–30.

Hawkins, M. G. and McGhee, J. D. (1995). elt-2, a second GATA factor from the nematode Caenorhabditis elegans. *J. Biol. Chem.* **270**, 14666–71.

Head, B. and Aballay, A. (2014). Recovery from an acute infection in C. elegans requires the GATA transcription factor ELT-2. *PLoS Genet.* **10**, e1004609.

Hedgecock, E. M. and White, J. G. (1985). Polyploid tissues in the nematode Caenorhabditis elegans. *Dev. Biol.* **107**, 128–133.

Hermann, G. J., Leung, B. and Priess, J. R. (2000). Left-right asymmetry in C. elegans intestine organogenesis involves a LIN-12/Notch signaling pathway. *Development* **127**, 3429–3440.

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

Horvitz, H. R. and Herskowitz, I. (1992). Mechanisms of asymmetric cell division: two Bs or not two Bs, that is the question. *Cell* **68**, 237–55.

Hristova, M., Birse, D., Hong, Y. and Ambros, V. (2005). The Caenorhabditis elegans heterochronic regulator LIN-14 is a novel transcription factor that controls the developmental timing of transcription from the insulin/insulin-like growth factor gene ins-33 by direct DNA binding. *Mol. Cell. Biol.* **25**, 11059–72.

Hsieh, J. and Fire, A. (2000). Recognition and silencing of repeated DNA. *Annu. Rev. Genet.* **34**, 187–204.

Huang, S., Shetty, P., Robertson, S. M. and Lin, R. (2007). Binary cell fate specification during C. elegans embryogenesis driven by reiterated reciprocal asymmetry of TCF POP-1 and its coactivator beta-catenin SYS-1. *Development* **134**, 2685–2695.

Hüsken, K., Wiesenfahrt, T., Abraham, C., Windoffer, R., Bossinger, O. and Leube, R. E. (2008). Maintenance of the intestinal tube in Caenorhabditis elegans: the role of the intermediate filament protein IFC-2. *Differentiation.* **76**, 881–96.

Hyun, S. (2013). Body size regulation and insulin-like growth factor signaling. *Cell. Mol. Life Sci.* **70**, 2351–65.

Inoue, H. and Nishida, E. (2010). The DM domain transcription factor MAB-3 regulates male hypersensitivity to oxidative stress in Caenorhabditis elegans. *Mol. Cell. Biol.* **30**, 3453–9.

Inoue, T. and Thomas, J. H. (2000). Targets of TGF-beta signaling in Caenorhabditis elegans dauer formation. *Dev. Biol.* **217**, 192–204.

Jafari, G., Burghoorn, J., Kawano, T., Mathew, M., Mörck, C., Axäng, C., Ailion, M., Thomas, J. H., Culotti, J. G., Swoboda, P., et al. (2010). Genetics of extracellular matrix remodeling during organ growth using the Caenorhabditis elegans pharynx model. *Genetics* **186**, 969–82.

Johnstone, O. and Lasko, P. (2001). Translational regulation and RNA localization in Drosophila oocytes and embryos. *Annu. Rev. Genet.* **35**, 365–406.

Johnstone, I. L., Shafi, Y. and Barry, J. D. (1992). Molecular analysis of mutations in the Caenorhabditis elegans collagen gene dpy-7. *EMBO J.* **11**, 3857–63.

Julien-Gau, I., Schmidt, M. and Kurz, C. L. (2014). f57f4.4p::gfp as a fluorescent reporter for analysis of the C. elegans response to bacterial infection. *Dev. Comp. Immunol.* **42**, 132–7.

Kaletta, T., Schnabel, H. and Schnabel, R. (1997). Binary specification of the embryonic lineage in Caenorhabditis elegans. *Nature* **390**, 294–8.

Kamei, C. N., Kempf, H., Yelin, R., Daoud, G., James, R. G., Lassar, A. B., Tabin, C. J. and Schultheiss, T. M. (2011). Promotion of avian endothelial cell differentiation by GATA transcription factors. *Dev. Biol.* **353**, 29–37.

Kennedy, B. P., Aamodt, E. J., Allen, F. L., Chung, M. A., Heschl, M. F. and McGhee, J. D. (1993). The gut esterase gene (ges-1) from the nematodes Caenorhabditis elegans and Caenorhabditis briggsae. *J. Mol. Biol.* **229**, 890–908.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996–1006.

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–7.

Kenyon, C. (2010). A pathway that links reproductive status to lifespan in Caenorhabditis elegans. *Ann. N. Y. Acad. Sci.* **1204**, 156–62.

Kerry, S., TeKippe, M., Gaddis, N. C. and Aballay, A. (2006). GATA transcription factor required for immunity to bacterial and fungal pathogens. *PLoS One* **1**, e77.

Kimble, J. and Sharrock, W. J. (1983). Tissue-specific synthesis of yolk proteins in Caenorhabditis elegans. *Dev. Biol.* **96**, 189–96.

Kimura, T., Takanami, T., Sakashita, T., Wada, S., Kobayashi, Y. and Higashitani, A. (2012). Innate immune genes including a mucin-like gene, mul-1, induced by ionizing radiation in Caenorhabditis elegans. *Radiat. Res.* **178**, 313–20.

Kniazeva, M., Crawford, Q. T., Seiber, M., Wang, C.-Y. and Han, M. (2004). Monomethyl branched-chain fatty acids play an essential role in Caenorhabditis elegans development. *PLoS Biol.* **2**, E257.

Kohara, Y. (2001). [Systematic analysis of gene expression of the C. elegans genome]. *Tanpakushitsu Kakusan Koso.* **46**, 2425–31.

Köppen, M., Simske, J. S., Sims, P. A., Firestein, B. L., Hall, D. H., Radice, A. D., Rongo, C. and Hardin, J. D. (2001). Cooperative regulation of AJM-1 controls junctional integrity in Caenorhabditis elegans epithelia. *Nat. Cell Biol.* **3**, 983–91.

Kormish, J. D., Gaudet, J. and McGhee, J. D. (2010). Development of the C. elegans digestive tract. *Curr. Opin. Genet. Dev.* **20**, 346–354.

Korswagen, H. C. (2002). Canonical and non-canonical Wnt signaling pathways in Caenorhabditis elegans: variations on a common signaling theme. *Bioessays* **24**, 801–10.

Kramer, J. M., French, R. P., Park, E. C. and Johnson, J. J. (1990). The Caenorhabditis elegans rol-6 gene, which interacts with the sqt-1 collagen gene to determine organismal morphology, encodes a collagen. *Mol. Cell. Biol.* **10**, 2081–9.

Kuervers, L. M., Jones, C. L., O'Neil, N. J. and Baillie, D. L. (2003). The sterol modifying enzyme LET-767 is essential for growth, reproduction and development in Caenorhabditis elegans. *Mol. Genet. Genomics* **270**, 121–131.

Kusch, M. and Edgar, R. S. (1986). Genetic studies of unusual loci that affect body shape of the nematode Caenorhabditis elegans and may code for cuticle structural proteins. *Genetics* **113**, 621–39.

Lee, S.-H., Wong, R.-R., Chin, C.-Y., Lim, T.-Y., Eng, S.-A., Kong, C., Ijap, N. A., Lau, M.-S., Lim, M.-P., Gan, Y.-H., et al. (2013). Burkholderia pseudomallei suppresses Caenorhabditis elegans immunity by specific degradation of a GATA transcription factor. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15067–72.

Lehmann, R. and Nüsslein-Volhard, C. (1991). The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. *Development* **112**, 679–91.

Lengyel, J. A. and Iwaki, D. D. (2002). It takes guts: the Drosophila hindgut as a model system for organogenesis. *Dev. Biol.* **243**, 1–19.

Leung, B., Hermann, G. J. and Priess, J. R. (1999). Organogenesis of the Caenorhabditis elegans intestine. *Dev. Biol.* **216**, 114–134.

Leung, C. K., Wang, Y., Deonarine, A., Tang, L., Prasse, S. and Choe, K. P. (2013). A negative-feedback loop between the detoxification/antioxidant response factor SKN-1 and its repressor WDR-23 matches organism needs with environmental conditions. *Mol. Cell. Biol.* **33**, 3524–37.

Levy, A. D., Yang, J. and Kramer, J. M. (1993). Molecular and genetic analyses of the Caenorhabditis elegans dpy-2 and dpy-10 collagen genes: a variety of molecular alterations affect organismal morphology. *Mol. Biol. Cell* **4**, 803–17.

Li, J. and Greenwald, I. (2010). LIN-14 inhibition of LIN-12 contributes to precision and timing of C. elegans vulval fate patterning. *Curr. Biol.* **20**, 1875–9.

Lin, R., Hill, R. J. and Priess, J. R. (1998). POP-1 and anterior-posterior fate decisions in C. elegans embryos. *Cell* **92**, 229–239.

Lo, M., Odom, R., Shi, Y. and Lin, R. (2004). Phosphorylation by the -Catenin / MAPK Complex Promotes 14-3-3-Mediated Nuclear Export of TCF / POP-1 in Signal-Responsive Cells in C . elegans. **117**, 95–106.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *bioRxiv*.

Lowry, J. A. and Atchley, W. R. (2000). Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *J. Mol. Evol.* **50**, 103–15.

Maduro, M. F. and Rothman, J. H. (2002). Making Worm Guts: The Gene Regulatory Network of the Caenorhabditis elegans Endoderm. *Dev. Biol.* **246**, 68–85.

Maduro, M. F., Meneghini, M. D., Bowerman, B., Broitman-Maduro, G. and Rothman, J. H. (2001). Restriction of mesendoderm to a single blastomere by the combined action of SKN-1 and a GSK-3beta homolog is mediated by MED-1 and -2 in C. elegans. *Mol. Cell* **7**, 475–85.

Maduzia, L. L., Yu, E. and Zhang, Y. (2011). Caenorhabditis elegans galectins LEC-6 and LEC-10 interact with similar glycoconjugates in the intestine. *J. Biol. Chem.* **286**, 4371–81.

Maeda, I., Kohara, Y., Yamamoto, M. and Sugimoto, A. (2001). Large-scale analysis of gene function in Caenorhabditis elegans by high-throughput RNAi. *Curr. Biol.* **11**, 171–6.

Mallo, G. V, Kurz, C. L., Couillault, C., Pujol, N., Granjeaud, S., Kohara, Y. and Ewbank, J. J. (2002). Inducible antibacterial defense system in C. elegans. *Curr. Biol.* **12**, 1209–14.

Marshall, S. D. and McGhee, J. D. (2001). Coordination of ges-1 expression between the Caenorhabditis pharynx and intestine. *Dev. Biol.* **239**, 350–363.

Maxwell, C. S., Antoshechkin, I., Kurhanewicz, N., Belsky, J. a. and Baugh, L. R. (2012). Nutritional control of mRNA isoform expression during developmental arrest and recovery in C. elegans. *Genome Res.* **22**, 1920–1929.

McGhee, J. D. (2007). The C. elegans intestine. *WormBook* 1–36.

McGhee, J. D. (2013). The Caenorhabditis elegans intestine. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 347–367.

McGhee, J. D., Sleumer, M. C., Bilenky, M., Wong, K., McKay, S. J., Goszczynski, B., Tian, H., Krich, N. D., Khattra, J., Holt, R. a., et al. (2007). The ELT-2 GATA-factor and the global regulation of transcription in the C. elegans intestine. *Dev. Biol.* **302**, 627–645.

McGhee, J. D., Fukushige, T., Krause, M. W., Minnema, S. E., Goszczynski, B., Gaudet, J., Kohara, Y., Bossinger, O., Zhao, Y., Khattra, J., et al. (2009). ELT-2 is the predominant transcription factor controlling differentiation and function of the C. elegans intestine, from embryo to adult. *Dev. Biol.* **327**, 551–565.

McKay, R. M., McKay, J. P., Avery, L. and Graff, J. M. (2003). C elegans: a model for exploring the genetics of fat storage. *Dev. Cell* **4**, 131–42.

Mello, C. C., Kramer, J. M., Stinchcomb, D. and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–70.

Micallef, L. and Rodgers, P. (2014). eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* **9**, e101717.

Mohun, T. and Sparrow, D. (1997). Early steps in vertebrate cardiogenesis. *Curr. Opin. Genet. Dev.* **7**, 628–33.

Moilanen, L. H., Fukushige, T. and Freedman, J. H. (1999). Regulation of metallothionein gene transcription. Identification of upstream regulatory elements and transcription factors responsible for cell-specific expression of the

metallothionein genes from Caenorhabditis elegans. *J. Biol. Chem.* **274**, 29655–29665.

Mörck, C. and Pilon, M. (2006). C. elegans feeding defective mutants have shorter body lengths and increased autophagy. *BMC Dev. Biol.* **6**, 39.

Morita, K., Chow, K. L. and Ueno, N. (1999). Regulation of body length and male tail ray pattern formation of Caenorhabditis elegans by a member of TGF-beta family. *Development* **126**, 1337–47.

Murakami, R., Okumura, T. and Uchiyama, H. (2005). GATA factors as key regulatory molecules in the development of Drosophila endoderm. *Dev. Growth Differ.* **47**, 581–9.

Murray, J. I., Bao, Z., Boyle, T. J., Boeck, M. E., Mericle, B. L., Nicholas, T. J., Zhao, Z., Sandel, M. J. and Waterston, R. H. (2008). Automated analysis of embryonic gene expression with cellular resolution in C. elegans. *Nat. Methods* **5**, 703–9.

Muskavitch, M. A. (1994). Delta-notch signaling and Drosophila cell fate choice. *Dev. Biol.* **166**, 415–30.

Nahmad, M. and Lander, A. D. (2011). Spatiotemporal mechanisms of morphogen gradient interpretation. *Curr. Opin. Genet. Dev.* **21**, 726–31.

Nelson, A. C. and Wardle, F. C. (2013). Conserved non-coding elements and cis regulation: actions speak louder than words. *Development* **140**, 1385–95.

Neves, A. and Priess, J. R. (2005). The REF-1 family of bHLH transcription factors pattern C. elegans embryos through Notch-dependent and Notch-independent pathways. *Dev. Cell* **8**, 867–879.

Neves, A., English, K. and Priess, J. R. (2007). Notch-GATA synergy promotes endoderm-specific expression of ref-1 in C. elegans. *Development* **134**, 4459–4468.

NEXTDB The Nematode Expression Pattern Database

Nigon, V. (1949). Les modalites de la reproduction et le determinisme de sexe chez quelque Nematode libres. *C. R. Hebd. Seances Acad. Sci.* **228**, 1161.

Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795–801.

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

Olsen, P. H. and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**, 671–80.

Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J. and Gronenborn, A. M. (1993). NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* **261**, 438–46.

Oshlack, A. and Wakefield, M. J. (2009). Transcript length bias in RNA-seq data confounds systems biology. *Biol. Direct* **4**, 14.

Owraghi, M., Broitman-Maduro, G., Luu, T., Roberson, H. and Maduro, M. F. (2010). Roles of the Wnt effector POP-1/TCF in the C. elegans endomesoderm specification gene network. *Dev. Biol.* **340**, 209–21.

Park, Y. S. and Kramer, J. M. (1994). The C. elegans sqt-1 and rol-6 collagen genes are coordinately expressed during development, but not at all stages that display mutant phenotypes. *Dev. Biol.* **163**, 112–24.

Park, M. and Krause, M. W. (1999). Regulation of postembryonic G(1) cell cycle progression in Caenorhabditis elegans by a cyclin D/CDK-like complex. *Development* **126**, 4849–4860.

Pauli, F., Liu, Y., Kim, Y. a, Chen, P.-J. and Kim, S. K. (2006). Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in C. elegans. *Development* **133**, 287–295.

Phillips, B. T., Kidd, A. R., King, R., Hardin, J. and Kimble, J. (2007). Reciprocal asymmetry of SYS-1/beta-catenin and POP-1/TCF controls asymmetric divisions in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3231–3236.

Posakony, J. W. (1994). Nature versus nurture: asymmetric cell divisions in Drosophila bristle development. *Cell* **76**, 415–8.

Pukkila-Worley, R. and Ausubel, F. M. (2012). Immune defense mechanisms in the Caenorhabditis elegans intestinal epithelium. *Curr. Opin. Immunol.* **24**, 3–9.

Raharjo, W. H., Ghai, V., Dineen, A., Bastiani, M. and Gaudet, J. (2011). Cell architecture: surrounding muscle cells shape gland cell morphology in the Caenorhabditis elegans pharynx. *Genetics* **189**, 885–97.

Raj, A., Rifkin, S. A., Andersen, E. and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–8.

Ramakrishnan, K., Ray, P. and Okkema, P. G. (2014). CEH-28 activates dbl-1 expression and TGF-β signaling in the C. elegans M4 neuron. *Dev. Biol.* **390**, 149–59.

Reddy, B. V. V. G. and Irvine, K. D. (2008). The Fat and Warts signaling pathways: new insights into their regulation, mechanism and conservation. *Development* **135**, 2827–38.

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 regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol.* **6**, R110.

Rehrauer, H., Opitz, L., Tan, G., Sieverling, L. and Schlapbach, R. (2013). Blind spots of quantitative RNA-seq: the limits for assessing abundance, differential expression, and isoform switching. *BMC Bioinformatics* **14**, 370.

Reinhart, B. J. and Ruvkun, G. (2001). Isoform-specific mutations in the Caenorhabditis elegans heterochronic gene lin-14 affect stage-specific patterning. *Genetics* **157**, 199–209.

Ren, H. and Zhang, H. (2010). Wnt signaling controls temporal identities of seam cells in Caenorhabditis elegans. *Dev. Biol.* **345**, 144–155.

Resnick, T. D., McCulloch, K. a. and Rougvie, A. E. (2010). miRNAs give worms the time of their lives: Small RNAs and temporal control in Caenorhabditis elegans. *Dev. Dyn.* **239**, 1477–1489.

Rhyu, M. S. and Knoblich, J. A. (1995). Spindle orientation and asymmetric cell fate. *Cell* **82**, 523–6.

Rhyu, M. S., Jan, L. Y. and Jan, Y. N. (1994). Asymmetric distribution of numb protein during division of the sensory organ precursor cell confers distinct fates to daughter cells. *Cell* **76**, 477–91.

Riddle, M. R., Weintraub, A., Nguyen, K. C. Q., Hall, D. H. and Rothman, J. H. (2013). Transdifferentiation and remodeling of post-embryonic C. elegans cells by a single transcription factor. *Development* **140**, 4844–9.

Roberts, A. F., Gumienny, T. L., Gleason, R. J., Wang, H. and Padgett, R. W. (2010). Regulation of genes affecting body size and innate immunity by the DBL-1/BMP-like pathway in Caenorhabditis elegans. *BMC Dev. Biol.* **10**, 61.

Rocheleau, C. E., Yasuda, J., Shin, T. H., Lin, R., Sawa, H., Okano, H., Priess, J. R., Davis, R. J. and Mello, C. C. (1999). WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in C. elegans. *Cell* **97**, 717–26.

Roh, H. C., Dimitrov, I., Deshmukh, K., Zhao, G., Warnhoff, K., Cabrera, D., Tsai, W. and Kornfeld, K. (2014). A modular system of DNA enhancer elements mediates tissue-specific activation of transcription by high dietary zinc in C. elegans. *Nucleic Acids Res.* **43**, 803–816.

Romney, S. J., Thacker, C. and Leibold, E. a. (2008). An Iron Enhancer Element in the FTN-1 gene directs iron-dependent expression in Caenorhabditis elegans intestine. *J. Biol. Chem.* **283**, 716–725.

Romney, S. J., Newman, B. S., Thacker, C. and Leibold, E. a. (2011). HIF-1 regulates iron homeostasis in caenorhabditis elegans by activation and inhibition of genes involved in iron uptake and storage. *PLoS Genet.* **7**,.

Ruvkun, G. and Giusto, J. (1989). The Caenorhabditis elegans heterochronic gene lin-14 encodes a nuclear protein that forms a temporal developmental switch. *Nature* **338**, 313–319.

Sassi, H. E., Renihan, S., Spence, A. M. and Cooperstock, R. L. (2005). Gene CATCHR--gene cloning and tagging for Caenorhabditis elegans using yeast homologous recombination: a novel approach for the analysis of gene expression. *Nucleic Acids Res.* **33**, e163.

Savage, C., Das, P., Finelli, A. L., Townsend, S. R., Sun, C. Y., Baird, S. E. and Padgett, R. W. (1996). Caenorhabditis elegans genes sma-2, sma-3, and sma-4 define a conserved family of transforming growth factor beta pathway components. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 790–4.

Savage-Dunn, C., Tokarz, R., Wang, H., Cohen, S., Giannikas, C. and Padgett, R. W. (2000). SMA-3 smad has specific and critical functions in DBL-1/SMA-6 TGFbeta-related signaling. *Dev. Biol.* **223**, 70–6.

Schroeder, D. F. and McGhee, J. D. (1998). Anterior-posterior patterning within the Caenorhabditis elegans endoderm. *Development* **125**, 4877–4887.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* **473**, 337–42.

Severance, S., Rajagopal, A., Rao, A. U., Cerqueira, G. C., Mitreva, M., El-Sayed, N. M., Krause, M. and Hamza, I. (2010). Genome-wide analysis reveals novel genes essential for heme homeostasis in Caenorhabditis elegans. *PLoS Genet.* **6**, e1001044.

Shapira, M., Hamlin, B. J., Rong, J., Chen, K., Ronen, M. and Tan, M.-W. (2006). A conserved role for a GATA transcription factor in regulating epithelial innate immune responses. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 14086–91.

Shetty, P., Lo, M.-C., Robertson, S. M. and Lin, R. (2005). C. elegans TCF protein, POP-1, converts from repressor to activator as a result of Wnt-induced lowering of nuclear levels. *Dev. Biol.* **285**, 584–92.

Shibata, Y., Fujii, T., Dent, J. A., Fujisawa, H. and Takagi, S. (2000). EAT-20, a novel transmembrane protein with EGF motifs, is required for efficient feeding in Caenorhabditis elegans. *Genetics* **154**, 635–46.

Shimizu, R. and Yamamoto, M. (2005). Gene expression regulation and domain function of hematopoietic GATA factors. *Semin. Cell Dev. Biol.* **16**, 129–36.

Shoichet, S. A., Malik, T. H., Rothman, J. H. and Shivdasani, R. A. (2000). Action of the Caenorhabditis elegans GATA factor END-1 in Xenopus suggests that similar mechanisms initiate endoderm development in ecdysozoa and vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4076–81.

Siegfried, K. R., Kidd, A. R., Chesney, M. a. and Kimble, J. (2004). The sys-1 and sys-3 Genes Cooperate with Wnt Signaling to Establish the Proximal-Distal Axis of the Caenorhabditis elegans Gonad. *Genetics* **166**, 171–186.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539.

Sinclair, J. and Hamza, I. (2010). A novel heme-responsive element mediates transcriptional regulation in Caenorhabditis elegans. *J. Biol. Chem.* **285**, 39536–39543.

Smit, R. B., Schnabel, R. and Gaudet, J. (2008). The HLH-6 transcription factor regulates C. elegans pharyngeal gland development and function. *PLoS Genet.* **4**, e1000222.

Sommermann, E. M., Strohmaier, K. R., Maduro, M. F. and Rothman, J. H. (2010). Endoderm development in Caenorhabditis elegans: the synergistic action of ELT-2 and -7 mediates the specification→differentiation transition. *Dev. Biol.* **347**, 154–66.

Son, L. T., Ko, K.-M., Cho, J. H., Singaravelu, G., Chatterjee, I., Choi, T.-W., Song, H.-O., Yu, J.-R., Park, B.-J., Lee, S.-K., et al. (2011). DHS-21, a dicarbonyl/L-xylulose reductase (DCXR) ortholog, regulates longevity and reproduction in Caenorhabditis elegans. *FEBS Lett.* **585**, 1310–6.

Stainier, D. Y. R. (2002). A glimpse into the molecular entrails of endoderm formation. *Genes Dev.* **16**, 893–907.

Sulston, J. E. and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev. Biol.* **56**, 110–56.

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

Suzuki, Y., Yandell, M. D., Roy, P. J., Krishna, S., Savage-Dunn, C., Ross, R. M., Padgett, R. W. and Wood, W. B. (1999). A BMP homolog acts as a dose-dependent regulator of body size and male tail patterning in Caenorhabditis elegans. *Development* **126**, 241–50.

Suzuki, M., Kobayashi-Osaki, M., Tsutsumi, S., Pan, X., Ohmori, S., Takai, J., Moriguchi, T., Ohneda, O., Ohneda, K., Shimizu, R., et al. (2013). GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation. *Genes to Cells* **18**, 921–933.

Tcherepanova, I., Bhattacharyya, L., Rubin, C. S. and Freedman, J. H. (2000). Aspartic proteases from the nematode Caenorhabditis elegans: Structural organization and developmental and cell-specific expression of asp-1. *J. Biol. Chem.* **275**, 26359–26369.

Thacker, C., Sheps, J. A. and Rose, A. M. (2006). Caenorhabditis elegans dpy-5 is a cuticle procollagen processed by a proprotein convertase. *Cell. Mol. Life Sci.* **63**, 1193–204.

Thomas-Chollier, M., Defrance, M., Medina-Rivera, A., Sand, O., Herrmann, C., Thieffry, D. and van Helden, J. (2011). RSAT 2011: regulatory sequence analysis tools. *Nucleic Acids Res.* **39**, W86–91.

Thorpe, C. J., Schlesinger, A., Carter, J. C. and Bowerman, B. (1997). Wnt signaling polarizes an early C. elegans blastomere to distinguish endoderm from mesoderm. *Cell* **90**, 695–705.

Tian, C., Sen, D., Shi, H., Foehr, M. L., Plavskin, Y., Vatamaniuk, O. K. and Liu, J. (2010). The RGM protein DRAG-1 positively regulates a BMP-like signaling pathway in Caenorhabditis elegans. *Development* **137**, 2375–84.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J. and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–5.

Travers, C., Sijia, H. and Lana X, G. (2014). Power analysis and sample size estimation for RNA-Seq differential expression. *RNA-Seq Blog* 1684–1696.

Tseng, W. F., Jang, T. H., Huang, C. Ben and Yuh, C. H. (2011). An evolutionarily conserved kernel of gata5, gata6, otx2 and prdm1a operates in the formation of endoderm in zebrafish. *Dev. Biol.* **357**, 541–557.

Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–10.

Viger, R. S., Guittot, S. M., Anttonen, M., Wilson, D. B. and Heikinheimo, M. (2008). Role of the GATA family of transcription factors in endocrine development, function, and disease. *Mol. Endocrinol.* **22**, 781–98.

Wang, J., Tokarz, R. and Savage-Dunn, C. (2002). The expression of TGFbeta signal transducers in the hypodermis regulates body size in C. elegans. *Development* **129**, 4989–98.

Watanabe, N., Nagamatsu, Y., Gengyo-Ando, K., Mitani, S. and Ohshima, Y. (2005). Control of body size by SMA-5, a homolog of MAP kinase BMK1/ERK5, in C. elegans. *Development* **132**, 3175–84.

Watanabe, N., Ishihara, T. and Ohshima, Y. (2007). Mutants carrying two sma mutations are super small in the nematode C. elegans. *Genes Cells* **12**, 603–9.

Wightman, B., Bürglin, T. R., Gatto, J., Arasu, P. and Ruvkun, G. (1991). Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. *Genes Dev.* **5**, 1813–24.

Wightman, B., Ha, I. and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855–862.

Yagi, R., Zhu, J. and Paul, W. E. (2011). An updated view on transcription factor GATA3-mediated regulation of Th1 and Th2 cell differentiation. *Int. Immunol.* **23**, 415–20.

Yan, J. (2007). M.Sc. Thesis.

Yang, X.-D., Huang, S., Lo, M.-C., Mizumoto, K., Sawa, H., Xu, W., Robertson, S. and Lin, R. (2011). Distinct and mutually inhibitory binding by two divergent β-catenins coordinates TCF levels and activity in C. elegans. *Development* **138**, 4255–65.

Yi, W. and Zarkower, D. (1999). Similarity of DNA binding and transcriptional regulation by Caenorhabditis elegans MAB-3 and Drosophila melanogaster DSX suggests conservation of sex determining mechanisms. *Development* **126**, 873–81.

Yin, J., Madaan, U., Park, A., Aftab, N. and Savage-Dunn, C. (2015). Multiple cis elements and GATA factors regulate a cuticle collagen gene in Caenorhabditis elegans. *Genesis*.

Yook, K., Harris, T. W., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., Davis, P., de la Cruz, N., Duong, A., Fang, R., et al. (2012). WormBase 2012: more genomes, more data, new website. *Nucleic Acids Res.* **40**, D735–41.

Yoshida, S., Morita, K., Mochii, M. and Ueno, N. (2001). Hypodermal expression of Caenorhabditis elegans TGF-beta type I receptor SMA-6 is essential for the growth and maintenance of body length. *Dev. Biol.* **240**, 32–45.

Yuan, J., Tirabassi, R. S., Bush, A. B. and Cole, M. D. (1998). The C. elegans MDL-1 and MXL-1 proteins can functionally substitute for vertebrate MAD and MAX. *Oncogene* **17**, 1109–18.

Zhang, S., Sun, W.-Y., Wu, J.-J. and Wei, W. (2014). TGF-β signaling pathway as a pharmacological target in liver diseases. *Pharmacol. Res.* **85**, 15–22.

Zheng, B., Wen, J.-K. and Han, M. (2003). Regulatory factors involved in cardiogenesis. *Biochem. Biokhimiia* **68**, 650–7.

Zheng, R., Rebolledo-Jaramillo, B., Zong, Y., Wang, L., Russo, P., Hancock, W., Stanger, B. Z., Hardison, R. C. and Blobel, G. a. (2013). Function of GATA factors in the adult mouse liver. *PLoS One* **8**, 1–19.

Zhu, J., Hill, R. J., Heid, P. J., Fukuyama, M., Sugimoto, A., Priess, J. R. and Rothman, J. H. (1997). end-1 encodes an apparent GATA factor that specifies the endoderm precursor in Caenorhabditis elegans embryos. *Genes Dev.* **11**, 2883–96.

Zhu, J., Fukushige, T., McGhee, J. D. and Rothman, J. H. (1998). Reprogramming of early embryonic blastomeres into endodermal progenitors by a Caenorhabditis elegans GATA factor. *Genes Dev.* **12**, 3809–14.

Zorn, A. M. and Wells, J. M. (2009). Vertebrate endoderm development and organ formation. *Annu. Rev. Cell Dev. Biol.* **25**, 221–51.

Zugasti, O. and Ewbank, J. J. (2009). Neuroimmune regulation of antimicrobial peptide expression by a noncanonical TGF-beta signaling pathway in Caenorhabditis elegans epidermis. *Nat. Immunol.* **10**, 249–56.

Appendices A: Chapter 2 Tables

Table 2.1 Frequency of Dpy Phenotype Rescue

Transgene	% non-Dpy	% Intermediate	n
dpy-7p::dpy-7	100%	0%	33
myo-2p::dpy-7 + marg-1p::dpy-7	0%	0%	21
rol-6p::dpy-7	74%	9%	46

Rescue of *dpy-7* mutants by different *dpy-7*–expressing transgenes, expressed as the percentage of transgenic animals showing rescue of the Dpy phenotype ('% non-Dpy'). Representative animals are shown in Supplemental Figure 2.4.

Table 2.2 Percentage of Wt and *sma-3(e491)* animals (+ or - pharyngeal *sma-3* minigene constructs) that had laid an egg by the time point indicated (hrs after initial egg laid)

Strain/Time Point	48 hrs	66 hrs	72 hrs	90 hrs	96 hrs	114 hrs	120 hrs	n
Wt	0%	8.7%	34.8%	100%	100%	100%	100%	23
sma-3(e491)	0%	0%	0%	20%	68%	100%	100%	25
sma-3(e491) myo-2p::sma-3 + marg-1p::sma-3 (-)	0%	0%	0%	42.9%	66.7%	100%	100%	21
sma-3(e491) myo-2p::sma-3 + marg-1p::sma-3 (+)	0%	0%	0%	88.2%	91.1%	100%	100%	34

Supplemental Table 2.1 Pharynx and Body Lengths of Various Strains

Strain	Pharynx Length (µm)	Body Length (µm)	n
Wt	140 ± 3	1253 ± 66	37
sma-3(e491)[a]	106 ± 2	629 ± 55	38
sma-3(wk30)	114 ± 3	776 ± 56	35
sma-6(e1482)	114 ± 3	841 ± 65	34
dpy-10(e128)	129 ± 4	749 ± 58	38
dpy-5(e61)	125 ± 4	790 ± 35	36
sma-3(e491)/+	130 ± 3	1115 ± 51	33
sma-3(wk30)/+	136 ± 3	1170 ± 50	33

Supplemental Table 2.2 Tissue Specificity of Promoters Used to Drive Rescue Constructs

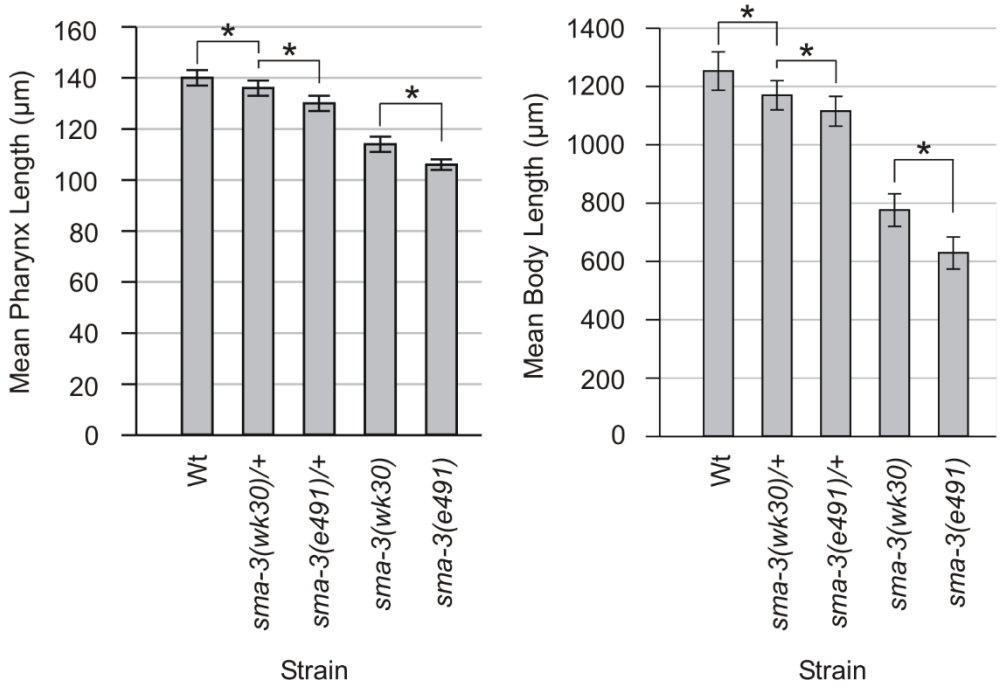
Promoter	Tissue(s) Active
sma-3	Hypodermis + Pharynx + Intestine
myo-2	Pharyngeal Muscle Cells
marg-1	Pharyngeal Marginal Cells
dpy-7	Hypodermis
rol-6	Hypodermis
K07C11.4	Pharynx + Intestine + Somatic Gonad
elt-3	Hypodermis

Supplemental Table 2.3 Pharynx and Body Lengths of Strains Imaged Using Various Anesthetics at Different Time Points After Egg Laying During Development

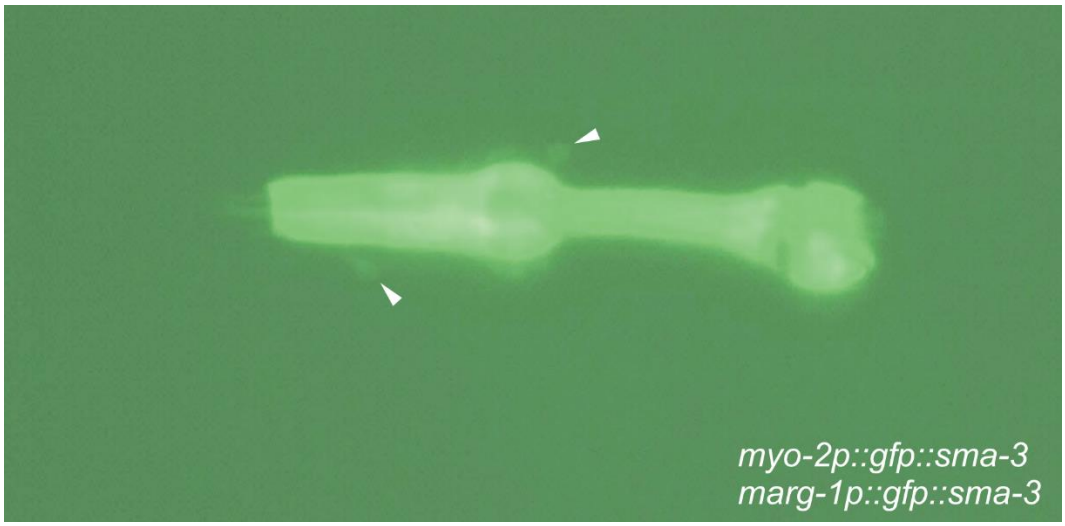
Strain	DNA Conc. (ng/μL)	Line	Pharynx Length (μm)		Body Length (μm)		n		Anesthetic/ Timepoint (hrs)
			-	+	-	+	-	+	
Wt						1069 ± 79		25	None/96
				142 ± 3		1279 ± 42		20	Na Az./96
				154 ± 4		1323 ± 69		27	Na Az./120
sma-3(e491)						700 ± 59		22	None/96
				113 ± 3		840 ± 66		28	Na Az./120
sma-3(e491) *sma-3p::sma-3*	20	A			658 ± 60	953 ± 82	22	23	None/96
			109 ± 4	130 ± 5	679 ± 56	1050 ± 68	32	34	Lev./96
		B	109 ± 4	130 ± 4	694 ± 57	1035 ± 61	32	33	Lev./96
sma-3(e491) *myo-2p::sma-3*	20	A	105 ± 3	116 ± 4	650 ± 48	801 ± 88	32	38	Lev./96
		B	109 ± 4	118 ± 4	665 ± 55	811 ± 97	33	38	Lev./96
sma-3(e491) *marg-1p::sma-3*	20	A	106 ± 2	114 ± 3	656 ± 37	784 ± 44	33	36	Lev./96
		B	106 ± 3	113 ± 4	661 ± 50	743 ± 72	34	34	Lev./96
sma-3(e491) *myo-2p::sma-3 +* *marg-1p::sma-3*	10 each	A	106 ± 3	119 ± 3	677 ± 48	844 ± 52	8	20	Lev./96
		B	106 ± 4	117 ± 5	658 ± 52	796 ± 53	10	20	Lev./96
	20 each	A	107 ± 3	121 ± 4	669 ± 45	870 ± 88	34	36	Lev./96
	30 each	A	107 ± 2	124 ± 4	633 ± 42	967 ± 76	36	36	Lev./96
					643 ± 52	865 ± 66	18	22	None/96
			109 ± 3	124 ± 4	736 ± 40	1023 ± 37	20	20	Na Az./96
			113 ± 4	131 ± 5	774 ± 51	1109 ± 93	26	29	Na Az./120
dbl-1(++); *sma-3(e491)* *myo-2p::sma-3 +* *marg-1p::sma-3*	30 each	A	112 ± 3	134 ± 5	782 ± 56	1356 ± 86	26	28	Na Az./120
sma-3(e491) *rol-6p::sma-3*	20	A	106 ± 3	107 ± 3	687 ± 49	712 ± 50	34	32	Lev./96
		B	104 ± 3	105 ± 3	630 ± 45	640 ± 49	32	35	Lev./96
	50	A	103 ± 3	104 ± 3	639 ± 7	632 ± 37	3	10	Lev./96
		B	106 ± 3	107 ± 3	644 ± 39	657 ± 38	11	10	Lev./96
sma-3(e491) *myo-2p::sma-3 +* *marg-1p::sma-3 +* *rol-6p::sma-3*	5 each	A	105 ± 3	125 ± 3	652 ± 44	972 ± 86	9	21	Lev./96
		B	103 ± 4	119 ± 3	641 ± 51	847 ± 50	11	21	Lev./96
	20 each	A	107 ± 2	126 ± 3	661 ± 49	1020 ± 90	35	37	Lev./96
		B	108 ± 2	128 ± 3	687 ± 60	1037 ± 69	36	36	Lev./96
sma-3(wk30) *elt-3p::GFP::sma-3*	20	A	114 ± 3	122 ± 4	829 ± 42	1011 ± 87	36	37	Lev./96
			121 ± 4	130 ± 5	867 ± 56	1178 ± 100	26	24	Na Az./120
sma-3(wk30) *dpy-7p::GFP::sma-3*	20	A		112 ± 4		908 ± 53		33	Lev./96
sma-3(wk30) *rol-6p::sma-3*	20	A	114 ± 4	114 ± 3	827 ± 56	827 ± 58	33	36	Lev./96
sma-3(wk30) *myo-2p::sma-3 +* *marg-1p::sma-3*	20 each	A	114 ± 3	122 ± 6	806 ± 36	907 ± 50	31	36	Lev./96
sma-3(wk30) *myo-2p::sma-3 +* *marg-1p::sma-3 +* *rol-6p::sma-3*	20 each	A	114 ± 4	130 ± 4	785 ± 36	1074 ± 80	32	34	Lev./96
sma-3(wk30) K07C11.4*p::sma-3 +* *rol-6p::sma-3*	20 each	A	112 ± 3	134 ± 5	794 ± 56	1191 ± 79	35	36	Lev./96

Appendices B: Chapter 2 Supplemental Figures

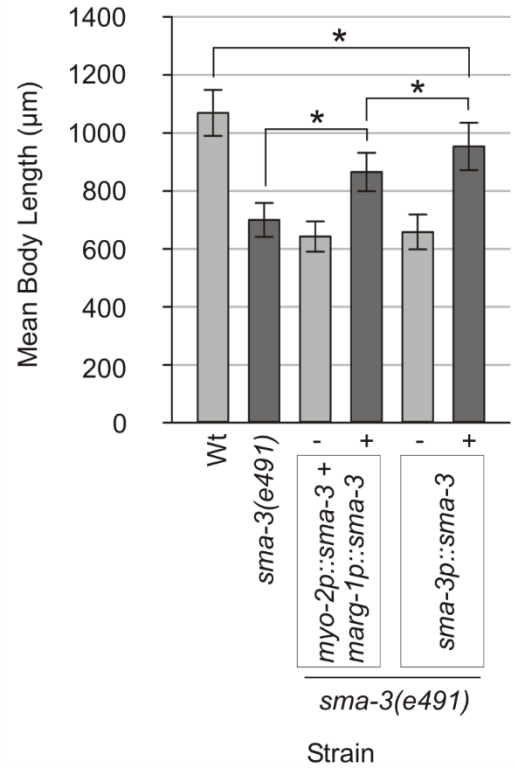
Supplemental Figure 2.1. Mean pharynx and body length measurements ± standard deviation of Wild type (Wt) N2, *sma-3* mutants and *sma-3* heterozygotes. Complete data is provided in Supplemental Table 2.1. * denotes statistically significant differences of p<0.001. All other differences in pharynx and body lengths between strains not directly indicated on the graphs are significant (p<0.001).


Mean Pharynx Length (µm)
160
140
120
100
80
60
40
20
0
*
*
*
Wt
sma-3(wk30)/+
sma-3(e491)/+
sma-3(wk30)
sma-3(e491)
Strain
Mean Body Length (µm)
1400
1200
1000
800
600
400
200
0
*
*
*
Wt
sma-3(wk30)/+
sma-3(e491)/+
sma-3(wk30)
sma-3(e491)
Strain

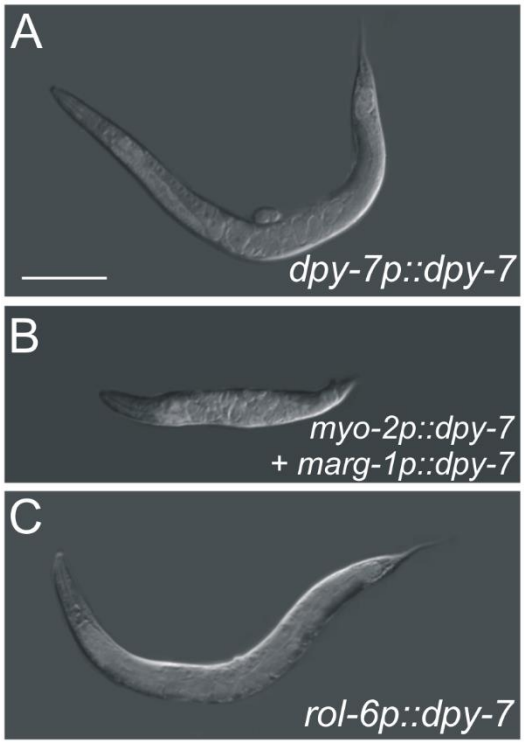
Supplemental Figure 2.2. Pharyngeal expression of the *sma-3* minigene carrying an in-frame N-terminal GFP tag under the control of the *myo-2* and *marg-1* promoters. Very weak expression is occasionally observed outside of the pharynx in some animals (arrowheads).


myo-2p::gfp::sma-3
marg-1p::gfp::sma-3

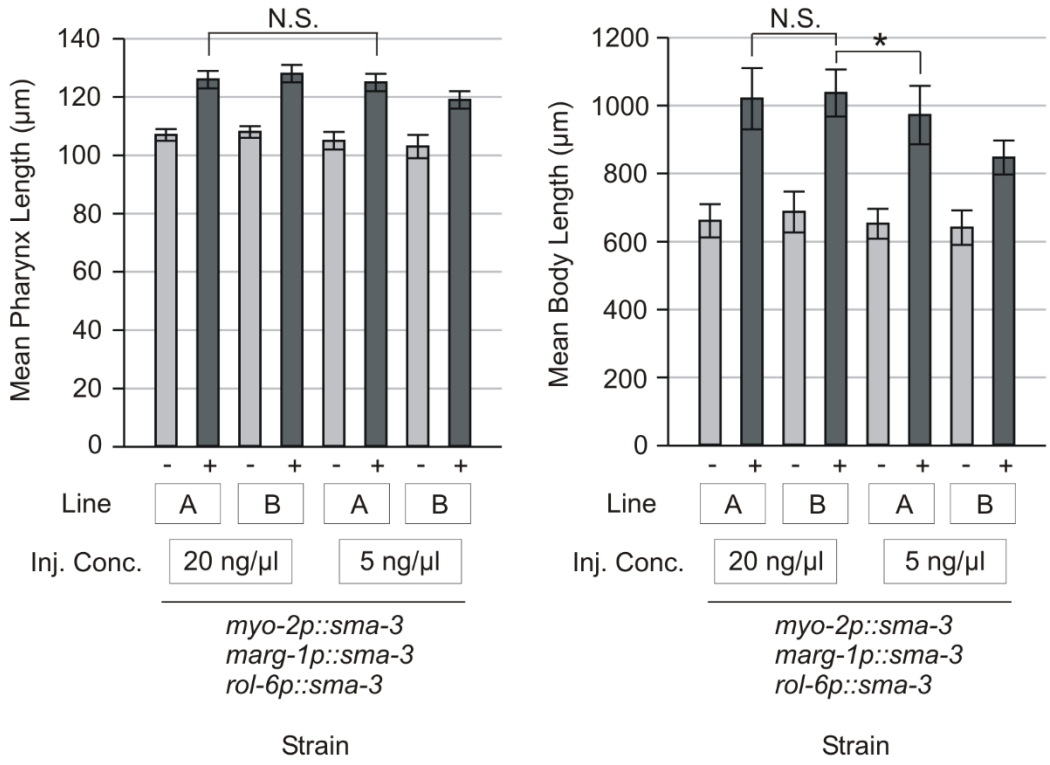
Supplemental Figure 2.3. Mean body length measurements ± standard deviation (measured under a dissecting microscope in the absence of anesthetic) of Wild type (Wt) N2, *sma-3(e491)* and *sma-3(e491)* animals from various *sma-3* minigene rescue experiments. Vertical labels indicate tissue specific promoter-*sma-3* minigene fusion rescue constructs in each strain. In each case, we measured animals carrying the transgenic array (+) and siblings that lacked the array (-), as before. Complete data for all lines is provided in Supplemental Table 2.3. All transgenic animal means (+) were statistically significantly different from non-transgenic sibling means (-) ($p<0.001$). * denotes significant differences of $p<0.001$. All other differences in body lengths between strains not directly indicated on the graphs are significant ($p<0.05$).


1400
1200
1000
800
600
400
200
0
Mean Body Length (µm)
*
*
*
Wt
sma-3(e491)
-
+
-
+
myo-2p::sma-3 + marg-1p::sma-3
sma-3p::sma-3
sma-3(e491)
Strain

Supplemental Figure 2.4. Rescue of *dpy-7* mutants by different transgenes. (A) Rescue of *dpy-7* by a *dpy-7p::dpy-7* transgene. (B) Expression of *dpy-7* under the control of the pharyngeal promoters *myo-2* and *marg-1* does not rescue the *dpy-7* phenotype. (C) Rescue of *dpy-7* by a *rol-6p::dpy-7* transgene. Complete data is provided in Table 1. Scale bar is 150µm.


A
dpy-7p::dpy-7
B
myo-2p::dpy-7
+ marg-1p::dpy-7
C
rol-6p::dpy-7

Supplemental Figure 2.5. Measurements are means ± standard deviation. Transgenic lines were established using injection mixes containing *myo-2p::sma-3*, *marg-1p::sma-3* and *rol-6p::sma-3*, each at a concentration of either 20 ng/µL or 5 ng/µL, as indicated. A and B indicate independently generated transgenic lines. All transgenic animal means (+) were statistically significantly different from non-transgenic sibling means (-) ($p<0.001$). All differences in pharynx length not directly indicated on the graph are significant ($p<0.001$), except where indicated by N.S. * denotes significant differences of $p<0.05$. All other differences in body length not directly indicated on the graph are significant ($p<0.05$) except where indicated by N.S. Complete data for multiple lines is provided in Supplemental Table 2.3.


N.S.
Mean Pharynx Length (µm)
140
120
100
80
60
40
20
0
- + - + - + - +
Line
A
B
A
B
Inj. Conc.
20 ng/µl
5 ng/µl
myo-2p::sma-3
marg-1p::sma-3
rol-6p::sma-3
Strain
N.S.
*
Mean Body Length (µm)
1200
1000
800
600
400
200
0
- + - + - + - +
Line
A
B
A
B
Inj. Conc.
20 ng/µl
5 ng/µl
myo-2p::sma-3
marg-1p::sma-3
rol-6p::sma-3
Strain

Appendices C: Chapter 4 Tables

Table 4.1 Strains used to generate genetically distinct populations for RNAseq

Strain	Background	GFP
Wt	N2	Negative
JM222	*elt-7(tm840) V*	Negative
JM147	*elt-2(ca15) X; caEx3[elt-2(+), rol-6(su1006), sur-5p::gfp]*	Positive and Negative
JM199	*elt-7(tm840)V; elt-2(ca15) X; caEx3[elt-2(+), rol-6(su1006), sur-5p::gfp]*	Positive and Negative

Table 4.2 Mean body length measurements of sorted L1 populations

Population	Body Length (µm)	n
Wt (s)	212 ± 15	35
elt-7(tm840) (s)	247 ± 13	22
elt-2(+++)	225 ± 19	29
elt-2(ca15)	230 ± 12	40
elt-7(tm840); elt-2(+++)	231 ± 40	28
elt-7(tm840); elt-2(ca15)	228 ± 20	19

Table 4.3 Mean total reads obtained from RNAseq per population and mean percentage of those reads that mapped to the C. elegans genome

Population	Mean Total Reads (Millions)	Mean % Mapped	n
Wt (f)	15.8 ± 2.9	92 ± 1.7	6
Wt (s)	20.0 ± 0.6	91 ± 0.8	4
elt-7(tm840) (f)	17.5 ± 0.9	93 ± 1.3	3
elt-7(tm840) (s)	18.2 ± 0.5	92 ± 0.4	3
elt-2(+++)	14.7 ± 6.0	96 ± 1.6	4
elt-2(ca15)	19.8 ± 5.0	95 ± 2.8	4
elt-7(tm840); elt-2(+++)	17.1 ± 0.5	96 ± 1.7	4
elt-7(tm840); elt-2(ca15)	19.0 ± 1.7	95 ± 1.0	4

Table 4.4 Differentially expressed gene counts in various comparisons arranged by log2 fold change and expression in the intestine

Control Strain	Experimental Strain	DE Genes	Up > 0.5	Down < -0.5	Up > 0.5 Intestinal	Down < -0.5 Intestinal
Wt (s)	*elt-7(tm840)*(s)	2923	622	340	126	38
Wt (s)	*elt-2(ca15)*	5236	1535	1418	331	174
Wt (s)	*elt-7(tm840); elt-2(ca15)*	4295	1350	1264	238	223
elt-7(tm840)(s)	*elt-7(tm840); elt-2(ca15)*	3644	1221	1107	200	209
elt-2(ca15)	*elt-7(tm840); elt-2(ca15)*	2472	534	599	60	171
Wt (s)	*elt-2*(+++)	1585	340	256	90	24

Table 4.5 Log2 fold changes and associated adjusted p values for the wild-type vs *elt-2(+++)* comparison and related information for intestinal genes that positively respond to ELT-2

Gene	Protein	Exp.	Log2 Fold Change	p adj	Promoter Size (kb)	Number TGATAA
act-5	Cytoplasmic Actin	IE	0.75	4.77E-06	2.000	10
asp-1	Aspartic protease	IE	0.99	9.24E-06	2.000	3
asp-3	Aspartic protease	IP	1.19	8.47E-13	1.638	4
asp-5	Aspartic protease	IE	1.17	1.09E-09	0.378	1
C08E3.13	Novel protein	IP	0.53	5.82E-04	1.494	2
C49C8.5	Uncharacterized	IE	0.74	3.81E-05	0.945	2
cdr-4	CaDmium Responsive TM	IP	0.61	8.28E-05	0.846	1
cebp-1	bZip domain-containing protein	IP	0.98	1.03E-11	2.000	4
clec-50	C-type lectin	IE	0.53	3.48E-03	1.262	1
clec-52	C-type lectin	IP	1.16	4.52E-09	1.246	3
clec-63	C-type lectin	IP	1.29	7.10E-05	0.541	3
clec-65	C-type lectin	IE	0.67	6.14E-03	0.929	2
clec-66	C-type lectin	IE	0.74	3.06E-03	1.121	2
clec-67	C-type lectin	IE	1.44	5.95E-08	0.581	3
clec-85	C-type lectin	IE	1.24	8.56E-07	1.524	3
cpr-1	Cysteine protease	IE	0.99	2.25E-04	2.000	5
cpr-3	Cysteine protease	IP	1.51	1.38E-32	2.000	5
cyp-13A7	cytochrome P450	IE	1.35	3.76E-11	2.000	4
ddl-2	Novel proline-rich protein	IP	0.54	5.30E-02	0.305	0
F49E12.9	Uncharacterized	IE	1.31	1.30E-06	2.000	3
fil-1	Lipase	IE	0.69	2.90E-02	1.342	4
gem-4	Ca[2+]-dependent phosphatidylserine binding	IP	0.70	1.98E-03	2.000	2

ifd-2	Intermediate filament protein	IE	0.71	1.81E-04	2.000	4
ifp-1	Intermediate filament protein	IE	0.57	1.59E-04	2.000	3
K12H4.7	Putative serine protease	IE	0.71	7.98E-03	2.000	0
lec-6	Beta-galactosyl-binding lectin	IP	1.00	2.24E-12	2.000	8
lipl-1	Lipase	IE	0.83	2.34E-03	1.104	3
lys-1	Putative lysozyme	IP	0.88	1.16E-05	2.000	4
lys-7	Lysozyme	IP	1.44	1.17E-06	0.491	1
lys-8	Lysozyme	IP	1.34	4.49E-19	1.485	6
M176.4	Uncharacterized	IP	0.67	2.24E-05	2.000	5
mtl-2	Metallothionein	IE	0.52	6.05E-02	0.580	3
nlp-16	Predicted neuropeptide	IP	0.53	5.97E-02	1.215	1
nlp-9	Neuropeptide-like protein	IP	0.70	6.53E-04	2.000	1
pho-1	Acid phosphatase	IE	0.78	9.94E-05	0.425	3
pmp-1	ABC transporter	IE	0.73	1.96E-03	0.379	2
pqn-25	Prion-like-(Q/N-rich)-domain-bearing	IP	0.69	4.20E-02	0.518	0
slcf-1	Monocarboxylate transporter	IE	0.80	1.65E-07	0.273	2
smd-1	S-adenosylmethionine decarboxylase	IP	0.69	1.75E-03	2.000	4
spp-3	SPP antimicrobial peptide	IP	0.62	2.92E-03	0.747	3
tag-244	Uncharacterized	IE	0.72	3.22E-06	0.703	1
tth-1	Thymosin beta ortholog	IP	0.76	2.93E-05	2.000	1
ugt-44	Ortholog of human UDP glycosyltransferase 3 family, polypeptide A1	IP	1.28	2.29E-06	2.000	6
W02A2.9	Uncharacterized	IP	0.65	2.68E-05	2.000	6
ZC416.6	Similar to bifunctional leukotriene A4 hydrolases/aminopeptidases	IE	0.57	1.27E-02	0.982	2
zip-10	bZIP transcription factor	IP	1.75	2.58E-23	2.000	4
ztf-22	Putative zinc finger transcription factor	IP	0.52	2.43E-04	2.000	2

Table 4.6 Log2 fold changes and associated adjusted p values for the wild-type vs *elt-2(ca15)* comparison and related information for intestinal genes that are exclusively activated by ELT-2

Gene	Protein	Exp.	Log2 Fold Change	p adj	Promoter Size (up to 2 kb)	Number TGATAA
alh-13	Aldehyde dehydrogenase	IE	-0.67	1.12E-03	2.000	5
asah-1	AcylSphingosine AmidoHydrolase	IE	-2.00	3.48E-06	0.209	2
asp-1	Aspartic protease	IE	-3.03	3.58E-42	2.000	3
asp-5	Aspartic protease	IE	-2.38	8.44E-52	0.378	1
C45G9.13	Uncharacterized	IP	-1.01	9.15E-15	2.000	1
C56G2.9	Uncharacterized	IP	-0.94	3.39E-08	2.000	1
clec-85	C-type lectin	IE	-1.26	1.23E-07	1.524	3
cyp-33E1	Cytochrome P450	IE	-0.82	4.74E-13	0.509	1
eat-4	Ortholog of mammalian BNPI vesicular glutamate transporter	IP	-0.65	4.21E-11	2.000	2
elo-6	Polyunsaturated fatty acid elongase	IP	-0.70	1.69E-04	1.805	3
F08A8.2	Acyl-coenzyme A oxidase	IE	-1.00	8.35E-06	0.187	0
F10E9.11	Uncharacterized	IP	-1.05	3.30E-15	0.490	1
glc-4	Glutamate-gated chloride channel	IP	-0.88	5.50E-14	2.000	1
gly-19	2/I N-acetylglucosaminyltransferase	IP	-0.94	2.62E-07	0.729	3
gpc-1	Heterotrimeric guanine nucleotide-binding protein gamma subunit	IP	-1.03	4.91E-12	2.000	2
haf-4	Half-molecular ABC transporter	IE	-4.56	5.27E-150	0.279	3
haf-9	Half-molecular ABC transporter	IE	-4.30	1.27E-209	2.000	2
K10C2.1	Uncharacterized	IE	-3.78	1.05E-135	0.221	1
lurp-1	Ly6 Urokinase plasminogen domain Receptor-related Protein	IP	-1.05	1.65E-24	2.000	0
math-35	Meprin-associated Traf homology domain containing	IE	-0.94	1.56E-06	0.517	1
mxl-1	bHLH transcription factor	IP	-0.71	6.97E-13	0.465	1
nca-2	Alpha-1 subunit of a voltage-insensitive cation leak channel	IP	-0.91	4.77E-20	2.000	1
nhr-108	Nuclear hormone receptor	IP	-1.23	3.29E-15	0.839	1
nlp-1	Neuropeptide-like protein	IP	-1.02	2.68E-49	2.000	2

ser-6	G protein-coupled receptor	IP	-0.74	4.16E-08	2.000	1
spp-1	Saposin (B) domain-containing protein	IE	-6.27	1.03E-128	0.450	4
vha-6	Vacuolar proton-translocating ATPase subunit	IP	-1.06	1.77E-14	0.755	2
Y75B8A.4	Lon protease homolog 2, peroxisomal	IE	-0.59	4.08E-04	1.291	3
zig-5	IG domain protein	IP	-0.65	3.59E-14	2.000	1
ZK550.6	Ortholog of the human gene Phytanoyl-COA hydroxylase	IE	-0.83	2.19E-19	1.820	5

Table 4.7 Log2 fold changes and associated adjusted p values for the wild-type vs *elt-7(tm840); elt-2(ca15)* comparison and related information for intestinal genes that are redundantly activated by ELT-2 or ELT-7

Gene	Protein	Exp.	Log2 Fold Change	p adj	Promoter Size (up to 2 kb)	Number TGATAA
aex-6	Rab small GTPase	IP	-0.52	1.65E-03	2.000	0
aqp-4	Aquaporin	IE	-3.16	4.66E-10	2.000	8
C50B6.7	Uncharacterized	IE	-1.38	1.00E-53	2.000	3
cdf-2	Cation Diffusion Facilitator	IE	-1.49	7.75E-14	1.371	4
clec-5	C-type lectin	IP	-3.00	6.44E-29	2.000	2
clec-65	C-type lectin	IE	-2.21	1.21E-02	0.929	2
dsc-4	Microsomal triglyceride transfer protein	IE	-2.19	3.03E-15	1.100	2
F10A3.4	Uncharacterized	IP	-3.11	6.53E-78	2.000	2
F13D12.6	Ortholog of human PPGB	IE	-2.43	6.80E-04	2.000	10
F57F4.4	ET Domain repeats	IE	-2.97	7.88E-34	2.000	4
flr-1	DEG/ENaC sodium channel	IE	-1.79	1.95E-19	0.355	2
fmo-1	Flavin-containing monoxygenase	IP	-0.77	1.86E-05	1.413	5
ges-1	Type B carboxylesterase	IE	-0.96	9.17E-12	2.000	3
gpx-1	Phospholipid hydroperoxide glutathione peroxidase	IP	-0.95	1.23E-03	1.024	1
haly-1	Histidine ammonia lyase	IP	-0.99	5.46E-09	1.701	1
hrg-1	Orthologous to vertebrate heme transporters	IE	-2.27	8.86E-16	2.000	2
ifb-2	Intermediate filament protein	IE	-2.10	1.10E-57	2.000	4
ifd-2	Intermediate filament protein	IE	-2.26	8.40E-15	2.000	4
inx-16	Pannexin	IE	-1.71	4.82E-23	0.946	4
K02D7.1	Ortholog of human nucleoside phosphorylase	IE	-0.54	2.70E-04	2.000	6
lys-8	Lysozyme	IP	-0.83	1.69E-03	1.485	6

mtl-1	Metallothioneins	IP	-1.01	7.26E-05	0.520	3
nlp-20	Neuropeptide-Like Protein	IP	-1.20	8.53E-13	2.000	1
nlp-36	Neuropeptide-Like Protein	IP	-0.95	2.38E-15	2.000	2
pept-1	Low affinity/high capacity oligopeptide transporter	IE	-2.03	1.28E-38	1.898	5
pmp-1	ABC transporter peroxisomal membrane related protein	IE	-2.31	2.25E-09	0.379	2
R09H10.5	Uncharacterized	IE	-2.03	1.37E-30	0.662	3
ttm-4	Uncharacterized	IE	-0.61	1.94E-05	1.223	2
Y106G6H.1	Uncharacterized	IE	-3.67	1.19E-46	0.714	1
Y82E9BR.23	Uncharacterized	IP	-2.46	6.94E-03	0.660	2
ZC416.6	Similar to bifunctional leukotriene A4 hydrolases/aminopeptidases	IE	-2.26	1.31E-03	0.982	2
ZK550.2	Uncharacterized	IE	-1.06	5.95E-09	2.000	8

Table 4.8 Log2 fold changes and associated adjusted p values for the wild-type vs *elt-7(tm840)*, wild-type vs *elt-2(ca15)* and wild-type vs *elt-7(tm840); elt-2(ca15)* comparisons and related information for intestinal genes that are additively activated by ELT-2 and ELT-7

Gene	Protein	Exp.	Wt vs *elt-7(tm840)*	Wt vs *elt-2(ca15)*	Wt vs *elt-7(tm840); elt-2(ca15)*	Promoter Size (up to 2 kb)	Number TGATAA
cgr-1	CRAL/TRIO and GOLD domain suppressor of activated Ras	IE	-0.08, (0.70)	-0.83, (1.58E-14)	-2.24, (1.62E-19)	1.384	4
elo-5	Polyunsaturated fatty acid elongase	IP	-0.29, (0.41)	-0.67, (3.41E-03)	-3.54, (1.78E-08)	2.000	5
**ndg-4*	NorDiHydroGuaiaretic acid resistant	IM	-0.66, (2.65E-04)	-0.88, (2.38E-09)	-2.19, (7.47E-37)	1.774	6
pal-1	Homeodomain Caudal ortholog	IP	-0.17, (0.35)	-0.60, (2.00E-05)	-1.31, (2.72E-18)	2.000	3
pcp-3	Serine-type peptidase	IP	-0.62, (3.09E-06)	-0.52, (3.42E-06)	-3.65, (1.35E-17)	1.880	7
ZC204.12	Uncharacterized	IP	0.07, (0.60)	-0.84, (1.56E-18)	-3.315, (2.02E-86)	2.000	2

Table 4.9: Log2 fold changes and associated adjusted p values for the wild-type vs *elt-7(tm840)*, wild-type vs *elt-2(ca15)* and *elt-2(ca15)* vs *elt-7(tm840); elt-2(ca15)* comparisons and related information for intestinal genes that are not synergistically activated by ELT-2 and ELT-7

Gene	Protein	Exp.	Wt vs *elt-7(tm840)*	Wt vs *elt-2(ca15)*	*elt-2(ca15)* vs *elt-7(tm840); elt-2(ca15)*
C01H6.4	Uncharacterized	IP	-0.65, (5.14E-04)	-0.54, (5.54E-05)	-0.76, (5.56E-06)*
mtl-2	Metallothionein	IE	-0.82, (3.68E-06)	-6.16, (2.01E-53)	1.57, (NA)*
pmp-2	ortholog of human ATP-binding cassette D (ALD) 3	IE	-0.54, (1.25E-03)	-1.34, (1.42E-21)	0.16, (0.57)
shw-3	Shaw-related subfamily, member 4, K^+ gated channel	IP	-0.73, (6.26E-04)	-0.88, (3.84E-08)	0.22, (0.40)*
sre-4	serpentine receptor class e	IP	-0.66, (3.12E-06)	-1.38, (6.87E-29)	0.06, (0.86)
ZK1320.3	Uncharacterized	IE	-0.69, (6.93E-05)	-5.28, (1.96E-51)	1.01, (NA)*

*****Base mean expression was below the cutoff of 100 in this DESeq2 comparison.

Table 4.10 Position frequency matrices from the different classes of activated intestinal genes and the originally identified intestinal gene promoter PFM from McGhee et al. (2009)

Original PFM*

Base	1	2	3	4	5	6	7	8	9	10
A	0.64	0.12	0.03	0.00	1.00	0.00	1.00	0.96	0.10	0.66
C	0.07	0.50	0.01	0.00	0.00	0.00	0.00	0.00	0.26	0.08
G	0.15	0.17	0.00	1.00	0.00	0.00	0.00	0.03	0.58	0.17
T	0.15	0.21	0.96	0.00	0.00	1.00	0.00	0.01	0.06	0.09

Exclusive Sites PFM

Base	1	2	3	4	5	6	7	8	9	10
A	0.49	0.19	0.00	0.00	1.00	0.00	1.00	1.00	0.25	0.56
C	0.11	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.14
G	0.09	0.18	0.00	1.00	0.00	0.00	0.00	0.00	0.42	0.07
T	0.32	0.30	1.00	0.00	0.00	1.00	0.00	0.00	0.21	0.23

Redundant Sites PFM

Base	1	2	3	4	5	6	7	8	9	10
A	0.45	0.24	0.00	0.00	1.00	0.00	1.00	1.00	0.31	0.47
C	0.15	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.17
G	0.10	0.16	0.00	1.00	0.00	0.00	0.00	0.00	0.29	0.18
T	0.29	0.33	1.00	0.00	0.00	1.00	0.00	0.00	0.17	0.18

Additive Sites PFM

Base	1	2	3	4	5	6	7	8	9	10
A	0.57	0.43	0.00	0.00	1.00	0.00	1.00	1.00	0.24	0.24
C	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.10
G	0.19	0.05	0.00	1.00	0.00	0.00	0.00	0.00	0.29	0.10
T	0.24	0.19	1.00	0.00	0.00	1.00	0.00	0.00	0.24	0.57

Table 4.11 Log2 fold changes and associated adjusted p values for the wild-type vs *elt-2(ca15)* comparison and related information for intestinal genes that are exclusively repressed by ELT-2

Gene	Protein	Exp.	Log2 Fold Change	p adj	Promoter Size (kb)	Number TGATAA
bath-41	BTB and MATH domains	IE	0.59	6.06E-06	0.298	3
C01H6.9	Kinase	IP	0.57	8.87E-03	0.600	0
C49C8.5	Uncharacterized	IE	0.61	1.02E-06	0.945	2
cdc-48.2	ATPase	IP	1.54	2.21E-126	1.603	2
cdf-1	Cation Diffusion Facilitator	IP	0.72	2.38E-11	2.000	2

cdr-4	CaDmium Responsive TM	IP	1.97	7.71E-58	0.846	1
cpr-3	Cysteine protease	IP	3.96	0.00	2.000	5
dct-18	Uncharacterized	IE	1.98	1.95E-129	2.000	4
djr-1.1	Glyoxylase	IE	0.60	2.26E-04	0.303	0
gem-4	Ca[2+]-dependent phosphatidylserine binding	IP	2.20	8.87E-54	2.000	2
itx-1	Caspr ortholog neurexin	IE	1.79	4.60E-92	2.000	2
lec-10	Galactose-binding lectin	IE	1.79	1.13E-79	1.198	0
lon-2	Heparan sulfate proteoglycan	IP	0.99	5.10E-17	2.000	3
mig-1	Frizzled-like receptor	IP	0.71	1.12E-12	2.000	3
nft-1	Ortholog of human FHIT	IP	0.71	2.11E-13	0.202	0
npp-11	Nucleoporin	IP	0.67	1.50E-11	1.282	1
npp-4	Nucleoporin	IE	0.58	3.06E-10	0.645	1
pgp-1	P-GlycoProtein ABC transporter	IP	1.95	2.03E-61	2.000	5
pqn-25	Prion-like-(Q/N-rich)-domain-bearing	IP	2.16	4.89E-25	0.518	0
rpn-9	Non-ATPase proteasome subunit	IP	1.02	5.56E-37	0.799	2
sec-24.1	Sec24 homolog	IE	0.63	6.61E-23	2.000	3
stc-1	ATPase similar to heat shock protein 70 family	IP	0.76	2.48E-17	0.503	0
ipla-7	Phospholipase A2	IP	1.29	1.07E-30	2.000	8
tth-1	Thymosin beta ortholog	IP	2.23	7.82E-89	2.000	1
ufd-3	Phospholipase A2-activating protein	IP	0.70	7.32E-18	0.746	0
W02A2.9	Uncharacterized	IP	1.73	1.32E-74	2.000	6
xpc-1	XPC DNA repair homolog	IP	0.50	2.03E-08	1.715	1
zip-10	bZIP transcription factor	IP	3.19	2.48E-101	2.000	4

Table 4.12 Log2 fold changes and associated adjusted p values for the wild-type vs *elt-7(tm840); elt-2(ca15)* comparison and related information for intestinal genes that are redundantly repressed by ELT-2 or ELT-7

Gene	Protein	Exp.	Log2 Fold Change	p adj	Promoter Size (kb)	Number TGATAA
C08E3.13	Novel protein	IP	1.42	3.45E-08	1.494	2
C30G12.2	Uncharacterized	IP	2.20	2.75E-31	2.000	4
cnp-2	CalciNeurin binding Protein	IE	2.17	2.01E-14	2.000	0
EEED8.2	Uncharacterized	IE	1.42	1.91E-10	1.466	3
lec-6	Beta-galactosyl-binding lectin	IP	3.00	1.79E-59	2.000	8
mnp-1	Matrix Non-Peptidase homolog	IP	1.70	9.20E-33	0.795	0
odd-2	Drosophila ODD-skipped- like	IP	0.58	1.06E-02	2.000	2
rig-3	NeuRonal IGCAM	IP	1.46	1.97E-17	2.000	1
ttr-46	Transthyretin-like protein	IP	0.72	1.35E-13	2.000	5

Table 4.13 Log2 fold changes and associated adjusted p values for the wild-type vs *elt-7(tm840)*, wild-type vs *elt-2(ca15)* and wild-type vs *elt-7(tm840); elt-2(ca15)* comparisons and related information for intestinal genes that are additively repressed by ELT-2 and ELT-7

Gene	Protein	Exp.	Wt vs *elt-7(tm840)*	Wt vs *elt-2(ca15)*	Wt vs *elt-7(tm840); elt-2(ca15)*	Promoter Size (kb)	Number TGATAA
C25E10.8	Putative secreted TIL-domain protease inhibitor	IE	-0.21, (0.43)	1.65, (1.63E-14)	4.08, (1.79E-81)	0.335	1
cebp-1	bZip domain-containing protein	IP	1.02, (7.73E-17)	1.66, (1.47E-76)	2.63,(4.76E-45)	2.000	4
cpr-1	Cysteine protease	IE	0.63, (4.85E-02)	1.20, (1.17E-04)	2.73, (1.05E-30)	2.000	5
lec-8	Glycolipid-binding galectin	IP	-0.15, (0.31)	0.97, (1.26E-27)	2.40, (1.45E-37)	1.493	2
rop-1	RNAi-binding protein	IP	-0.11, (0.43)	1.07, (3.20E-33)	2.42, (5.94E-119)	1.312	2

Table 4. 14 Log2 fold changes and adjusted p values the wild-type vs *elt-7(tm840)*, wild-type vs *elt-2(ca15)* and wild-type vs *elt-7(tm840); elt-2(ca15)* comparisons for intestinal genes identified as regulated by both ELT-2 and ELT-7 by Sommermann et al. (2010)

Gene	Protein	Exp.	Wt vs *elt-7(tm840)*	Wt vs *elt-2(ca15)*	Wt vs *elt-7(tm840); elt-2(ca15)*
ajm-1	Apical junction molecule	M	NA	NA	NA
cdf-1	Cation diffusion facilitator	IP	-0.05 (0.79)	0.72 (2.38E-11)	0.81 (1.19E-10)
erm-1	Ortholog ERM cytoskeletal linkers	M	0.46 (0.20)	0.93 (4.19E-03)	-0.46 (0.32)
ges-1	Type B carboxylesterase	IE	-0.42 (8.21E-03)	0.17 (7.97E-02)	-0.96 (9.17E-12)
ifb-2	Intermediate Filament	IE	0.38 (1.64E-03)	0.66 (1.54E-11)	-2.10 (1.10E-57)
itr-1	Inositol (1,4,5) trisphosphate receptor	M	0.21 (0.15)	-0.02 (0.93)	-0.78 (1.37E-05)
itx-1	Caspr ortholog	IP	0.26 (3.18E-02)	1.79 (4.60E-92)	0.88 (1.73E-12)
let-767	Steroid dehydrogenase	IP	NA	NA	NA

Table 4.15 Differentially expressed intestinal transcription factors

Gene	Protein	Wt vs *elt-7(tm840)*	Wt vs *elt-2(ca15)*	Wt vs *elt-7(tm840); elt-2(ca15)*
2L52.1	Similarity to GLI-family zinc-finger transcription factors	Down	Down	Down
athp-1	PHD finger transcription factor	Up	Up	Up
bar-1	Beta-catenin	No Change	Down	Down
C01F6.9	C2H2 zinc finger	No Change	Up	Up
C34D10.2	CCCH zinc finger	Up	Up	No Change
C52E12.1	C2H2 zinc finger	No Change	Up	Up
ceh-100	Homeobox transciption factor	Up	Up	Up
ceh-37	Homeodomain transcription factor	No Change	Down	Up
cey-1	Cold-shock/Y-box domain	Up	Up	Up
cey-4	Cold-shock/Y-box domain	Up	Up	Up
dhhc-1	DHHC zinc finger	Down	Up	Up
eea-1	C2H2 zinc finger	Up	No Change	No Change
ets-4	Ortholog of human SAM pointed domain containing ETS transcription factor	No Change	Up	Down
F13C5.2	Bromodomain containing	Up	Down	No Change
F17C11.1	bZIP transcription factor	No Change	No Change	Up
F21A9.2	C2H2 zinc finger	No Change	Up	Down
F26H9.2	RPEL domain containing	No Change	Up	Up

F55B11.4	Zinc Finger - C2H2 - 1 finger	Down	Up	Down
F57A8.1	Winged Helix	No Change	No Change	No Change
fkh-7	Forkhead transcription factor	Up	Up	Up
hmg-1.1	HMG transcription factor	Up	No Change	No Change
hmg-11	HMG transcription factor	Up	No Change	No Change
hmg-12	HMG transcription factor	No Change	Up	Up
irx-1	Homeodomain transcription factor	Up	No Change	Up
K09A11.1	Zinc Finger - BED	No Change	Down	Down
K11D12.12	C2H2 zinc finger	No Change	Down	Down
lfi-1	Large coiled-coil protein	Up	Down	Down
mbf-1	Multiprotein bridging factor transcriptional coactivator	No Change	Up	Up
mdl-1	bHLH transcription factor	Up	No Change	Up
mgl-2	bZIP transcription factor	Down	Down	Down
moe-3	CCCH zinc finger	No Change	Up	No Change
mxl-1	bHLH transcription factor	No Change	Down	Down
mxl-2	bHLH transcription factor	Up	Up	No Change
mxl-3	bHLH transcription factor	Up	Up	Up
nhr-100	Nuclear hormone receptor	Down	Down	Down
nhr-101	Nuclear hormone receptor	Down	Down	Down
nhr-105	Nuclear hormone receptor	Down	Down	Down
nhr-108	Nuclear hormone receptor	Down	Down	Down
nhr-139	Nuclear hormone receptor	No Change	Down	Down
nhr-140	Nuclear hormone receptor	No Change	Down	Down
nhr-143	Nuclear hormone receptor	Down	Up	Down
nhr-153	Nuclear hormone receptor	Down	No Change	No Change
nhr-154	Nuclear hormone receptor	Down	No Change	No Change
nhr-16	Nuclear hormone receptor	No Change	Down	Down
nhr-161	Nuclear hormone receptor	Down	Down	Down
nhr-170	Nuclear hormone receptor	No Change	Up	Down
nhr-183	Nuclear hormone receptor	No Change	Up	Up
nhr-19	Nuclear hormone receptor	Down	Down	Down
nhr-201	Nuclear hormone receptor	No Change	Up	Up
nhr-203	Nuclear hormone receptor	No Change	Up	Up
nhr-206	Nuclear hormone receptor	Down	No Change	Down
nhr-208	Nuclear hormone receptor	No Change	Down	Down
nhr-209	Nuclear hormone receptor	Down	Down	Down
nhr-210	Nuclear hormone receptor	No Change	Up	No Change
nhr-212	Nuclear hormone receptor	No Change	Up	Up
nhr-22	Nuclear hormone receptor	No Change	Down	Down
nhr-232	Nuclear hormone receptor	No Change	Down	Down
nhr-3	Nuclear hormone receptor	Down	Up	Up

nhr-32	Nuclear hormone receptor	Down	Down	Down
nhr-4	Nuclear hormone receptor	Down	Down	Down
nhr-41	Nuclear hormone receptor	Up	Up	No Change
nhr-42	Nuclear hormone receptor	Down	Down	Down
nhr-44	Nuclear hormone receptor	No Change	Down	Down
nhr-55	Nuclear hormone receptor	No Change	Up	Up
nhr-57	Nuclear hormone receptor	No Change	No Change	Up
nhr-61	Nuclear hormone receptor	Up	Up	Up
nhr-63	Nuclear hormone receptor	No Change	Up	Up
nhr-78	Nuclear hormone receptor	No Change	Down	No Change
nhr-8	Nuclear hormone receptor	No Change	Down	Down
nhr-80	Nuclear hormone receptor	Up	Down	Down
nhr-86	Nuclear hormone receptor	No Change	Up	Up
nhr-88	Nuclear hormone receptor	Up	No Change	No Change
nhr-90	Nuclear hormone receptor	No Change	Down	Down
nhr-98	Nuclear hormone receptor	No Change	Up	Up
nhr-99	Nuclear hormone receptor	No Change	Up	Down
pal-1	Caudal ortholog, Homeodomain transcription factor	No Change	Down	Down
peb-1	FLYWCH zinc finger	No Change	Up	Up
php-3	Posterior Hox gene paralog	No Change	No Change	Down
pop-1	TCF/LEF transcription factor	Up	No Change	No Change
pqm-1	C2H2 zinc finger	No Change	Up	Down
sdc-2	Unknown	No Change	Down	Down
sea-2	Zinc finger transcription factor	Up	No Change	No Change
sma-2	Smad	No Change	Down	Down
uaf-2	CCCH zinc finger	Up	Up	Up
ubxn-1	C2H2 zinc finger	Up	Up	Up
unc-130	Forkhead transcription factor	Up	Up	Up
Y48C3A.12	TSC-22/dip/bun	Up	Up	Up
Y57A10A.31	C2H2 zinc finger	No Change	Up	Up
Y79H2A.3	C2H2 zinc finger	Up	Up	Up
zip-2	bZIP transcription factor	Up	Up	Up
zip-3	bZIP transcription factor	Up	Up	No Change
zip-5	bZIP transcription factor	No Change	Down	Down
ztf-13	Zinc finger transcription factor	No Change	Up	Up
ztf-22	Zinc finger transcription factor	Up	Up	No Change
ztf-8	Zinc finger transcription factor	No Change	Up	Up