UNIVERSITY OF CALIFORNIA
RIVERSIDE

Robustness of Gut Development in the Nematode, *Caenorhabditis elegans*

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Cell, Molecular, and Developmental Biology

by

Hailey Heiyan Choi

December 2017

Dissertation Committee:

Dr. Morris F. Maduro, Chairperson
Dr. Patricia Springer
Dr. Venugopala Reddy Gonehal

The Dissertation of Hailey Heiyan Choi is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

First and foremost, I would like to thank my graduate advisor Morris Maduro for inviting me to join his lab five years ago. By extending that invitation he not only gave me an opportunity to further my graduate studies, but also provided a home for me, a nurturing environment for me to learn and grow as a scientist and as a person. Thank you for your guidance, for encouraging me to learn new skills, for all the life advice, and for teaching me by example how to be a good educator. I would also like to thank Gina Broitman-Maduro. Her guidance and support was vital during my graduate career. Melissa was right; Gina is the "heart" of the lab. Thank you for all that you did for me, and for all that you do for the lab. Thank you for teaching me laboratory techniques, for being patient with me, and for your friendship. I want to thank the current and past undergraduates that I have worked with, Francisco Carranza, Christian Turner, Kollan Doan, and Raudel Acosta, for all their assistance and the fun spirit they brought to the lab.

I would like to thank my Committee members Dr. Patricia Springer and Dr. Venugopala Gonehal for being a part of my committee and for their guidance and support since my very first Annual Research Progress Evaluation meeting. I would not have pursued a graduate degree if it wasn't for the guidance and encouragement of Dr. Jeffery Bachant, Jeffery Julius, and Dr. Hovik Gasparyan during my undergraduate studies. Thank you for letting me join your lab, for helping me realize how much I like doing research, and helping me get into graduate school. Hovik was the graduate student that I worked with the most as an undergraduate. He was the person that pushed me to apply to

the CMDB program at UCR. He started off as being a great mentor, and as the years went by he became one of my closest friends.

I would like to thank David Carter from the Microscopy Core, Holly Clark and Clay Clark from the Genomics Core. I would like to also acknowledge our collaborators in the Yanai lab for helping me troubleshoot the CEL-seq protocol. The text of this dissertation, in part, is a reprint of the material as it appears in Partially compromised specification causes stochastic effects on gut development in *C. elegans*, May 2017. The co-author Morris Maduro listed in that publication directed and supervised the research which forms the basis for this dissertation. Figures from Chapter 1 were taken from the following resources with permission: Figure 1-1 from Chalancon et al.,2012, Figure 1-5 from WormAtlas, (Altun 2005), Figure 1-6 WormBook, (Zarkower 2006), and Figure 1-7 from Maduro et al. 2015.

My graduate experience would not have been as enjoyable if I was not able to work alongside my friends on the second floor of the Genomics Building. Thank you, Dr. Marcella Teixeira, for being the older sister that I never had, for your witty banter, and for picking me up when I was having a bad day in lab. I cannot think of my lab without thinking about the neighboring Le Roch lab. I feel like I am an honorary member of your lab. I want to thank Dr. Serena Cervantes, Jacque Prudhomme, Michael Lee, Aster Escalante, and Chris Connor for sharing my graduate experience with me. I will always remember and cherish our side conversations in-between your time points and as I worked at the dissecting scope. I also want to thank Dr. Evelien Bunnik for being a great

mentor to me, and for her patience as she taught me how to do my own bioinformatics analysis.

Next, I want to thank all my friends from graduate school: Maggie Lu, Gayani Batugedara, Kevin Rodriguez, Lily Maxham, Graciel Diamante, Dustin Brewton, Teresa Wen, Oscar Gonzalez, and my CMDB cohort. Thank you for helping me through all the tough times that came during graduate school, and for all the memories that we've made together over the years. I wouldn't have made it through without the encouragement and support of every one of you.

Next, I want to thank all my friends outside of graduate school: Elissa Liong, Justine Tran, Cindy Tran, Adela Ng and Anthony Choy. Thank you for the support and for believing in me. I appreciated you all taking the time to check up on me frequently, and your words of encouragement helped me throughout graduate school.

To my cousins, Doris, Edward, Tony, Jim, Samantha, Sherri, Gene, Sara, Timmy, Rachel (陳華騫), Phoebe, and Gloria, thank you for always letting me know that you're proud of me, for being understanding about me missing family events for graduate school, but also for keeping me grounded throughout, and reminding me of the importance of life outside of graduate school. Thank you for all the love and support and for believing in me. I wouldn't be where I am today without your help, guidance, support, and love. I want to especially thank Doris for always being there for me ever since I was a little kid, whether it was for school or something else. You were always there, and I'll always remember that. To Rachel, even though you were a year younger, I always looked

up to you. When we were growing up you were smart, hardworking, and motivated. You set the bar, and it inspired me to be a better student.

To my dad 蔡邦忠, my mom 蔡黃笑蘭, and my brother, Kai, words cannot express how grateful I am for all that you have done for me, and for the sacrifices you've had to make, so that I can be where I am right now. Thank you for doing your best to give me whatever I needed, and for always putting me first. Thank you for teaching me the value of hard work. None of this would have been possible without your love and support. My goal in life is to make you proud.

DEDICATION

I dedicate this dissertation to my Dad 蔡邦忠, My Mom 蔡黃笑蘭, my brother Kai, my cousin Doris, my Grandmother 鄺彩屏, and my Uncle 黃漢傑. For believing in me when I didn't believe in myself, and for your support throughout my academic journey.

ABSTRACT OF THE DISSERTATION

Robustness of Gut Development in the Nematode, *Caenorhabditis elegans*

by

Hailey Heiyan Choi

Doctor of Philosophy, Graduate Program in Cell, Molecular, and Developmental Biology
University of California, Riverside, December 2017
Dr. Morris F. Maduro, Chairperson

Robust gene expression is crucial for proper embryonic development, and yet it is subject to extrinsic and intrinsic sources of noise. We investigated mechanisms by which the *C. elegans* embryo buffers these stochastic variations in gene expression during endoderm specification. In this system, the factors that contribute to the specification and differentiation of the progenitors of intestine are known, allowing us to probe effects of upstream stochasticity on known downstream events. The ability of this system to buffer noise during early embryogenesis was tested by targeting early acting inputs in the endoderm gene regulatory cascade and measuring its effect on tissue specification and commitment to differentiation. To study the downstream effects of animals that were on the verge of nearly missing endoderm differentiation, we utilized a series of allelic mutants. These mutant strains specify varied numbers of gut cells, between 0%-100% of the time. Using these strains, we are currently investigating E lineage phenotypes during early and mid-embryogenesis, as well as the young adult stage. We have screened for maternal factors to identify potential modifiers of gut specification, and showed that

specification of endoderm is not an all-or-none event at the level of E blastomere. Instead, it can be displaced as a binary choice at later time points within the E lineage – a phenomenon that we are calling "stratified specification". Last, we find that fully differentiated intestine retains a memory of their “near-miss” specification, which results in abnormal adult phenotype, due to differential expression of genes in the intestine. Thus far, we have found that several diverse maternal pathways, including transcriptional regulation, metabolism, and apoptosis influence the ability of embryos to make gut, suggesting that embryonic gene expression is highly sensitive to many sources of variation. Collectively our data show that gut specification is not an "all-or-none" event, and that if specification does not occur properly, defects can manifest into adulthood. The results have implications in other developmental pathways in *C. elegans* and in any metazoan system in which gene regulatory networks drive important developmental events.

Table of Contents

Chapter 1

Introduction

Chapter 2

An RNAi screen for modifiers of gut specification identifies pathways in metabolism and gene expression

Chapter 3

Partially compromised specification causes stochastic effects on gut development in *C. elegans*

List of figures

Chapter 3: Partially compromised specification causes stochastic effects on gut development in *Caenorhabditis elegans*

List of figures

Chapter 4: Adult Survivors of Hypomorphic Specification have Defects in Stress Response and Metabolism

List of tables

Chapter 2: An RNAi screen for modifiers of gut specification identifies pathways in metabolism and gene expression

List of tables

Chapter 3: Partially compromised specification causes stochastic effects on gut development in *Caenorhabditis elegans*

List of tables

Chapter 4: Adult Survivors of Hypomorphic Specification have Defects in Stress Response and Metabolism

CHAPTER 1

INTRODUCTION

1.1 Stochastic Gene Expression

Stochasticity, or the variability of gene expression due to random events, can be observed all around us. It is present from the minute differences in the facial features of identical twins to the emergence of phenotypically distinct subgroups in an isogeneic population of cells. Stochasticity has been thought to have detrimental effects on cellular processes (McAdams, Arkin 1997; Elowitz et al. 2002; Kaern et al. 2005; Raj, van Oudenaarden 2008). It also has profound effects on stem cell differentiation (Chang et al. 2008), cancer development (Brock, Chang, Huang 2009), and responses to apoptosis-inducing factors (Bastiaens 2009; Spencer et al. 2009). Stochastic mechanisms can also be advantageous to the viability of the cell, through adaptations that allow the cells to survive in the presence of perturbations, such as bet-hedging in bacterial phenotypes (Kussell et al. 2005; Fraser, Kaern 2009).

Gene expression is inherently stochastic and subject to variability due to several factors, including intrinsic noise, such as the abundance and availability of transcription factors and extrinsic environmental fluctuations (Fig. 1-1). Thus, randomness in

transcription and translation leads to cell-to-cell variation in mRNA and protein levels (Raj, van Oudenaarden 2008; Chalancon et al. 2012). Stochastic gene expression occurs in both prokaryotes (Elowitz et al. 2002; Ozbudak et al. 2002; Kaern et al. 2005) and eukaryotes (Ozbudak et al. 2002; Raser, O'Shea 2004; Kaern et al. 2005). In cultured, undifferentiated cells, this can lead to asymmetric cell divisions and spontaneous differentiation, even if cells are exposed to the same environment. The stochastic cell fate decisions can generate non-genetic cellular diversity, which may be critical for metazoan development. Additionally, microbes may also utilize such stochastic variations in gene expression to optimize resource utilization and survive in a fluctuating, frequently stressful environment (Balazsi et al., 2011; Ozbudak et al., 2002).

The presence of both intrinsic and extrinsic noise were experimentally proven in *Escherichia coli* and *Bacillus subtilis*. The two-reporter method was used to study the noise that accompanies gene expression in *E. coli*. The levels of cyan and yellow fluorescent protein expression were quantified from identical promoters on the same chromosome. The results showed that randomness in gene expression resulted in increased noise in protein levels of an isogenic population of *E. coli*. Relative to total noise, the contribution of extrinsic and intrinsic components varied with expression level (Elowitz et al. 2002). Other studies in *B. subtilis* showed that the dominant noise in protein levels is due to the various steps of transcription (McAdams, Arkin 1997; Thattai, van Oudenaarden 2001; Ozbudak et al. 2002).

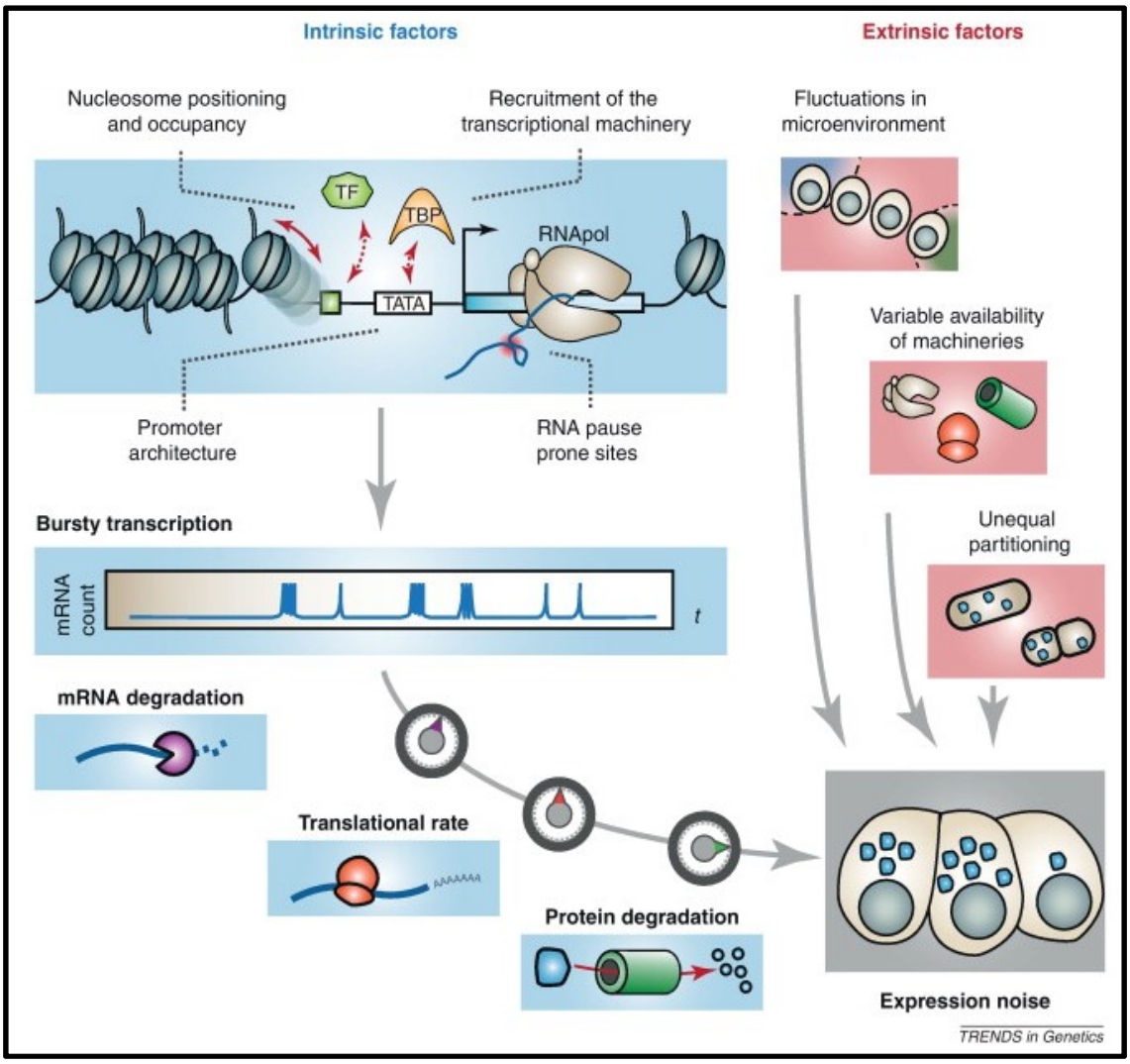


Figure 1-1: Stochastic noise is due to intrinsic or extrinsic factors. Intrinsic factors include the structure of chromatin, recruitment transcriptional/ translational machinery, the rate of transcription/translation, rate of mRNA/ protein degradation. Extrinsic factors encompass the variation within the cellular environment, presence of essential factors, and unequal portioning of products during cell division.

This figure is a reprint from Trends in Genetics, Vol 28 Issue 5, Chalancon et al., Interplay between gene expression noise and regulatory network architecture, Pages 12., Copyright (2012), with permission from Elsevier.

1.1.1 Extrinsic Noise

Animal cells undergo many rounds of division and are exposed to various environmental conditions or extrinsic noise, which may result in phenotypic differences among the cells. Extrinsic noise is subdivided into two categories: noise that is specific to a particular gene or pathway (Raser, O'Shea 2004; Pedraza, van Oudenaarden 2005; Raser, O'Shea 2005) and global noise, which accounts for the fluctuation of factors that affect all gene expression.

In the wild, single-celled species are exposed to constant changes in temperature, pH, and availability of nutrients. For cells to thrive in those dynamic environments, they have evolved to take advantage of the stochasticity which accompanies gene expression. In order for a population of cells to increase their chance of survival in an ever changing physical environment, individual cells will undergo stochastic gene expression to acquire multiple phenotypes, to ensure that a subpopulation of the cells will be able to overcome any changes that might come their way (Balaban et al. 2004; Levy, Ziv, Siegal 2012). In an experiment where two mutant yeast strains that adapt to their environment at different rates were exposed to different conditions, it was evident that the population of cells that were fast-switching had a fitness advantage and outgrew the slow-switching cells, when environmental changes occurred rapidly. Conversely, in an environment that hardly fluctuated, the slow-switching cell population outgrew the fast-switching cells (Acar, Mettetal, van Oudenaarden 2008). Aside from environmental changes, varying concentrations of molecules and chemicals within the environment also contribute to stochastic noise. The uneven distribution of material across a population of cells can

result in shifts of molecules, which are used by cells as positional information and for cell-cell signaling (Athale, Chaudhari 2011).

During mitotic cell division, chromosomes will duplicate, and the cell splits to form two genetically identical daughter cells. During the anaphase stage, mitotic spindles help guide the duplicated chromosomes to opposite sides of the mother cell. The cell then undergoes telophase and cytokinesis and gives rise to two genetically identical daughter cells that display cellular variability. During cytokinesis, only the chromosomes are ensured to be evenly divided among the daughter cells, but the number of macromolecules and organelles that each daughter cell receives is potentially unequal. This source of inheritance noise is the result of uneven partitioning of cellular components between the daughter cells, and can lead to extrinsic differences within the cell population (Huh, Paulsson 2011). The random assortment of cellular components may generate significant cell-to-cell variation in many different processes. (Johnston et al. 2012) showed that uneven distribution of mitochondria between two daughter cells can lead to skewed level of cellular energy (ATP), which causes downstream effects on ATP dependent processes. ATP is necessary for both transcription and translation, and the skewed levels of ATP among a population of cells can cause differential gene expression (Shahrezaei, Ollivier, Swain 2008). This suggests that extrinsic noise can synergize with intrinsic noise and generate greater variation within an isogenic population.

1.1.2 Intrinsic Noise

Sources of intrinsic noise include factors that affect gene expression directly at a molecular level. This innate noise randomly arises from fluctuations of molecules necessary for transcription and translations. Gene expression is complex and requires multiple sequential steps. The start of transcription occurs when a transcription factor (TF) binds to cis-regulatory elements within or near the promoter of the gene of interest. This binding event is the result of both Brownian motion within the cell (Saffman, Delbruck 1975) and the random encounters between the molecules, some of which are of low copy number. This is followed by synthesis and processing of mRNAs, export of the mRNA to the cytoplasm, translation, protein folding, and ultimately degradation of these products. Each of these steps is susceptible to variations, and together they make the biochemical processes that regulate transcription and translation inherently stochastic (Kaern et al. 2005).

Aside from the abundance of readily available molecules, the size of the cell is also a contributing factor to stochastic gene expression. The "finite- number effect" is used to describe the accumulation of molecular-level noise in cellular regulation. There is a fundamental relationship between size and noise: molecular-level noise increases when the area containing the molecule decreases (Kaern et al., 2005; Kampen, 1992). This relationship can be defined, using a molecule that moves freely between the nucleus and cytoplasm. Within the cell, the concentration of molecules in both the nucleus and cytoplasm is roughly the same when the cell reaches a state of equilibrium, however, due

to the size difference between the nucleus and cytoplasm, the movement of molecule across the membrane will cause different effects. For example, if there 25 molecules in the nucleus and 100 times more molecules in the cytoplasm, the addition of 5 molecules to the nucleus will cause greater change, than that same addition to cytoplasm. The differential effects seen in gene expression can be the result of varying numbers of molecules in the nucleus versus the cytoplasm (Kaern et al. 2005).

Other factors that contribute to gene expression noise are created during the transition from transcription to translation. In one example, a cell that contains high levels of expressed mRNA and protein will display small fluctuation of noise. In a different cell, a fivefold decrease in transcriptional rate, it will lead to a lower amount of expressed mRNA, leading to decrease in the amount of protein synthesized, and will lead to fluctuation in protein abundance between cells. However, in an event where there was a fivefold decrease in transcription, but the rate of translation is increased to mimic the overall production of protein in the earlier example, the increase in gene expression noise in spite of large protein abundance is associated with increased changes in mRNA level, which synergistically increase the fluctuation in protein synthesis (Ullah, Wolkenhauer 2010). When the expression of two genes on average is equal, but varies in the efficiency of translation and mRNA (transcript) number, the one with the higher translational efficiency and lower mRNA abundance will display greater fluctuations in protein concentration. Since the average number of protein molecules used for translation and the cell volume are kept fixed, the variability in gene expression noise is due to increased fluctuations in mRNA abundance, leading to greater fluctuations in the

rate of protein synthesis. Depending on the rate of translational efficiency and the relative change between rates of transcription and translation, it can result in synergistic fluctuation of gene expression noise (Blake et al. 2003); (Ozbudak et al. 2002); (Raser, O'Shea 2004). Although stochastic gene expression can be attributed to levels of mRNA and/ or proteins, the variation in gene expression is impacted more by mRNA level than protein levels, due to the more stable nature of proteins (Newman et al. 2006).

1.2 Developmental Robustness

Higher organisms start as a single cell, with some vertebrate species looking strikingly similar at the early stages of embryogenesis. However, at the end of embryonic development, each species looks distinctively different. Although members of the same species are not phenotypically identical, they carry features that clearly distinguish them from other species. Robustness plays an essential role in development, by ensuring that key features develop in a specific way. Robustness is a genetic term in biology. It can be a quantitative trait and can be used as a relative measure, to compare the same phenotype across different incoming environmental variations (Felix, Barkoulas 2015). The role of robustness of gene regulation has become the topic of interest in previous years, particularly in the role it plays in multifactorial human diseases (Kitano 2004; Gibson 2009).

Developmental robustness is defined as the ability of an embryo to develop normally in the presence of variation. Over the course of its development, an organism is equipped to withstand many sources of variability from both its internal and external

environment. Evolution has selected for this trait, to allow development of individuals that belong to the same species to produce a stereotyped phenotypic outcome (Fusco, Minelli 2010). An organism is often faced with environmental perturbations or experiences limiting amounts of molecular molecules, but this may or may not result in a change in its phenotypic outcome. The lack of phenotype is associated with the term canalization, coined by C.H. Waddington. Canalization is synonymous with developmental robustness, where both define specific trajectories in cell fate choice during a tightly regulated developmental process in a spatiotemporal manner. Genetic variation or environmental noise is "buffered" up to a certain threshold of change and will not affect development. However, even the slightest amount above the acceptable threshold will cause cells to change their cell fate trajectory (Waddington 1942).

The mechanisms of robustness have been tested extensively in the *Drosophila*, in experiments that exposed the animals to high temperature (heat), high salt, and through mutations in regulatory genes (Waddington 1959; Rutherford, Lindquist 1998; Waddington 1953; Scharloo 1991). Robustness can be broken down into the adaptive and intrinsic types. Adaptive robustness occurs when an organism has evolved to reduce the effects of mutations, by preventing phenotypic change. However, partial adaptive robustness also has its advantages: by accumulating mutations that do not exhibit an outward phenotype, the organism may become pre-adapted to various changes which can help them to evolve and survive under unfavorable conditions (Masel, Siegal 2009). The other form of robustness is best described as an intrinsic quality, such as the mechanisms

of robustness that are built into gene regulatory networks, to ensure timely activation of essential developmental genes.

The fundamental components that make up a robust system are derived from the combination of: system control in terms of feedback loop, a failsafe mechanism in terms of redundancy, and modularity (Kitano 2004). Feedback loops can be positive or negative self-regulators, where a negative regulator will decrease the original input signal and a positive regulator will multiply it (Freeman 2000). A balance between positive and negative regulators is necessary to generate a stable yet sensitive circuit, and plays a role in signal transduction and transcription. Redundancy can be either structural (the wiring of multiple transcription factors to one gene) or functional (the duplication of essential genes). By allowing multiple transcription factors to bind to the promoter of the same gene, it ensures that the gene will be activated, even if one or more of the transcription factors are missing. The duplication of essential genes also serves a similar function, where the accumulation of mutations in one copy of the gene can be compensated for by the duplicated gene and will lead to little phenotypic effect. In many organisms, development can be broken down into modules. These modules can be physical, functional, spatial, or temporal, where individual "module" or "unit" has a role in a specific development plan, such as being a part of a metabolic networks, signal transduction and developmental regulatory networks (West-Eberhard 2003; Kitano 2004). While individual modules function independently to execute a set of plans, robustness is ensured when modules reinforce each other during crucial developmental time periods (Benitez, Alvarez-Buylla 2010), (Bolker 2000; Sharma et al. 2016)

(Klingenberg et al. 2001). Individually, each module is effective at maintaining robustness by limiting changes and mutations to a specific region, thus minimizing the effects. However, the modular trait also allows for the emergence of diverse phenotypes through changes in interaction between modules and evolution of modules through accumulation of mutations (Kitano 2004).

To reach a robust state, a system needs to obtain the ability to buffer various levels of changes that it will receive. Proteins have adapted to tolerate incorporation of different amino acids, metabolic networks have evolved to bypass missing intermediate steps, and gene regulatory network can remain functional even when some of the transcriptional wiring is removed (Rennell et al. 1991; Edwards, Palsson 2000; von Dassow et al. 2000; Bhattacherjee, Biswas 2010). Buffering mechanisms can be simple, such as the ability of *Saccharomyces cerevisiae* to buffer changes in varying temperature (Levy, Siegal 2008; Balazsi, van Oudenaarden, Collins 2011), and its ability to perform chemotaxis over a range of chemo-attractant concentrations (Barkai, Leibler 1997; Alon et al. 1999; Yi et al. 2000). They can also be quite complex such as the buffering mechanism that accompanies the segmental polarity genes of *D. melanogaster* (von Dassow et al. 2000; Ingolia 2004) and the co-opted robustness that leads to diseases such as cancer and diabetes (Kitano 2003; Kitano et al. 2004).

Robustness in a biological system comes in many layers and levels, and is put in place through hierarchically structured feedback loops that rely heavily on both structural and functional redundancy (Keller 2002). The way an organism can combine and gain

the beneficiary effects of those features is through the use of gene regulatory networks (GRN) (Wagner 1996; Siegal, Bergman 2002; Bergman, Siegal 2003). Where a mutation in one pathway occurs can be compensated for by other processes in the system (Wagner 2005). This dissertation examines the effects of a gene regulatory network when contributors of robustness are removed from a highly robust gene regulatory network.

1.3 Gene Regulatory Network

The genetic blueprint of most organisms is made up of Deoxyribose Nucleic Acid (DNA), which contains the base sequences that will ultimately be transcribed to RNA and translated to protein. Imbedded in the sequences are a series of genomic regulatory codes, which are self-guided instructions to execute the expression of genes (Istrail, Davidson 2005). These instructions come in the form of molecular regulators that work in conjunction to maintain the expression of genes in the proper spatiotemporal matter. Such regulators can be made up of DNA, RNA, or protein. However, the main regulators of GRNs are Transcription Factors (TF), a type of DNA binding protein which can have a positive or negative impact on gene expression (Jeziorska, Jordan, Vance 2009). Together, a set of instructions, TFs, and sequences of protein coding genes can be defined as a GRN. These networks can be further subdivided into two main categories: gene expression networks derived from transcript profiling and transcription factor networks derived from identification of DNA binding sites for each transcription factor (Dewey GT, 2013). The focus of this dissertation will be on the latter type of GRN. Formation

and timely activation of these networks is an essential component in the differentiation of cells, tissues, and organs during development.

1.3.1 Properties

Two genes are connected if the expression of one gene modulates the expression of another gene. The interaction between those two genes comes from the interplay between binding of Transcription Factors (TFs) to a specific short sequence within *cis*-Regulatory Modules (CRMs). The regulation of gene expression comes from the CRMs, which are information processing centers embedded into the genomic sequence (Davidson 2001). CRMs are DNA sequences that are several hundred base pairs, located either upstream, downstream, or within the introns (Jack et al. 1991; Zimmerman et al. 1994; Bien-Willner, Stankiewicz, Lupski 2007; Tumpel et al. 2008). While a single CRM can manage the expression of many genes, a single gene can also have many CRMs (Ben-Tabou de-Leon, Davidson 2007; Teif 2010). As they are named, CRMs are made up of individual units that work independently of one another. They generally contain multiple binding sites, ranging from 4-8 (Arnone, Davidson 1997). Where a given TF binds to multiple binding sites (homotypic CRMs), the increased sensitivity of a factor is attributed to how many sites are present (Howard 2004). The interaction of TFs and CRMs follows a set of logic rules: AND logic, OR logic, or NOT logic. When two different transcription factors need to be present at the same time and interact with one other, it is defined by AND logic- factor A AND factor B need to bind to the CRM to activate gene expression. If the binding of either Factor A OR factor B is sufficient to

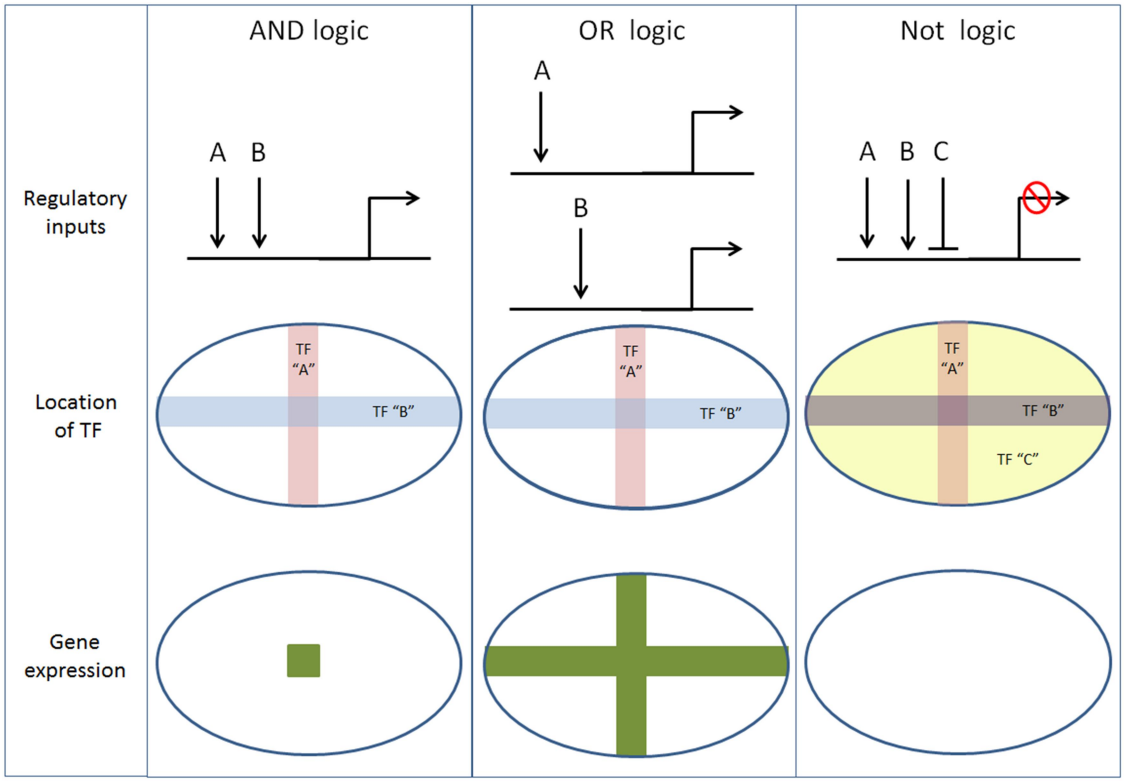


Figure 1-2: The three logic rules governing gene expression are "AND", "OR", and "NOT" logic. "AND" logic requires two separate transcription factors (A and B) to be present in the same spatiotemporal manner in order to drive gene expression. "OR" logic requires either of the transcription factors (A or B). "NOT" logic occurs when in the presence of a separate transcription factor (C) gene expression does NOT occur, even in the presence of transcriptional activator A and B.

drive gene expression, then it follows OR Logic. Lastly, if the binding of factor C represses gene expression, then it defined as NOT logic (Fig. 1-2).

CRMs are not driven by positive or negative inputs alone. During the developmental process, both positive and negative inputs work in conjunction to ensure that gene expression occurs at the right place and time. An example of the cooperative nature of positive and negative inputs on CRMs can be seen in the expression of the even-skipped gene of *Drosophila melanogaster.* During embryonic development, nuclei in the embryo are replicated without cytokinesis, which forms a syncytial blastoderm, where one cytoplasm is shared by all existing nuclei. The syncytium allows maternal and zygotic factors to diffuse throughout the embryo, setting up the various gradients before the embryo is cellularized. This results in an anterior to posterior gradient of the transcription factors Biciod and Hunchback, and a posterior to anterior gradient of Caudal. The three transcription factors contribute to the activation of gap genes: *giant*, *kruppel*, and *knirps* (Gilbert 2000). Expression of the even-skipped gene is highly robust, and forms a pattern of seven thin stripes. The expression of each stripe is orchestrated by CRMs. The best understood module is of even-skipped stripe 2. The precise expression of even skipped stripe 2 is generated by transcriptional activators: Bicoid and Hunchback, as well as transcriptional repressors: Giant and Krupple (Fig. 1-3). The anterior gradient of Bicoid and Hunchback can activate the expression of even- skipped 2, however the Giant and Kruppel are necessary to form the anterior and posterior boundaries of the stripe (Small, Blair, Levine 1992; Arnosti et al. 1996; Gray, Levine 1996).

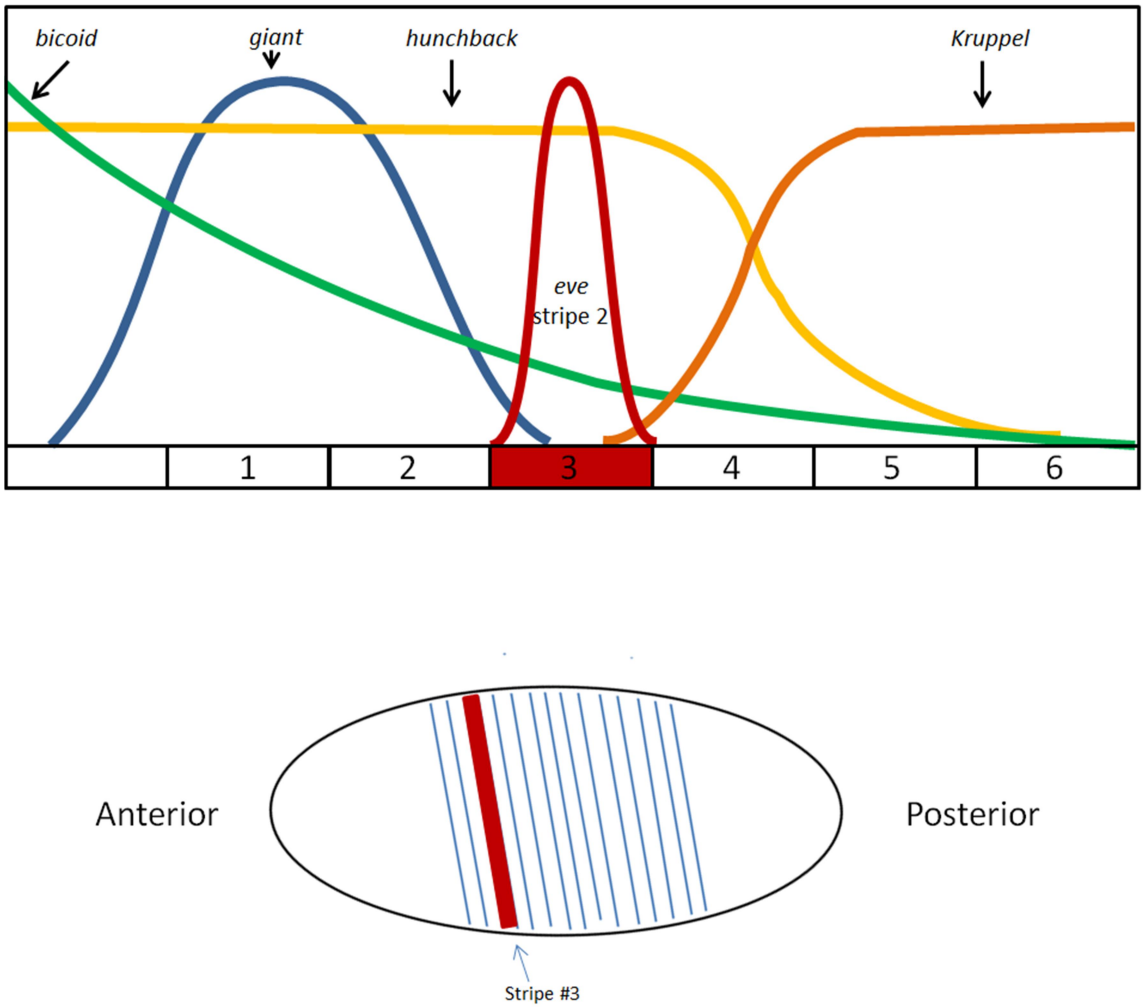


Figure 1-3: The expression of even-skipped at third parasegment is regulated by the enhancer region of *eve* stripe 2. The enhancer region of eve stripe 2 contains binding sites for *biciod, giant, hunchback*, and *Kruppel*. Expression patterns of the four transcription factors have already been set during early embryogenesis. Activation of *eve* stripe 2 is driven by transcription factors *biciod* and *hunchback*. The borders of *eve* stripe 2 are determined by the *giant* (anterior boarder) and *Kruppel* (posterior border).

1.3.2 Redundancy

Gene regulatory networks can ensure robust activation of their downstream targets through multiple redundant factors. To ensure timely activation of essential genes, GRNs have evolved to contain redundant interactions of transcription factors to individual genes (Macneil, Walhout 2011). In *Drosophila melanogaster*, researchers have identified redundant wiring at the level of enhancers. They identified "shadow enhancers" that are located far from the transcriptional start site and are functionally redundant to "primary enhancer" in several genes. (Hong, Hendrix, Levine 2008; Frankel et al. 2010; Cannavo et al. 2016). In several genes necessary for proper trichrome (hair-like) formations, *shavenbaby,* was found to be regulated by both primary and shadow enhancers (Frankel et al. 2010). In a mutant background where the function of one of the primary enhancers was disrupted in combination with the removal of a shadow enhancer, there was dramatic reduction of trichrome formation. When animals were grown at varying temperatures, disrupting transcription factor activity, also resulted in decreased number of trichrome.

Robustness of gene expression can also be derived from the ability of different transcription factors within the same family binding to the same *cis*- regulatory DNA element, as well as a different family of transcription factors interacting with the same enhancer of a CRM. Both mechanisms can be seen in *C. elegans*. The pair of transcription factors FLH-1 and FLH-2 are both capable of binding to the promoter region of a group of microRNA genes (Ow et al. 2008). Loss-of-function in either one of

these transcription factors did not show any effects. However, a double mutant resulted in increased larval lethality. The endoderm GRN in *C. elegans* is a great example of a highly robust GRN, which results in a tightly regulated process controlled by a highly-controlled network of transcription factors (Maduro et al. 2007). At the top of the GRN is the maternal factor SKN-1, which activates a set of GATA-type transcription factors. Together, the transcription factors function to activate the gut master regulator *elt-2* at the correct spatiotemporal context, to ensure proper gut gene programming. Interestingly, in an isogenic population of animals that contains the same mutation that effects the endoderm gene regulatory network, some animals can specify intestine while others cannot. This phenomenon, where only a subgroup of isogenic animals exhibits a phenotype is defined as penetrance. Varying levels of penetrance can be seen when the endoderm GRN is mutated. When the function of *skn-1* is removed, it results in ~20% of embryos specifying a gut, but a mutation in *end-3* can yield ~95% of embryo making an intestine. Work done by Raj et al. (2010) elucidated the mechanism that allows for partial penetrance of certain mutations. They looked at the level of *end-1* mRNA in a *skn-1* background, and saw that the levels varied from embryo to embryo. By quantifying the amount of *end-1* mRNA within each cell, they concluded that a threshold amount of *end-1* mRNA transcript is necessary for activation of *elt-2*. Thus, to guarantee a robust expression of *elt-2* across all wild-type animals, the endoderm GRN is wired in a way where multiple transcription factors activate the same essential gene, to ensure the transcript level is well above the threshold necessary for activation.

Aside from duplicating *cis*-regulatory regions and generating redundant wiring of transcription factors, robustness of GRN is in part due to genetic redundancy. Genetic redundancy is when two or more genes serve the same function, and the absence of one of these genes will have little phenotypic effects (Cadigan, Grossniklaus, Gehring 1994; Gibson, Spring 1998; Winzeler et al. 1999). By having two almost identical genes operating in the same regulatory network, it provides a built-in backup for the network, in case one of the gene functions gets disrupted (Tautz 1992; Wilkins 1997). It has been shown in yeast, that duplication of genes is attributed to robustness against null mutations, but the mechanism of genetic robustness is not based solely on gene duplication alone; it is also through the function of duplicated genes in a compensatory regulation network (Gu et al. 2003).

Genetic redundancy is a predominant feature in the *C. elegans* endoderm GRN, where the four GATA-type genes are made up of two pairs of redundant genes. The *med-1* and *med-2* genes encode transcription factors that are high up in the endoderm GRN and are essential for both mesodermal and endodermal cell fates (Maduro 2006). Both *med* genes are expressed during the four-cell stage of embryogenesis. The absence of these genes results in embryonic arrest, as well as the failure to specify posterior pharynx in 100% of embryos and intestine in ~50% (Maduro et al. 2007). The other GATA- type genes *end-1* and *end-3* have redundant functions in the specification of the endoderm. In an *end-3* mutant, the number of gut cells in each embryo varies, with 5% of embryos undergoing embryonic arrest. In the absence of either one of the *end* genes, the development of the embryos are not dramatically affected, while a double mutant will

lead to 100% of embryos not specifying an intestine (Maduro et al. 2005; Maduro et al. 2007). In all, the effectiveness and robustness of gene regulatory networks can be credited to the built-in layers of structural and functional redundancy.

1.4 *C. elegans* as a Model System

Model organisms are widely used to study a range of biological phenomena. Insights gained from extensively studied non-human species provide valuable information to questions that cannot be directly examined in humans. Various model organisms are used in research, ranging from unicellular bacteria to multicellular vertebrates such as mice. These organisms are individually unique, but collectively share the same characteristic that make it advantageous to study, due to their small size, short generation time, short life cycle, high fertility rate, small genome, and higher susceptibility to genetic modification (Leonelli, Ankeny 2013).

Both unicellular and multicellular model organisms contribute greatly to our current understanding of conserved genes and their functions. *Saccharomyces cerevisiae* is a unicellular organism, whose genes encode very similar proteins to mammals (Botstein, Fink 1988). It is no surprise that they encode similar proteins necessary for cellular processing, such as cytoskeletal proteins. However, this organism also shows functional conservation of two mammalian RAS-proto-oncogenes (Powers et al. 1984). In the absence of both genes, *S. cerevisiae* cells are not viable (Kataoka et al. 1984). Further testing showed that conservation of those genes existed not only at the sequence level, but also at the functional level (Kataoka et al. 1985). Aside from the RAS genes,

20% of human disease genes are homologous to *Saccharomyces cerevisiae* genes. Many of these diseases are caused by mutations in basic cellular processes such as DNA repair, cell division or control of gene expression (B Alberts 2002; Brown 2002).

Of all the model organisms used by the scientific community, *Drosophila melanogaster* is one of the most studied. The extensive research conducted throughout the past century has expanded the basic knowledge in molecular biology and the field of genetics. Of the 289 known human disease genes, 60% have homologs in *Drosophila* (Rubin et al. 2000; Tickoo, Russell 2002). Knowledge gained by studying this model system may help provide insight in better understanding molecular regulations and possibly advance the process in identifying cures for genetic diseases.

Like *Drosophila, Caenorhabditis elegans* is another powerful invertebrate model organism. 60-80% of human genes have an orthologue in the *C. elegans* genome (Kaletta, Hengartner 2006) and 40% of genes known to be associated with human diseases have clear orthologues in *C. elegans* (Culetto, Sattelle 2000). Studies done in *C. elegans* have paved the way for a better understanding of development in human health and disease.

1.4.1 C. elegans as a model to study Robustness

C. elegans are soil dwelling, non-pathogenic, transparent, free living nematodes that can be found all over the world. They were first used by Sydney Brenner to study animal development and neurobiology. The organism is a powerful system for genetic studies, due to the ease of genetic manipulation. Once a mutant has been shown to breed

true, the location of the mutation can be mapped using classical genetic tools (Brenner 1974). Traditionally, *C. elegans* was used to conduct forwards genetic screens, due to the range of phenotypes that it displays. In the early 2000s the genome sequence of *C. elegans* was constructed, providing more genetic information for researchers. The current advances in genome editing such as CRISPR/Cas9, makes generating precise mutations in a genome even easier.

In addition to being a powerful system for genetic studies, *C. elegans* has many advantages as a model system and is an ideal organism to study stochastic gene expression and robust noise buffering mechanisms for several reasons. First, *C. elegans* is a hermaphroditic species that can produce a large brood of isogenic embryos. Additionally, the worms are transparent, allowing reporter constructs to be easily seen and the invariant cell number and development allows cells to be easily followed throughout embryogenesis. Lastly, they have a short development and generation time, thus allowing rapid analysis of gene expression profiles. These advantages make *C. elegans* a suitable system for our studies.

1.5 *C. elegans* Development

Embryonic development of *C. elegans* is highly stereotyped. After fertilization, the zygote undergoes several rounds of holoblastic cleavage to generate seven founder cells. Following a mosaic developmental plan, the descendants of each founder cell will commit to a specific fate (Sulston et al. 1983). The cells are directed to adopt the appropriate fates through cell-autonomous activation of sets of genes that define their

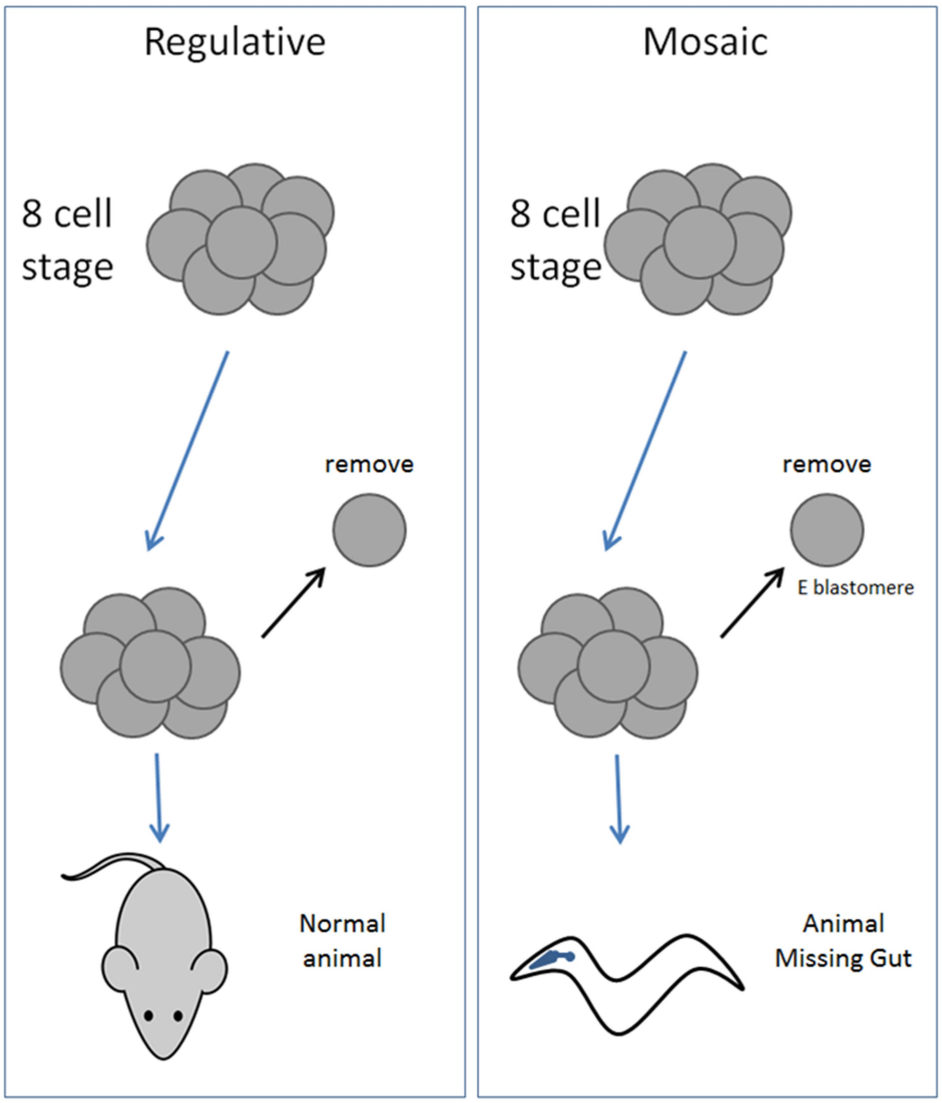


Figure 1-4: The two main types of development are regulative and mosaic. To test whether an animal follows either type of development, a single cell is removed during the eight-cell stage and left to develop under normal conditions. After development, if the animal appears normal, then it follows a regulative mode of development (left panel). In the same scenario, if the animal develops and is missing a body part, then it follows a mosaic mode of development (right panel)

lineage. The cells attain differential gene expression and adopt various fates via sequential activation of transcription factors that belong to a specific lineage and/or combination of co-activating transcription factors, defined as a gene regulatory network (GRN) (Davidson, 2010);(Macneil, Walhout 2011). Development can be broadly classified as regulative or mosaic (Fig. 1-4). To distinguish between the two types, an experiment must be done during the early stages of development. If an embryonic cell is ablated or removed at the eight-cell stage, but the remaining cells still give rise to a normal and complete animal, then the animal follows a regulative mode of development. However, if the animal develops with a missing tissue type, then development of that animal is considered mosaic. The difference between the two types of development results from the time when there is a commitment to adopt a particular cell fate during embryogenesis. In a regulatory mode of development, specification occurs later. Therefore, when a cell is removed at the eight-cell stage, none of the cells have committed to a distinct cell fate yet and can compensate for the loss of a cell. *C. elegans* follows a mosaic mode of development, where cell specification occurs early on. By the eight-cell stage, the *C. elegans* embryos have seven founder cells, and each founder cell will give rise to a specific lineage. Therefore, when a cell is removed at the eight-cell stage in *C. elegans*, it is essentially removing a whole cell lineage, and will results in an animal that is missing certain tissue-types.

This dissertation will focus on the E lineage and its core gene regulatory network. At the four-cell stage, *C. elegans* is comprised of the cells ABa, ABp, EMS, and P_2. The EMS blastomere is the precursor to the MS and E founder cells; the descendants of MS

give rise to mesoderm and posterior pharynx cells, while the descendants of E will differentiate into intestinal cells (Sulston et al. 1983). The difference between the two EMS daughter cells is due to the polarization of the P_2 cell, and P_2's contribution to the Wnt signaling in the endoderm gene regulatory network (Goldstein 1993; Thorpe et al. 1997; Schlesinger et al. 1999).

1.5.1 Life cycle

C. elegans has a rapid generation time of three days when grown at 25°C. The embryos are 25µm in length. When hatched, the larvae are .25mm in length and will be 1.5 mm when they reach adulthood and contain only 959 somatic cells. They are direct developers; the hatched juveniles resemble the adult and growth occurs through successive molts as they increase in size. The embryos are laid by gravid hermaphrodites when the embryos reach the ~22 cell stage, and it takes ~16 hours to complete embryogenesis and give rise to hatched L1 larvae. The animals begin to eat and will undergo multiple molting stages (L1-L4), until they become a fertile adult. The L1 stage is ~ 16 hours long, while the other stages take ~ 12 hours. At the end of each stage there will be a period of sleep-like inactivity called lethargus (Raizen et al. 2008), where a new cuticle is made. Lethargus ends with the molting of the old cuticle. In an unfavorable environment, such as overcrowding or depletion of nutrients, L2 animals will activate an alternative life cycle. The animals will skip the L3 molting stage and enter a dauer stage. Dauer animals are thinner than regular L3 animals; they form a cuticle layer that surrounds the entire animal. The formation of the cuticle will halt its development by

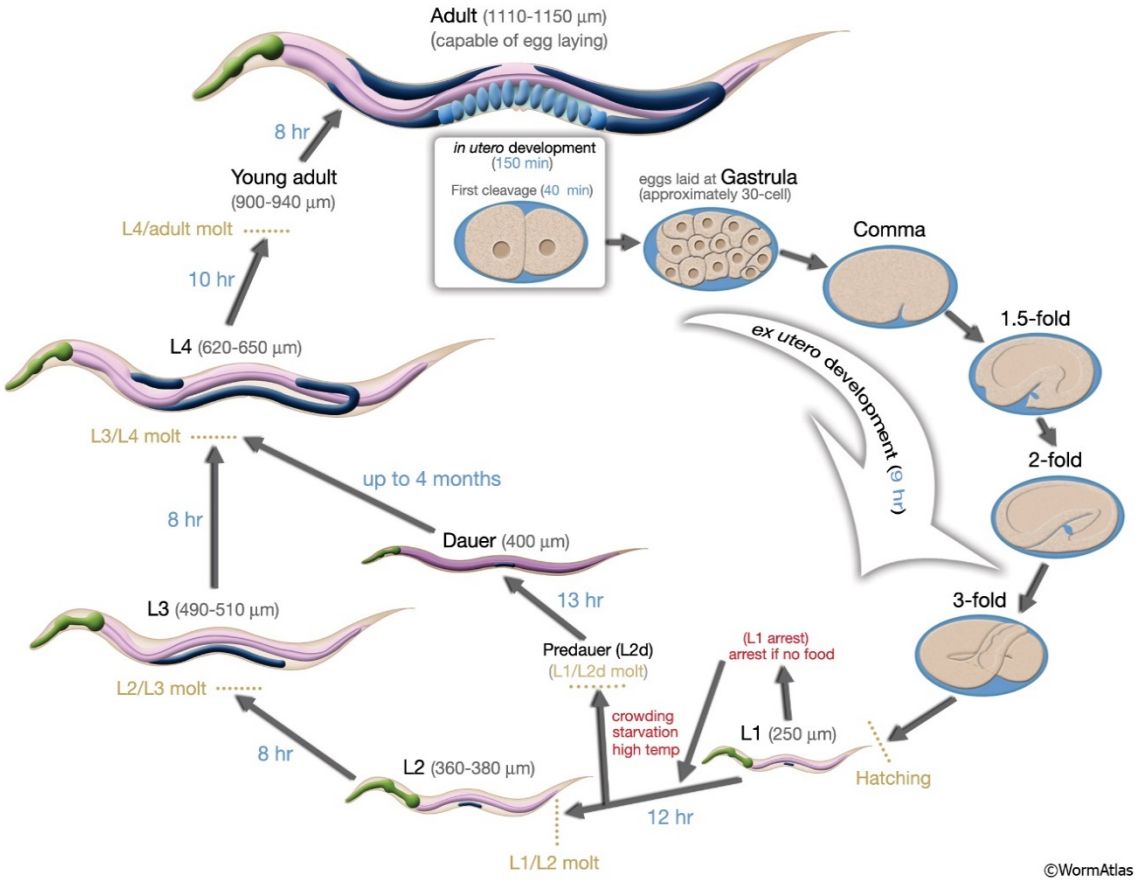


Figure 1-5: Life cycle of *C. elegans* at 22°C. Fertilization is marked as 0 minutes. The number labelled in blue denotes the number of hours the animals stays in each life stage. Embryos remain *in utero* for 150 minutes. During this time the first cell cleavage occurs at 40 minutes, and gastrulation begins around the time the embryos is laid. The length of the animal at each stage is marked next to the stage name in micrometers.

This figure is a reprinted from WormAtlas, (Altun 2005)., Handbook of *C. elegans* Anatomy, http://www.wormatlas.org/ver1/handbook/contents.htm, Copyright (2002-2006)

plugging up the mouth and anus, preventing the animal from eating and defecating. Animals can stay in the dauer stage for months, until they sense a more favorable environment. When dauer animals are moved to a new bacterial lawn, they will shed the protective cuticle and resume the regular life cycle as an L4 animal (DL Riddle 1997).

1.5.2 Sexual forms

C. elegans comes in two sexual forms, self-fertilizing hermaphrodites and males (Fig. 1-6). Hermaphrodites are females whose gonads temporarily produce sperm before switching fates to produce oocytes. When a hermaphrodite animal reaches the L4 stage, its gonad will form an ovotestis. Upon reaching sexual maturity (adult stage) it will produce haploid sperm for the first few hours; these are stored in the spermatheca, then the germline of the hermaphrodite will switch fate from sperm production to produce oocytes for the remainder of its reproductive life cycle. The natural form of breeding in *C. elegans* is to self-fertilize, but occasionally mating with males does occur. The majority of the population are self-fertilizing hermaphrodites, although males arise at a frequency of <0.2% (Corsi, Wightman, Chalfie 2015). This rare occurrence of males is due to a rare meiotic non-disjunction of the X chromosome. Hermaphrodites store enough sperm to produce up to 300 self-fertilized progenies, but generate enough oocytes to produce over 1000 progenies if mated with a male. *C. elegans* are diploid animals that contain five autosomal chromosomes. In *C. elegans* there is no Y chromosome, and sex

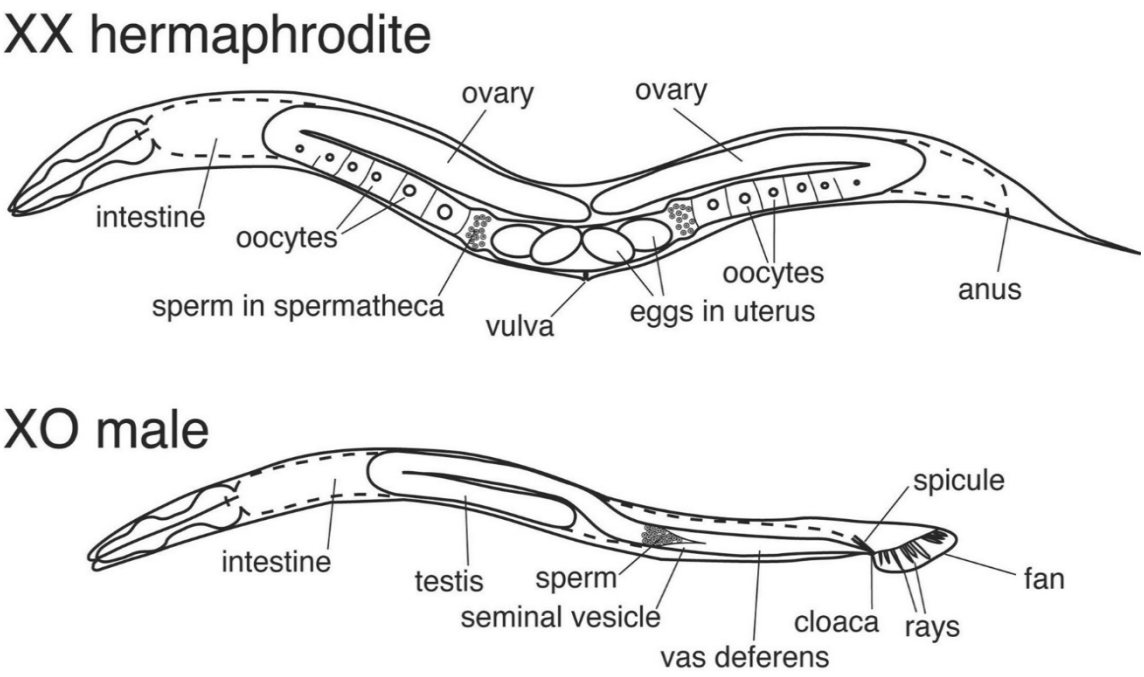


Figure 1-6: The sexual forms of *C. elegans* are male and hermaphrodite. While the majority of their tissues and organs are similar, they differ in overall body size and structure, such as gonad and tail.

This figure is a reprinted from WormBook., (Zarkower 2006), Somatic sex determination,http://www.wormbook.org/chapters/www_somaticsexdeterm/somaticsexdeterm.html

of the animal is determined by the X to autosome (X:A) ratio (Zarkower 2006). The chromosomal differences between the sexes are the single X chromosome carried by the male and the two X chromosomes carried by the hermaphrodites.

Although both sexes utilize the Wnt signaling pathway in the development of their reproductive tissues, they also have distinct anatomical differences, such as variation in their somatic gonad, secondary mating structures, and body size. The somatic gonad runs parallel to the intestine, in the center of the animal. While hermaphrodites have two somatic gonad arms, which are U-shaped tubes that are a mirror image of one another, males only have one U-shaped lobe. Other differences are the secondary sexual mating structures, which are represented by a vulva in hermaphrodites and a fan-shaped tail in males (Emmons 2005b), Herman 2006). The vulva has two functions. It is both the point of sperm entry from the male and where the embryos are laid. The vulva develops from the epidermis on the ventral side of the animal (Sternberg 2005). The small physique of the adult males is due to the absence of embryos and the second gonad arm. The flat fan-shaped tail that they have contains 18 projections of neurons and is surrounded by support cells called rays (Emmons 2005a). Dosage compensation in *C. elegans* requires signaling pathways that equalize the X-linked expression between the sexes by down-regulating X-linked gene expression by 50% on both chromosomes in XX animals, through formation of heterochromatin via methylation of the X chromosome (Meyer 2005).

1.6 Intestine Development

The intestine is the largest organ in *C. elegans*. It stretches along the length of the animal, makes up roughly one third of its somatic mass, and is connected to the pharynx and hindgut. The intestine is the product of a robust gene regulatory network, to ensure that it is properly developed. The organ is essential for digestion of nutrients. The cuticle lined structure of the animal allows it to mechanically breakdown food into smaller pieces. Enzymes and macromolecules are broken down and absorbed within the lumen structure. The gut is fully functional at hatching (first larval stage), and grows as the animal undergoes its molting stages.

1.6.1 Endoderm Gene Regulatory Network (GRN)

The core endoderm Gene Regulatory Network (GRN) of *C. elegans* is summarized in Figure 1-7. It is made up of both maternal and zygotic factors that work synergistically to activate the gut differentiation factor, *elt-2*. Upon activation of *elt-2*, it will proceed to activate all intestine specific genes. Maternal factors are transcripts or proteins, packaged into the oocyte by the mother. These factors can be founded during the earliest stages of embryogenesis and are generally required for embryo viability. The maternal factor SKN-1 is the earliest contributor of both mesoderm and endoderm cell fate. SKN-1 is a bZIP/ homeodomain transcription factor, which was the first to be identified as an early regulator of blastomeric fate in *C. elegans* (Bowerman, Eaton, Priess 1992; Blackwell et al. 1994). The maternal transcript *skn-1* is present throughout the embryo, but its protein expression is asymmetric (Bowerman et al. 1993; Seydoux,

Fire 1994). At the two-cell stage, there is a higher level of SKN-1 protein in the P_1 cell, which will divide to become the EMS and P_2 embryonic cells.

Roughly an hour into embryogenesis, the EMS blastomere divides to produce the MS (mesoderm) and E (endoderm) founder cells. The MS founder cell will give rise to cells that will eventually become mesoderm and posterior pharynx, while the E founder cell will divide and produce all the cells that make up the intestine (Sulston et al. 1983). SKN-1 functions at the top of the endoderm GRN as a transcription factor. Its main role in the endoderm gene regulatory network is to activate its zygotic targets, the *med-1* and *med-2* genes (for mesendoderm determination). Activation of the genes *med-1,2* by SKN-1 is the transition from maternal to zygotic regulation of mesendoderm specification (Maduro et al. 2001). The genes *med-1* and *med-2* are a pair of nearly identical genes, which are located on different chromosomes. The activation of the *med* genes by SKN-1 begins the endoderm specification pathway and transitions the pathways from maternal regulators to zygotic ones. The *med* genes are two of the four GATA-type transcription factors that function in the endoderm GRN, named after the consensus DNA binding site GTATACTYYY that they associate with (Broitman-Maduro, Maduro, Rothman 2005; Maduro et al. 2005; Lowry et al. 2009).

SKN-1 proteins bind directly to the two *med* genes, and high ectopic expression of *skn-1* throughout the embryo is sufficient to turn on the *med* genes in non-EMS cells. This causes non-EMS descendants to adopt a mesendodermal fate. In the absence of the SKN-1 protein, roughly 80% of embryos will not generate any endodermal cells.

Although *skn-1* is expressed in the P_2 cell, it does not cause the cell to activate mesendodermal genes. Another maternally provided transcription factor, PIE-1, acts as a global repressor of transcription through the germline (P) lineage (Mello et al. 1996; Seydoux et al. 1996; Batchelder et al. 1999), where PIE-1 inhibits the activation of the two *med* genes by SKN-1 in the P2 lineage (Maduro et al. 2001). MED-1, 2 have an alternative role and functions to prevent the cell from differentiating to a different tissue type. In an embryo without SKN-1 and MED 1-2 proteins, the E and MS cells will adopt a C like fate and produce body wall muscles and hypodermis (Hunter, Kenyon 1996).

The C cell is one of the P_2 daughter cells (Sulston et al. 1983; Bowerman, Eaton, Priess 1992; Maduro et al. 2001). The specification factor necessary for the differentiation of the C fate is the maternally provided transcription factor PAL-1. This protein is found in all early P_1 descendants, including E, MS, and C (Hunter, Kenyon 1996). In the absence of SKN-1 and MED-1,2 the EMS daughter cells will adopt an alternative fate, with the help of the PAL-1 transcription factor. This suggests that the default specification program of the E and MS cell when mesendoderm specification is blocked is to adopt a C fate, in the presence of the PAL-1 transcription factor. It has been experimentally proven that when EMS blastomere is cultured in isolation, it divides into two MS like daughter cells. The results are similar to an embryo where the physical interaction between posterior side of the EMS cell and P_2 is blocked. This suggests that the interaction between the EMS and P_2 cell facilities cell-cell communication, required for the specification of the E cell (Schierenberg 1987; Goldstein 1992). The induction of the E cell by the P_2 cell is due to maternally provided WNT and MAPK signaling

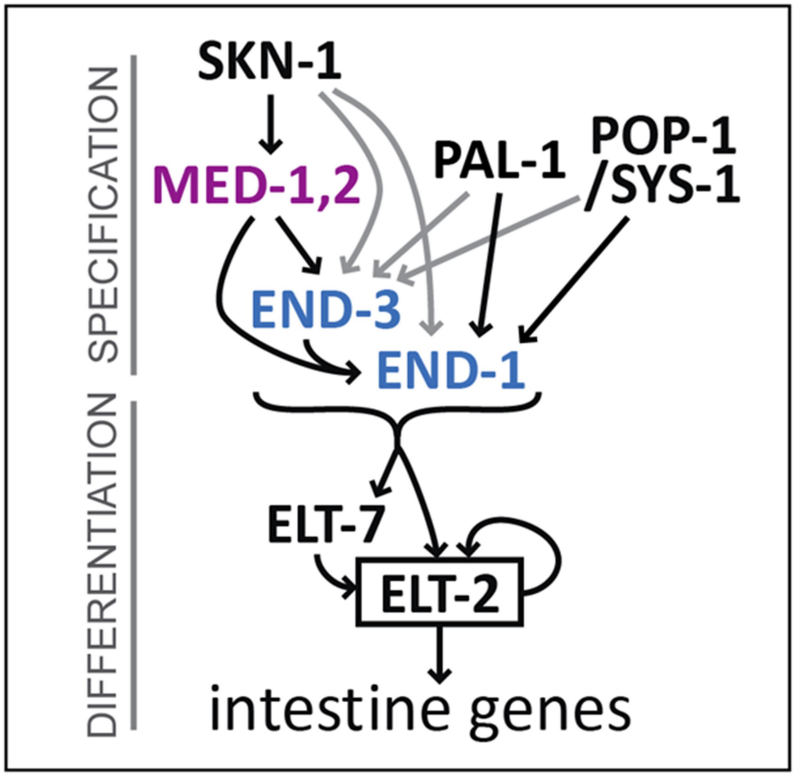


Figure 1-7: Simplified pathway showing hierarchy of transcription factors, modified from (Maduro et al. 2015). The endoderm specification strains described in this work perturb the overall contributions made by the MED-1,2 and END-1,3 regulators in a way that does not affect other lineages.

components (Rocheleau et al. 1997; Thorpe et al. 1997; Meneghini et al. 1999; Rocheleau et al. 1999; Shin et al. 1999). The loss of any of these components will results in the production of two MS-like daughter cells, like the isolation of the EMS blastomere. Wnt signaling functions to differentiate the MS and E cell through modification of the nuclear effector POP-1 (Huang et al. 2007; Maduro et al. 2007; Phillips et al. 2007). Differentiation between the EMS daughter cells requires the canonical Wnt signaling pathway (Fig. 1-8), whose default setting is to inhibit gene expression. In *C. elegans's* canonical Wnt pathway, POP-1 is a member of the TCF/LEF class HMG class protein and acts as a transcriptional regulator (Fig.1-9) (Brunner et al. 1997; Korswagen, Clevers 1999). POP-1 represses expression of Wnt targeted genes by interacting with the co-repressor GROUCHO and binding to the promoter region. After Wnt signaling occurs the divergent *β-catenin* SYS-1 is stabilized in the cytoplasm and, upon reaching a certain threshold amount, *β-catenin* will migrate into the nucleus. POP-1 will preferentially bind to *β-catenin*, thereby removing the inhibition of the Wnt targeted genes, and together POP-1/*β-catenin* act as co activators to begin transcription of the Wnt target genes.

The transduction of the Wnt signaling pathway ultimately effects the nuclear expression of POP-1 in the E and MS daughter cells. Although POP-1 is present in both MS and E, it has been shown that in MS the POP-1transcription factor is localized in the nucleus, where it acts as a repressor of *end-3*. The repression of *end-3* by POP-1 in the MS cell, prevents the specification of the E fate. In contrast, the E cell contains low expression of nuclear POP-1. Due to the cell-cell interaction between the P2 and E cell, Wnt signaling facilitates the phosphorylation of POP-1 in the nucleus. Phosphorylation of

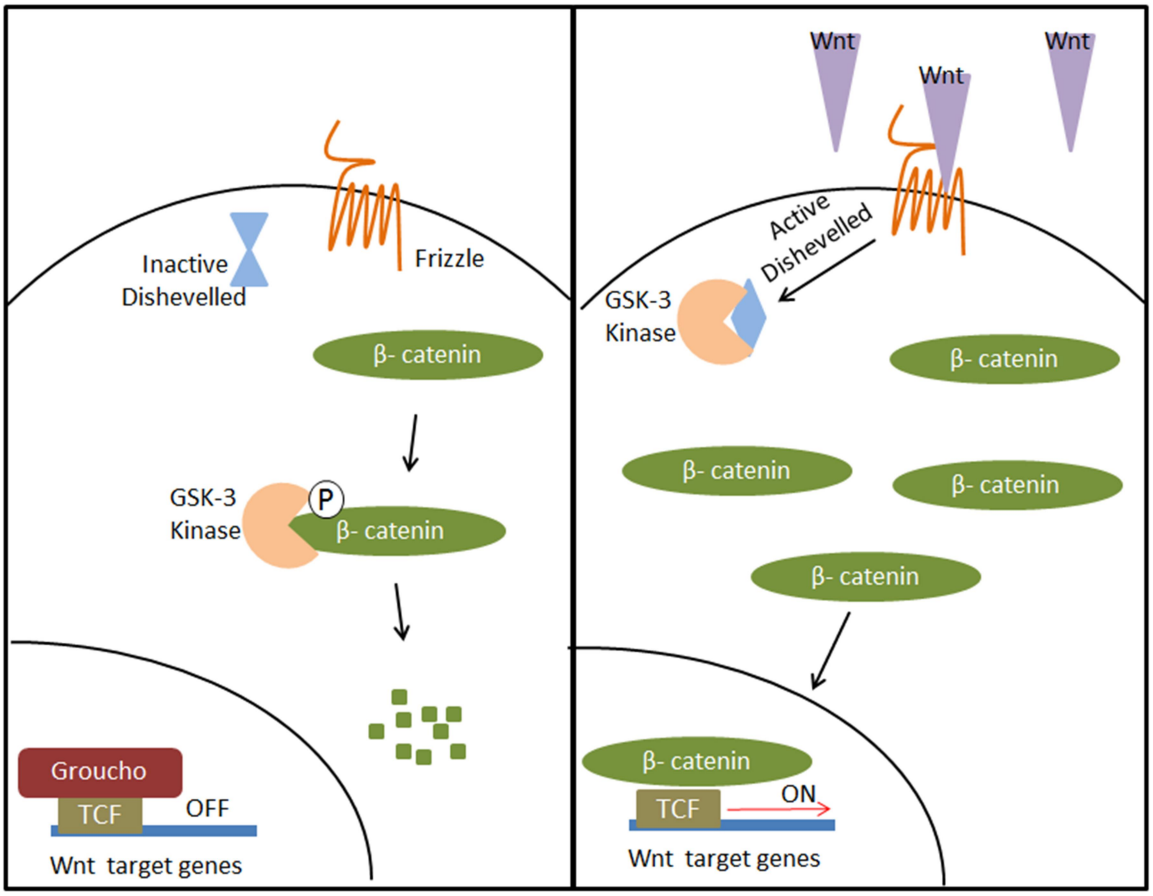


Figure 1-8: The diagram depicted on the left occurs in the absence of the Wnt ligand. In the absence of the Wnt ligands the Dishevelled protein is inactive. β-catenin in the cytoplasmic region is phosphorylated by GSK-3 and targeted for degradation. In the nucleus, TCF associates with the protein GROUCHO and inhibits expression of Wnt targeted genes. The image on the right depicts a scenario where the Wnt ligand is present. The Wnt ligand will bind to the Frizzle receptor and activate Dishevelled. Upon activation Dishevelled will prevent GSK-3 from phosphorylating β-catenin. β-catenin will accumulate in the cytoplasm, and enter the nucleus once it reaches a threshold amount. Once β-catenin enters the nucleus it will interact with TCF, and function together as a transcriptional activator of Wnt target genes.

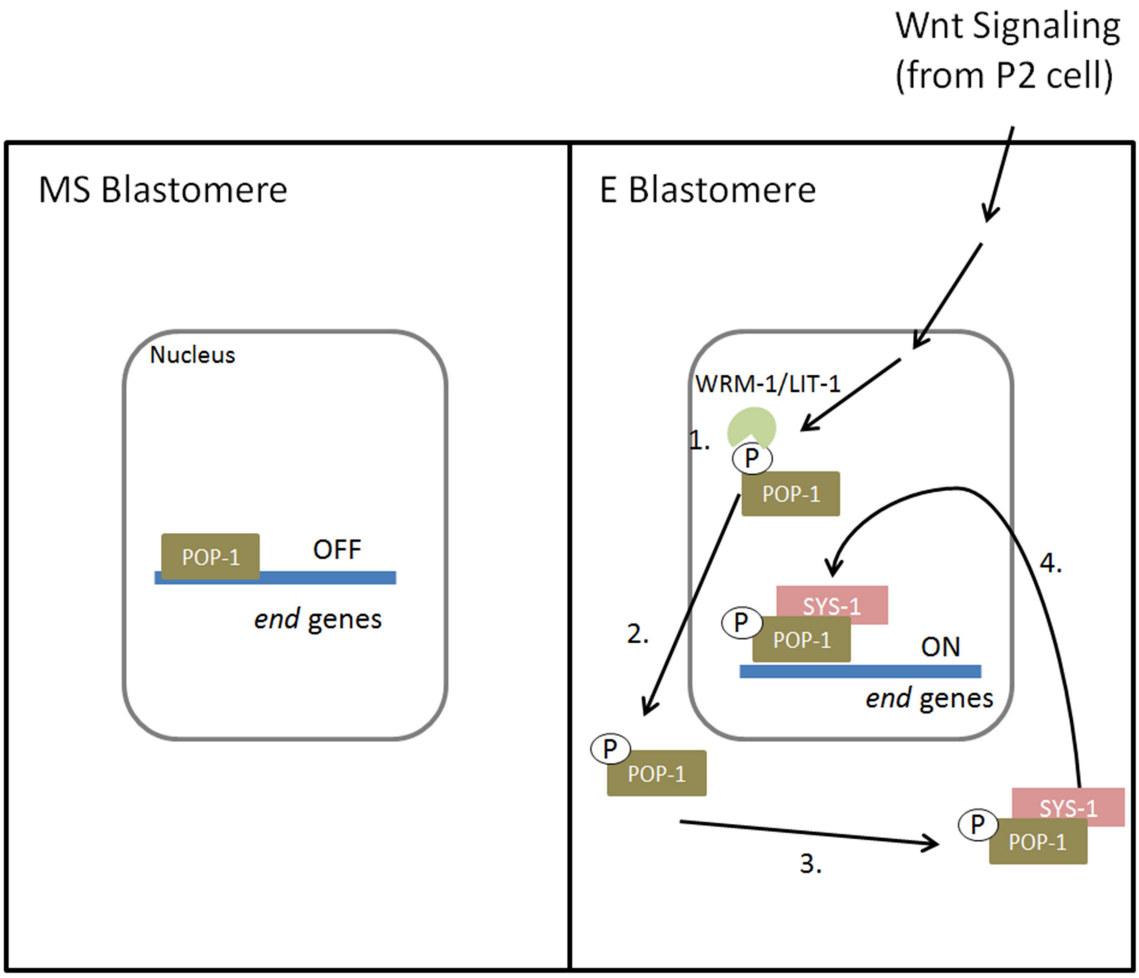


Figure 1-9: The MS blastomere does not receive Wnt signaling due to its proximity to the P2 cell. The transcription factor POP-1 binds to the promoter of the *end* genes and repress specification of the endoderm fate. The E blastomere receives Wnt signaling from its neighboring cell P2, which leads to the following events: 1) WRM-1/LIT-1 phosphorylates POP-1 and removes its repression on the *end genes.* 2) Phosphorylated POP-1 leaves the nucleus. 3) POP-1 associates with SYS-1. 4) POP-1 + SYS-1 enters the nucleus and functions as the transcriptional activator of the *end* genes. This is not the main regulatory input into *end* activation, however, as even in the complete absence of *pop-1* endoderm is still made. Rather, input from POP-1/SYS-1 adds to that from SKN-1/MED-1,2.

POP-1 removes its repression of *end-3*, and POP-1 is subsequently transported to the cytoplasm. The phosphorylated POP-1 in the cytoplasm interacts with the SYS-1 protein, and together the POP-1/SYS-1 complex also acts as a transcriptional activator of *end-3* (Huang et al., 2007; Kidd et al., 2005; Shetty et al., 2005). This input is in parallel with input provided by SKN-1/MED-1,2 (Maduro et al., 2015).

The endoderm lineage is specified when the specification genes *end-3* and its downstream target *end-1* are activated. Together the pair of *end* genes activate *elt-2*, the differentiation factor of the endoderm. To ensure the gut specification genes are activated, the endoderm gene regulatory network contains parallel pathways, utilizing transcription factors such as SKN-1, MED-1, 2, and POP-1/SYS-1 to activate *end-1, 3*. Upon the activation of *elt-2*, it will maintain its expression through auto-regulation, and activate 1000's of genes that are responsible for the structure and function of the gut (Zhu et al. 1998; Maduro et al. 2005; Sommermann et al. 2010).

1.6.2 Structure of the Gut

The gut is derived from a single progenitor cell, the E founder cell. The E cell undergoes a series of divisions to give rise to 20 gut cells at the end of embryogenesis. The 20 gut cells are encapsulated in 9 intestinal rings (1-9), from anterior to posterior (Sulston et al. 1983). The first intestinal ring contains 4 gut nuclei, while intestinal rings 2-9 each contain 2 gut nuclei. The intestinal rings are named based on their position: right, left, dorsal, ventral. During embryogenesis, cells begin to migrate. This leads to some intestinal rings adopting nearly dorsal/ ventral orientation (Leung, Hermann, Priess

1999; Asan, Raiders, Priess 2016). During the first larval stage 14 of the nuclei hosted in int 3-9 will undergo an additional nuclear division to become bi-nucleated, thereby increasing the total number of gut nuclei to ~30-34 per animal.

The interior of the gut consists of the gut lumen; which lines the digestive tract on the apical surface of the intestine cell. The lumen is a flattened, oval shape structure, lined with microvilli. The shape and structure of the gut lumen allows it to support *C. elegans* sinusoidal movement (Asan, Raiders, Priess 2016). The earliest signs of gut lumen formation occur at the 16E stage, the time when nuclei are repositioning themselves to align from anterior to posterior along the midline, while other organelles move away. The lumen is formed in segments within each intestinal ring, where it begins as a small region in the center of the ring to separate the cells (Leung, Hermann, Priess 1999). The lumen relies on the proteins which make up the microvilli and the establishment of apicobasal polarity to support its structure. The protein Ezrin/Radixin/Moesin ERM-1 is a membrane skeletal linker protein, that appears inside the forming lumen surface. It maintains the uniform width of the intestine, and works in conjunction with the actin gene *act-5* and B-H-spectrin gene *sma-1* to stabilize the apical membrane onto the cytoskeleton (Gobel et al. 2004; Maduro 2017). The formation and maintenance of the lumen is dependent on both trafficking of endosomal vesicles and lipid biogenesis. The clathrin coat protein CHC-1 is required for the apical positioning of the lumen. In the absence of *chc-1*, the expression of ERM-1 is no longer localized to the apical surface of the intestinal rings, but is distributed throughout the gut primordia (Zhang et al. 2012). The synthesis of specialized lipids such as glycosphingolipids are

essential for the formation of the lumen, due to their role in sorting lumen components to the apical membrane. In a *let-767* mutant, where glycosphingolipid synthesis is knocked down, the results resemble that of a *chc-1* mutant. Other factors also contribute to the regulation of the lumen. For example, a mutation in LET-413 results in the lumen relocating to the basolateral surface of the intestine (Zhang et al. 2011; Zhang et al. 2012).

The exterior of the gut is surrounded by basement membrane. With the exception of the foregut and hindgut, the intestine is not attached to the body (Leung, Hermann, Priess 1999). Although all gut cells are derived from the same precursor cell, and they have a similar overall appearance, they do display different functions. The more anterior cells are rich in ER and secretory vesicles and contain less storage granules, and the posterior cells have more storage granules and less secretory vesicles (Borgonie 1995). This suggests the primary function of the anterior gut is to excrete enzymes to help breakdown the food intake, while the posterior gut functions to absorb the nutrients.

1.6.3 Functions

The main role of the intestine is to break down macromolecules and to process them to extract chemical energy. The digestive tract runs along almost the entire length of the animal. Unlike humans, where the intestinal crypt maintains a constant population of new gut cells, in *C. elegans*, the cells that make up the fully formed intestine will serve the animal for the rest of its lifespan (Maduro 2017). The digestive tract is compartmentalized into the pharynx, intestine, and hindgut, with a small number of valve

cells separating each region. The pharynx is a muscular organ that pulsates to move food along, it is analogous to the esophagus in humans. The pharynx itself is segmented into two. The posterior pharynx acts as a grinder to mechanically breakdown food that moved through the pharynx intestinal valve and into the anterior lumen of the intestine (Mango 2007). As the substances moves along the intestinal lumen, the contents are digested by enzymes and absorbed. Defecation occurs in the hindgut area, through a set of contractions of the body muscles. Defecation occurs about every 45 seconds, and roughly 43% of the entire intestine volume is expelled (Liu, Thomas 1994; Ghafouri, McGhee 2007). On average, the contents of the intestine only stay in the animal for 1-2 minutes. Other important functions of the gut include the storage of dietary fat via lipid droplets through the lysosomal compartments in the gut cytoplasm (Lee 2002).

Upon reaching the eight E cell stage, the intestine begins to undergo endocytosis to transport the rest of the yolk from the embryo into the gut primordium (Bossinger, Wiegner, Schierenberg 1996; Yu et al. 2006). Endocytosis and trafficking also play a vital role in the intestine of the hatched animal. The *C. elegans* diet consists of mainly rotting vegetation and bacteria, and upon ingestion, molecules that aid in digestion and protection of the animals are released into the lumen of the intestine (McGhee 2007). The food is broken down and the nutrients are absorbed through transporters by endocytosis, while the remnants are passed along through the intestines to be expelled through defecation, or used for the synthesis of vitellogenins, which will be packaged into oocytes as their source of energy throughout embryogenesis (Robinson 2008; Lemieux, Ashrafi 2015).

Due to the nature of *C. elegans* diet, its intestine can detect and respond to infections. It is capable of innate immune response by activating the MAPK signaling pathway to respond to bacterial infection (Kim et al. 2002; Kim et al. 2004), and can fight off viral infection via the RNAi mechanism (Felix et al. 2011; Guo et al. 2013). Aside from the MAPK signaling pathway, the insulin-like signaling pathway also contributes to host immunity as well as to regulating metabolism, stress responses, and longevity (Gravato-Nobre, Hodgkin 2005; Ewbank 2006; Ermolaeva, Schumacher 2014). The lifespan of the animal can be extended from increased pathogen resistance (Garsin et al. 2003; Troemel et al. 2006) and by caloric restriction (Lee et al. 2006; Smith et al. 2008; Greer, Brunet 2009). The roles of innate immunity and stress response have also been linked to aging and longevity in *C. elegans*. Like humans, *C. elegans* immune responses decline with age, and older animals are more susceptible to pathogens (Laws et al. 2004).

The function of the pharynx and anterior gut is to breakdown the bacteria and to aid in the absorption of nutrients. Older animals lose their ability to breakdown bacteria, allowing them to colonize in the intestine. When animals are fed dead bacteria they exhibit an extended lifespan, which suggests the inability to breakdown bacteria and their colonization causes death (Garigan et al. 2002). The expression of PMK-1, which is part of the MAPK pathway necessary for immune response decreases with age , which sheds light on why immune response declines with age.

1.7 Conclusion

Robust expression of genes during early embryogenesis is crucial for proper embryonic development to obtain normal phenotypes. During this delicate process, gene regulatory networks are subjected to sources of noise, including intrinsic noise such as the abundance and availability of transcription factors and extrinsic differences in its environment. Organisms have evolved robust gene expression due to the structure of gene regulatory networks (GRNs). This dissertation aims to utilize *C. elegans'* well-studied endoderm gene regulatory network to investigate mechanisms by which the embryo buffers stochastic variation in gene expression during early development. These projects aim to exploit the mosaic mode of development and stereotyped cell divisions of the organism to study cell fate specification within the endoderm lineage and to identify differentially expressed genes that attributed to loss of integrity and functionally of differentially intestine tissue. To approach these research questions, it is necessary to have an experimental platform to associate upstream noise with downstream effects for quantitative *in vivo* studies in a set of isogenic strains that display a range of incomplete penetrance defects in gut specification. To study the downstream effects of animals that were at the verge of not specifying gut cells and nearly missed "near-miss" endoderm differentiation, a series of allelic mutants generated in the Maduro lab were used. A single-copy transgene containing the E progenitor specifying gene *end-1* and *end-3*, with a deletion of one or more MED-1,2 binding sites, was inserted into the genome of animals that contain an *end-1* and *end-3* null background. We also generated additional partial specification strains through the use of *end-3* single mutants and *med-1; end-3*

double mutants. This resulted in a series of mutant strains that are capable of specifying a wide range of intestine cells between 0%-100% of the time. The results establish that endoderm specification is not an all-or-none event at the level of the E blastomere, but rather can be displaced as a binary choice at later time points within the E lineage – a phenomenon that we are calling "stratified specification". Furthermore, we find that fully differentiated intestine retains a memory of their “near-miss” specification, resulting in an abnormal adult phenotype and maintenance of an alternative set of expressed genes. The work suggests that perturbation of some regulatory networks does not necessarily become corrected by later reinforcement of organ identity.

1.8 References

Acar, M, JT Mettetal, A van Oudenaarden. 2008. Stochastic switching as a survival strategy in fluctuating environments. Nat Genet 40:471-475.

Alon, U, MG Surette, N Barkai, S Leibler. 1999. Robustness in bacterial chemotaxis. Nature 397:168-171.

Altun, ZF. 2005. Handbook of C. elegans Anatomy. In WormAtlas.

Arnone, MI, EH Davidson. 1997. The hardwiring of development: organization and function of genomic regulatory systems. Development 124:1851-1864.

Arnosti, DN, S Barolo, M Levine, S Small. 1996. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development 122:205-214.

Asan, A, SA Raiders, JR Priess. 2016. Morphogenesis of the C. elegans Intestine Involves Axon Guidance Genes. PLoS Genet 12:e1005950.

Athale, CA, H Chaudhari. 2011. Population length variability and nucleoid numbers in Escherichia coli. Bioinformatics 27:2944-2948.

B Alberts, AJ, J Lewis. 2002. Studying Gene Expression and Function. Molecular Biology of the Cell. New York: Garland Science.

Balaban, NQ, J Merrin, R Chait, L Kowalik, S Leibler. 2004. Bacterial persistence as a phenotypic switch. Science 305:1622-1625.

Balazsi, G, A van Oudenaarden, JJ Collins. 2011. Cellular decision making and biological noise: from microbes to mammals. Cell 144:910-925.

Barkai, N, S Leibler. 1997. Robustness in simple biochemical networks. Nature 387:913-917.

Bastiaens, P. 2009. Systems biology: When it is time to die. Nature 459:334-335.

Batchelder, C, MA Dunn, B Choy, Y Suh, C Cassie, EY Shim, TH Shin, C Mello, G Seydoux, TK Blackwell. 1999. Transcriptional repression by the Caenorhabditis elegans germ-line protein PIE-1. Genes Dev 13:202-212.

Ben-Tabou de-Leon, S, EH Davidson. 2007. Gene regulation: gene control network in development. Annu Rev Biophys Biomol Struct 36:191.

Benitez, M, ER Alvarez-Buylla. 2010. Dynamic-module redundancy confers robustness to the gene regulatory network involved in hair patterning of Arabidopsis epidermis. Biosystems 102:11-15.

Bergman, A, ML Siegal. 2003. Evolutionary capacitance as a general feature of complex gene networks. Nature 424:549-552.

Bhattacherjee, A, P Biswas. 2010. Neutrality and evolvability of designed protein sequences. Phys Rev E Stat Nonlin Soft Matter Phys 82:011906.

Bien-Willner, GA, P Stankiewicz, JR Lupski. 2007. SOX9cre1, a cis-acting regulatory element located 1.1 Mb upstream of SOX9, mediates its enhancement through the SHH pathway. Hum Mol Genet 16:1143-1156.

Blackwell, TK, B Bowerman, JR Priess, H Weintraub. 1994. Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. Science 266:621-628.

Blake, WJ, KA M, CR Cantor, JJ Collins. 2003. Noise in eukaryotic gene expression. Nature 422:633-637.

Bolker, JA. 2000. Modularity in development and why it matters to evo-devo. American Zoologist 40:770-776.

Borgonie, G. 1995. Ultrastructure of the intestine of the bacteriophagous nematodes Caenorhabditis elgans, Panagrolaimus superbus and Acrobeloides maximus (Nematoda: Rhabditida). Fundamental and applied nematology:123-133.

Bossinger, O, O Wiegner, E Schierenberg. 1996. Embryonic gut differentiation in nematodes: endocytosis of macromolecules and its experimental inhibition. Roux Arch Dev Biol 205:494-497.

Botstein, D, GR Fink. 1988. Yeast: an experimental organism for modern biology. Science 240:1439-1443.

Bowerman, B, BW Draper, CC Mello, JR Priess. 1993. The maternal gene skn-1 encodes a protein that is distributed unequally in early C. elegans embryos. Cell 74:443-452.

Bowerman, B, BA Eaton, JR Priess. 1992. skn-1, a maternally expressed gene required to specify the fate of ventral blastomeres in the early C. elegans embryo. Cell 68:1061-1075.

Brenner, S. 1974. The genetics of Caenorhabditis elegans. Genetics 77:71-94.

Brock, A, H Chang, S Huang. 2009. Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. Nat Rev Genet 10:336-342.

Broitman-Maduro, G, MF Maduro, JH Rothman. 2005. The noncanonical binding site of the MED-1 GATA factor defines differentially regulated target genes in the C. elegans mesendoderm. Dev Cell 8:427-433.

Brown, T. 2002. Chapter 14, Mutation, Repair and Recombination. Genomes: Oxford: Wiley-Liss.

Brunner, E, O Peter, L Schweizer, K Basler. 1997. pangolin encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in Drosophila. Nature 385:829-833.

Cadigan, KM, U Grossniklaus, WJ Gehring. 1994. Functional redundancy: the respective roles of the two sloppy paired genes in Drosophila segmentation. Proc Natl Acad Sci U S A 91:6324-6328.

Cannavo, E, P Khoueiry, DA Garfield, P Geeleher, T Zichner, EH Gustafson, L Ciglar, JO Korbel, EE Furlong. 2016. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. Curr Biol 26:38-51.

Chalancon, G, CN Ravarani, S Balaji, A Martinez-Arias, L Aravind, R Jothi, MM Babu. 2012. Interplay between gene expression noise and regulatory network architecture. Trends Genet 28:221-232.

Chang, HH, M Hemberg, M Barahona, DE Ingber, S Huang. 2008. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 453:544-547.

Corsi, AK, B Wightman, M Chalfie. 2015. A Transparent Window into Biology: A Primer on Caenorhabditis elegans. Genetics 200:387-407.

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

Davidson, EH. 2001. Genomic regulatory systems : development and evolution. San Diego: Academic Press.

DL Riddle, TB, BJ Meyer, et al. 1997. Section II, The Dauer State. C. elegans II. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Edwards, JS, BO Palsson. 2000. Robustness analysis of the Escherichia coli metabolic network. Biotechnol Prog 16:927-939.

Elowitz, MB, AJ Levine, ED Siggia, PS Swain. 2002. Stochastic gene expression in a single cell. Science 297:1183-1186.

Emmons, SW. 2005a. Male development. In: W The C. elegans Research Community, editor. WormBook, ed.

Emmons, SW. 2005b. Sexual behavior of the Caenorhabditis elegans male. Neurobiology of C. Elegans 69:99-+.

Ermolaeva, MA, B Schumacher. 2014. Insights from the worm: the C. elegans model for innate immunity. Semin Immunol 26:303-309.

Ewbank, JJ. 2006. Signaling in the immune response. WormBook:1-12.

Felix, MA, A Ashe, J Piffaretti, et al. 2011. Natural and experimental infection of Caenorhabditis nematodes by novel viruses related to nodaviruses. PLoS Biol 9:e1000586.

Felix, MA, M Barkoulas. 2015. Pervasive robustness in biological systems. Nat Rev Genet 16:483-496.

Frankel, N, GK Davis, D Vargas, S Wang, F Payre, DL Stern. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466:490-493.

Fraser, D, M Kaern. 2009. A chance at survival: gene expression noise and phenotypic diversification strategies. Mol Microbiol 71:1333-1340.

Freeman, M. 2000. Feedback control of intercellular signalling in development. Nature 408:313-319.

Fusco, G, A Minelli. 2010. Phenotypic plasticity in development and evolution: facts and concepts. Introduction. Philos Trans R Soc Lond B Biol Sci 365:547-556.

Garigan, D, AL Hsu, AG Fraser, RS Kamath, J Ahringer, C Kenyon. 2002. Genetic analysis of tissue aging in Caenorhabditis elegans: a role for heat-shock factor and bacterial proliferation. Genetics 161:1101-1112.

Garsin, DA, JM Villanueva, J Begun, DH Kim, CD Sifri, SB Calderwood, G Ruvkun, FM Ausubel. 2003. Long-lived C. elegans daf-2 mutants are resistant to bacterial pathogens. Science 300:1921.

Ghafouri, S, JD McGhee. 2007. Bacterial residence time in the intestine of Caenorhabditis elegans. Nematology 9:87-91.

Gibson, G. 2009. Decanalization and the origin of complex disease. Nat Rev Genet 10:134-140.

Gibson, TJ, J Spring. 1998. Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. Trends Genet 14:46-49; discussion 49-50.

Gilbert, S. 2000. Early Drosophila Development. Developmental Biology. Sunderland (MA): Sinauer Associates.

Gobel, V, PL Barrett, DH Hall, JT Fleming. 2004. Lumen morphogenesis in C. elegans requires the membrane-cytoskeleton linker erm-1. Dev Cell 6:865-873.

Goldstein, B. 1992. Induction of gut in Caenorhabditis elegans embryos. Nature 357:255-257.

Goldstein, B. 1993. Establishment of gut fate in the E lineage of C. elegans: the roles of lineage-dependent mechanisms and cell interactions. Development 118:1267-1277.

Gravato-Nobre, MJ, J Hodgkin. 2005. Caenorhabditis elegans as a model for innate immunity to pathogens. Cell Microbiol 7:741-751.

Gray, S, M Levine. 1996. Short-range transcriptional repressors mediate both quenching and direct repression within complex loci in Drosophila. Genes Dev 10:700-710.

Greer, EL, A Brunet. 2009. Different dietary restriction regimens extend lifespan by both independent and overlapping genetic pathways in C. elegans. Aging Cell 8:113-127.

Gu, Z, LM Steinmetz, X Gu, C Scharfe, RW Davis, WH Li. 2003. Role of duplicate genes in genetic robustness against null mutations. Nature 421:63-66.

Guo, X, R Zhang, J Wang, SW Ding, R Lu. 2013. Homologous RIG-I-like helicase proteins direct RNAi-mediated antiviral immunity in C. elegans by distinct mechanisms. Proc Natl Acad Sci U S A 110:16085-16090.

Hong, JW, DA Hendrix, MS Levine. 2008. Shadow enhancers as a source of evolutionary novelty. Science 321:1314.

Huang, S, P Shetty, SM Robertson, R Lin. 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.

Huh, D, J Paulsson. 2011. Random partitioning of molecules at cell division. Proc Natl Acad Sci U S A 108:15004-15009.

Hunter, CP, C Kenyon. 1996. Spatial and temporal controls target pal-1 blastomere-specification activity to a single blastomere lineage in C. elegans embryos. Cell 87:217-226.

Ingolia, NT. 2004. Topology and robustness in the Drosophila segment polarity network. PLoS Biol 2:e123.

Istrail, S, EH Davidson. 2005. Logic functions of the genomic cis-regulatory code. Proc Natl Acad Sci U S A 102:4954-4959.

Jack, J, D Dorsett, Y Delotto, S Liu. 1991. Expression of the cut locus in the Drosophila wing margin is required for cell type specification and is regulated by a distant enhancer. Development 113:735-747.

Jeziorska, DM, KW Jordan, KW Vance. 2009. A systems biology approach to understanding cis-regulatory module function. Semin Cell Dev Biol 20:856-862.

Johnston, IG, B Gaal, RP Neves, T Enver, FJ Iborra, NS Jones. 2012. Mitochondrial variability as a source of extrinsic cellular noise. PLoS Comput Biol 8:e1002416.

Kaern, M, TC Elston, WJ Blake, JJ Collins. 2005. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet 6:451-464.

Kaletta, T, MO Hengartner. 2006. Finding function in novel targets: C. elegans as a model organism. Nat Rev Drug Discov 5:387-398.

Kataoka, T, S Powers, S Cameron, O Fasano, M Goldfarb, J Broach, M Wigler. 1985. Functional homology of mammalian and yeast RAS genes. Cell 40:19-26.

Kataoka, T, S Powers, C McGill, O Fasano, J Strathern, J Broach, M Wigler. 1984. Genetic analysis of yeast RAS1 and RAS2 genes. Cell 37:437-445.

Keller, EF. 2002. Developmental robustness. Ann N Y Acad Sci 981:189-201.

Kim, DH, R Feinbaum, G Alloing, et al. 2002. A conserved p38 MAP kinase pathway in Caenorhabditis elegans innate immunity. Science 297:623-626.

Kim, DH, NT Liberati, T Mizuno, H Inoue, N Hisamoto, K Matsumoto, FM Ausubel. 2004. Integration of Caenorhabditis elegans MAPK pathways mediating immunity and stress resistance by MEK-1 MAPK kinase and VHP-1 MAPK phosphatase. Proc Natl Acad Sci U S A 101:10990-10994.

Kitano, H. 2003. Cancer robustness: tumour tactics. Nature 426:125.

Kitano, H. 2004. Biological robustness. Nat Rev Genet 5:826-837.

Kitano, H, K Oda, T Kimura, Y Matsuoka, M Csete, J Doyle, M Muramatsu. 2004. Metabolic syndrome and robustness tradeoffs. Diabetes 53 Suppl 3:S6-S15.

Klingenberg, CP, AV Badyaev, SM Sowry, NJ Beckwith. 2001. Inferring developmental modularity from morphological integration: analysis of individual variation and asymmetry in bumblebee wings. Am Nat 157:11-23.

Korswagen, HC, HC Clevers. 1999. Activation and repression of wingless/Wnt target genes by the TCF/LEF-1 family of transcription factors. Cold Spring Harb Symp Quant Biol 64:141-147.

Kussell, E, R Kishony, NQ Balaban, S Leibler. 2005. Bacterial persistence: a model of survival in changing environments. Genetics 169:1807-1814.

Laws, TR, SV Harding, MP Smith, TP Atkins, RW Titball. 2004. Age influences resistance of Caenorhabditis elegans to killing by pathogenic bacteria. FEMS Microbiol Lett 234:281-287.

Lee, DL. 2002. The biology of nematodes. London: Taylor & Francis.

Lee, GD, MA Wilson, M Zhu, CA Wolkow, R de Cabo, DK Ingram, S Zou. 2006. Dietary deprivation extends lifespan in Caenorhabditis elegans. Aging Cell 5:515-524.

Lemieux, GA, K Ashrafi. 2015. Insights and challenges in using C. elegans for investigation of fat metabolism. Crit Rev Biochem Mol Biol 50:69-84.

Leonelli, S, RA Ankeny. 2013. What makes a model organism? Endeavour 37:209-212.

Leung, B, GJ Hermann, JR Priess. 1999. Organogenesis of the Caenorhabditis elegans intestine. Dev Biol 216:114-134.

Levy, SF, ML Siegal. 2008. Network hubs buffer environmental variation in Saccharomyces cerevisiae. PLoS Biol 6:e264.

Levy, SF, N Ziv, ML Siegal. 2012. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. PLoS Biol 10:e1001325.

Liu, DW, JH Thomas. 1994. Regulation of a periodic motor program in C. elegans. J Neurosci 14:1953-1962.

Lowry, JA, R Gamsjaeger, SY Thong, W Hung, AH Kwan, G Broitman-Maduro, JM Matthews, M Maduro, JP Mackay. 2009. Structural analysis of MED-1 reveals unexpected diversity in the mechanism of DNA recognition by GATA-type zinc finger domains. J Biol Chem 284:5827-5835.

Macneil, LT, AJ Walhout. 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res 21:645-657.

Maduro, MF. 2006. Endomesoderm specification in Caenorhabditis elegans and other nematodes. Bioessays 28:1010-1022.

Maduro, MF. 2017. Gut development in C. elegans. Semin Cell Dev Biol 66:3-11.

Maduro, MF, G Broitman-Maduro, H Choi, F Carranza, A Chia-Yi Wu, SA Rifkin. 2015. MED GATA factors promote robust development of the C. elegans endoderm. Dev Biol 404:66-79.

Maduro, MF, G Broitman-Maduro, I Mengarelli, JH Rothman. 2007. Maternal deployment of the embryonic SKN-1-->MED-1,2 cell specification pathway in C. elegans. Dev Biol 301:590-601.

Maduro, MF, RJ Hill, PJ Heid, ED Newman-Smith, J Zhu, JR Priess, JH Rothman. 2005. Genetic redundancy in endoderm specification within the genus Caenorhabditis. Dev Biol 284:509-522.

Maduro, MF, MD Meneghini, B Bowerman, G Broitman-Maduro, JH Rothman. 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 7:475-485.

Mango, SE. 2007. The C. elegans pharynx: a model for organogenesis. WormBook:1-26.

Masel, J, ML Siegal. 2009. Robustness: mechanisms and consequences. Trends Genet 25:395-403.

McAdams, HH, A Arkin. 1997. Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A 94:814-819.

McGhee, JD. 2007. The C. elegans intestine. WormBook:1-36.

Mello, CC, C Schubert, B Draper, W Zhang, R Lobel, JR Priess. 1996. The PIE-1 protein and germline specification in C. elegans embryos. Nature 382:710-712.

Meneghini, MD, T Ishitani, JC Carter, N Hisamoto, J Ninomiya-Tsuji, CJ Thorpe, DR Hamill, K Matsumoto, B Bowerman. 1999. MAP kinase and Wnt pathways converge to downregulate an HMG-domain repressor in Caenorhabditis elegans. Nature 399:793-797.

Meyer, B. 2005. X-Chromosome dosage compensation. WormBook, ed. The C. elegans Research Community, WormBook.

Newman, JR, S Ghaemmaghami, J Ihmels, DK Breslow, M Noble, JL DeRisi, JS Weissman. 2006. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441:840-846.

Ow, MC, NJ Martinez, PH Olsen, HS Silverman, MI Barrasa, B Conradt, AJ Walhout, V Ambros. 2008. The FLYWCH transcription factors FLH-1, FLH-2, and FLH-3 repress embryonic expression of microRNA genes in C. elegans. Genes Dev 22:2520-2534.

Ozbudak, EM, M Thattai, I Kurtser, AD Grossman, A van Oudenaarden. 2002. Regulation of noise in the expression of a single gene. Nat Genet 31:69-73.

Pedraza, JM, A van Oudenaarden. 2005. Noise propagation in gene networks. Science 307:1965-1969.

Phillips, BT, AR Kidd, 3rd, R King, J Hardin, J Kimble. 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.

Powers, S, T Kataoka, O Fasano, M Goldfarb, J Strathern, J Broach, M Wigler. 1984. Genes in S. cerevisiae encoding proteins with domains homologous to the mammalian ras proteins. Cell 36:607-612.

Raizen, DM, JE Zimmerman, MH Maycock, UD Ta, YJ You, MV Sundaram, AI Pack. 2008. Lethargus is a Caenorhabditis elegans sleep-like state. Nature 451:569-572.

Raj, A, A van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135:216-226.

Raser, JM, EK O'Shea. 2004. Control of stochasticity in eukaryotic gene expression. Science 304:1811-1814.

Raser, JM, EK O'Shea. 2005. Noise in gene expression: origins, consequences, and control. Science 309:2010-2013.

Rennell, D, SE Bouvier, LW Hardy, AR Poteete. 1991. Systematic mutation of bacteriophage T4 lysozyme. J Mol Biol 222:67-88.

Robinson, R. 2008. For mammals, loss of yolk and gain of milk went hand in hand. PLoS Biol 6:e77.

Rocheleau, CE, WD Downs, R Lin, C Wittmann, Y Bei, YH Cha, M Ali, JR Priess, CC Mello. 1997. Wnt signaling and an APC-related gene specify endoderm in early C. elegans embryos. Cell 90:707-716.

Rocheleau, CE, J Yasuda, TH Shin, R Lin, H Sawa, H Okano, JR Priess, RJ Davis, CC Mello. 1999. WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in C. elegans. Cell 97:717-726.

Rubin, GM, MD Yandell, JR Wortman, et al. 2000. Comparative genomics of the eukaryotes. Science 287:2204-2215.

Saffman, PG, M Delbruck. 1975. Brownian motion in biological membranes. Proc Natl Acad Sci U S A 72:3111-3113.

Schierenberg, E. 1987. Reversal of cellular polarity and early cell-cell interaction in the embryos of Caenorhabditis elegans. Dev Biol 122:452-463.

Schlesinger, A, CA Shelton, JN Maloof, M Meneghini, B Bowerman. 1999. Wnt pathway components orient a mitotic spindle in the early Caenorhabditis elegans embryo without requiring gene transcription in the responding cell. Genes Dev 13:2028-2038.

Seydoux, G, A Fire. 1994. Soma-germline asymmetry in the distributions of embryonic RNAs in Caenorhabditis elegans. Development 120:2823-2834.

Seydoux, G, CC Mello, J Pettitt, WB Wood, JR Priess, A Fire. 1996. Repression of gene expression in the embryonic germ lineage of C. elegans. Nature 382:713-716.

Shahrezaei, V, JF Ollivier, PS Swain. 2008. Colored extrinsic fluctuations and stochastic gene expression. Mol Syst Biol 4:196.

Sharma, AI, KO Yanes, L Jin, SL Garvey, SM Taha, Y Suzuki. 2016. The phenotypic plasticity of developmental modules. Evodevo 7:15.

Shin, TH, J Yasuda, CE Rocheleau, R Lin, M Soto, Y Bei, RJ Davis, CC Mello. 1999. MOM-4, a MAP kinase kinase kinase-related protein, activates WRM-1/LIT-1 kinase to transduce anterior/posterior polarity signals in C. elegans. Mol Cell 4:275-280.

Siegal, ML, A Bergman. 2002. Waddington's canalization revisited: developmental stability and evolution. Proc Natl Acad Sci U S A 99:10528-10532.

Small, S, A Blair, M Levine. 1992. Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J 11:4047-4057.

Smith, ED, TL Kaeberlein, BT Lydum, J Sager, KL Welton, BK Kennedy, M Kaeberlein. 2008. Age- and calorie-independent life span extension from dietary restriction by bacterial deprivation in Caenorhabditis elegans. BMC Dev Biol 8:49.

Sommermann, EM, KR Strohmaier, MF Maduro, JH Rothman. 2010. Endoderm development in Caenorhabditis elegans: the synergistic action of ELT-2 and -7 mediates the specification-->differentiation transition. Dev Biol 347:154-166.

Spencer, SL, S Gaudet, JG Albeck, JM Burke, PK Sorger. 2009. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature 459:428-432.

Sternberg, PW. 2005. Vulval development. WormBook:1-28.

Sulston, JE, E Schierenberg, JG White, JN Thomson. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol 100:64-119.

Tautz, D. 1992. Redundancies, development and the flow of information. Bioessays 14:263-266.

Teif, VB. 2010. Predicting gene-regulation functions: lessons from temperate bacteriophages. Biophys J 98:1247-1256.

Thattai, M, A van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A 98:8614-8619.

Thorpe, CJ, A Schlesinger, JC Carter, B Bowerman. 1997. Wnt signaling polarizes an early C. elegans blastomere to distinguish endoderm from mesoderm. Cell 90:695-705.

Tickoo, S, S Russell. 2002. Drosophila melanogaster as a model system for drug discovery and pathway screening. Curr Opin Pharmacol 2:555-560.

Troemel, ER, SW Chu, V Reinke, SS Lee, FM Ausubel, DH Kim. 2006. p38 MAPK regulates expression of immune response genes and contributes to longevity in C. elegans. PLoS Genet 2:e183.

Tumpel, S, F Cambronero, C Sims, R Krumlauf, LM Wiedemann. 2008. A regulatory module embedded in the coding region of Hoxa2 controls expression in rhombomere 2. Proc Natl Acad Sci U S A 105:20077-20082.

Ullah, M, O Wolkenhauer. 2010. Stochastic approaches in systems biology. Wiley Interdiscip Rev Syst Biol Med 2:385-397.

von Dassow, G, E Meir, EM Munro, GM Odell. 2000. The segment polarity network is a robust developmental module. Nature 406:188-192.

Waddington, CH. 1942. Canalization of development and the inheritance of acquired characters. Nature 150:563-565.

Wagner, A. 1996. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. Biol Cybern 74:557-567.

Wagner, A. 2005. Robustness, evolvability, and neutrality. FEBS Lett 579:1772-1778.

West-Eberhard, MJ. 2003. Developmental plasticity and evolution. Oxford ; New York: Oxford University Press.

Wilkins, AS. 1997. Canalization: a molecular genetic perspective. Bioessays 19:257-262.

Winzeler, EA, DD Shoemaker, A Astromoff, et al. 1999. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285:901-906.

Yi, TM, Y Huang, MI Simon, J Doyle. 2000. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. Proc Natl Acad Sci U S A 97:4649-4653.

Yu, X, S Odera, CH Chuang, N Lu, Z Zhou. 2006. C. elegans Dynamin mediates the signaling of phagocytic receptor CED-1 for the engulfment and degradation of apoptotic cells. Dev Cell 10:743-757.

Zarkower, D. 2006. Somatic sex determination. WormBook:1-12.

Zhang, H, N Abraham, LA Khan, DH Hall, JT Fleming, V Gobel. 2011. Apicobasal domain identities of expanding tubular membranes depend on glycosphingolipid biosynthesis. Nat Cell Biol 13:1189-1201.

Zhang, H, A Kim, N Abraham, LA Khan, DH Hall, JT Fleming, V Gobel. 2012. Clathrin and AP-1 regulate apical polarity and lumen formation during C. elegans tubulogenesis. Development 139:2071-2083.

Zhu, J, T Fukushige, JD McGhee, JH Rothman. 1998. Reprogramming of early embryonic blastomeres into endodermal progenitors by a Caenorhabditis elegans GATA factor. Genes Dev 12:3809-3814.

Zimmerman, L, B Parr, U Lendahl, M Cunningham, R McKay, B Gavin, J Mann, G Vassileva, A McMahon. 1994. Independent regulatory elements in the nestin gene direct transgene expression to neural stem cells or muscle precursors. Neuron 12:11-24.

CHAPTER 2

An RNAi screen for modifiers of gut specification identifies pathways in metabolism and gene expression

2.1 Abstract

Gene expression is subject to various sources of noise. These include intrinsic differences in the number of transcription factor molecules available to bind their promoters and extrinsic differences in the environment, such as temperature. We have developed a new system to examine how gene networks buffer the noise inherent to networks using the well-studied endoderm specification network. In this network, input from maternal and zygotic genes ultimately influences the activation of the embryonic E specification genes, *end-1* and *end-3*. Activation of these genes leads to activation of *elt-2* and *elt-7*, which direct a program of endoderm differentiation. Activation of the terminal gut regulator, *elt-2*, has been hypothesized to occur as a result of accumulation of threshold amounts of *end-1* mRNA (Raj et al. 2010). One way to examine how embryos deal with sub- or near-threshold amounts of *end-1* and *end-3* is to compromise their activation, and examine whether they can make gut afterwards. To partially compromise the activation of *end-1* and *end-3*, we utilized mutants that contain molecular null version of those alleles, and used them to generate strains carrying single-copy insertions of *end-1* and/or *end-3* in which the binding sites for the MED-1,2 factors, which activate *end-1*

and *end-3* expression, have been mutated. We performed RNAi screens for maternal factors that may influence the ability of this strain to specify gut. We hypothesize that such genes might act generally in gene expression, or may be more restricted to endoderm specification. In order to screen for genes affecting endoderm specification, we have selected some 3000 clones that are known to be expressed in the germline from the Vidal Feeding Library and have tested whether they can enhance or repress the number of GFP-expressing cells in a compromised endoderm-specification strain.

2.2 Introduction

During early embryogenesis, it is essential for the correct genes to be activated in the proper spatiotemporal context, and robust expression of gene regulatory networks is crucial for proper embryonic development. Gene expression is inherently stochastic and subject to variability due to intrinsic and extrinsic factors, and randomness in transcription and translation leads to cell-to-cell variation in mRNA and protein levels (Raj, van Oudenaarden 2008; Chalancon et al. 2012). In cultured, undifferentiated cells, this can lead to asymmetric cell divisions and spontaneous differentiation even if cells are exposed to the same environment. To counteract these effects, cells have evolved noise buffering mechanisms, which are responsible for ensuring developmental robustness. For example, the *Drosophila* gap genes function in a noise buffering pathway to reduce phenotypic variability in presence of either genetic or environmental perturbation (Gursky, Surkova, Samsonova 2012). In all living systems, stochastic gene expression is inevitable, and organisms are constantly undergoing changes at a genetic level.

The model system *C. elegans* is an ideal organism to study stochastic gene expression and mechanisms that buffer noise during embryonic development for several reasons. *C. elegans* is a hermaphroditic species that can produce isogenic embryos, RNAi mediated knock down can be done by feeding, and *C. elegans* has a short development and generation time, allowing rapid analysis of gene expression profiles. These advantages make *C. elegans* a suitable system for our studies. Embryonic development of *C. elegans* is highly stereotyped. After fertilization, the zygote undergoes several rounds of holoblastic cleavage that gives rise to 6 founder cells. Following a mosaic development plan, the descendants of each founder cell will commit to a specific fate (Sulston et al. 1983). The cells are directed to adopt the appropriate fates through cell-autonomous activation of sets of genes that define their lineage. The cells attain differential expression and adopt various fates via sequential activation of transcription factors that belong to a specific lineage and/or combination of co-activating transcription factors, defined as a gene regulatory network (GRN) (Davidson 2010; Macneil, Walhout 2011). A group of maternal and zygotic inputs lead to high activation of *end-1,3* that exceed a proposed threshold amount that is required for the activation of the endoderm differentiating genes *elt-2,7* (McGhee et al. 2009; Raj et al. 2010; Sommermann et al. 2010). While the upstream genes important for specification are transiently expressed, the expression of *elt-2* and *elt-7* are maintained by cross and auto-regulation. Upon activation of the *elt* genes, cells in the E lineage will commit to an endoderm fate (Bowerman et al. 1993; Maduro et al. 2005a; Kormish, Gaudet, McGhee 2010; Sommermann et al. 2010).

The connectivity of the network and duplication of essential genes are likely the key components that ensure robust phenotypes. Previous studies have been performed in *C. elegans* to identify factors that regulate development. A forward genetic screen utilized ethyl methanesulfonate (EMS) to identify modifiers of pharynx development that can suppress a partial loss of *pha-4* function (Updike, Mango 2007). We have taken an alternate approach, in order to identify modifiers of gut specification. The *C. elegans* system is known for its ability to conduct rapid screens by RNAi (Fraser et al. 2000; Maeda et al. 2001). We have taken advantage of the reverse genetics approach to conduct our targeted high throughput RNAi screen to identify modifiers that either enhance or suppress weak gut specification. For this targeted screen, only maternal genes are selected from the VIDAL feeding library for screening (Rual et al. 2004). Due to the early specification of the endoderm lineage, we hypothesize that maternal factors are the ones that play a role in buffering of noise during embryogenesis. The *end-1,3*(MED-) strains allow us to utilize the endoderm GRN of *C. elegans* as a model, with which we can ask more general questions about stochastic gene expression and understand how it is buffered in living organisms.

2.3 Material and Methods

2.3.1 Worm Maintenance and Strains Used

The Maduro lab generated an allelic series of mutants which contain various *end-1,3*(MED-) strains (Choi, Broitman-Maduro, Maduro 2017) that were used for the duration of the RNAi screen. The genotypes of the strains are as follows: MS1815: *end-1*(ok558) *end-3*(ok1448) V; *unc-119(ed3 or ed4)* III; *irSi7* [*end-1*(1-2-) *, *Cb-unc-*

119(+)] II; *wIs84* [*elt-2*::GFP, *rol-6D*] X; *irEx568* [*end-1,3(+), sur-5*::dsRed]. MS1799: *unc-119*(*ed3* or *ed4*) III; *end-1*(ok558) *end-3(ok1448)* V; *irIs98* [*end-1*(1-2-), *end-3*(1-2-3-4-), Cb-*unc-119*(+)] IV; wIs84 [*elt-2*::GFP, *rol-6D*] X. MS2330: *med-1*(ir66) wIs84 [*elt-2*::NLS::GFP::lacZ, *rol-6*D] X; *end-3*(ok1448) V. *These indicate the mutation of MED sites, two in *end-1* or four in *end-3*.

2.3.2 Feeding RNAi

Each Individual RNAi feeding bacterial strains selected from the VIDAL feeding library (Rual et al. 2004) and the control bacterial strain were cultured in 1mL of LB + 1µl of Carbencillin in a 37°C shaker. Cultures were centrifuged the following day at 14,000 rpm for 2 minutes, and 800µl of supernatant was removed. The sample was re-suspended and used to seed 3-cm plates containing IPTG to induce dsRNA expression. The plates were grown at room temperature for an additional three days to provide time for expression of specific dsRNA. Gravid worms were cleaned and synchronized using a mild solution of bleach and sodium hydroxide (Stiernagle 2006), and ~200 embryos were deposited onto each RNAi feeder plate. Embryos were allowed to hatch and feed on the RNAi feeding bacteria strain for 3 days in 20°C. On the third day 25-30 gravid animals were transferred to a regular *E. coli* OP50 seeded 3-cm plate, and left to lay embryos for ~7 hours at room temperature before they were removed from the plate. Embryos were stored at 4 °C overnight and analyzed the following day.

2.3.3 Microscopy, Imaging, and Data Analysis

Embryos were mounted in clusters on a standard 5% agar pad atop of a microscope slide. Conventional epifluorescence and differential interference microscopy

were performed on an Olympus BX-51 microscope, imaged through an LMscope adapter (Micro Tech Lab, Graz, Austria) and Canon Rebel T1 Digital Camera using software supplied with the camera. Data collected during the week were normalized to the expression of the *elt-2*::GFP from the control (empty vector) sample and in the process the data was transformed into the $\log_2$ scale. Functional analysis was performed on a subset of gene using the DAVID (Huang da, Sherman, Lempicki 2009b; Huang da, Sherman, Lempicki 2009a) and PANTHER (Mi et al. 2013; Mi et al. 2017) software.

2.4 Results

2.4.1 High throughput RNAi Screen in end-1,3(MED-1) mutants.

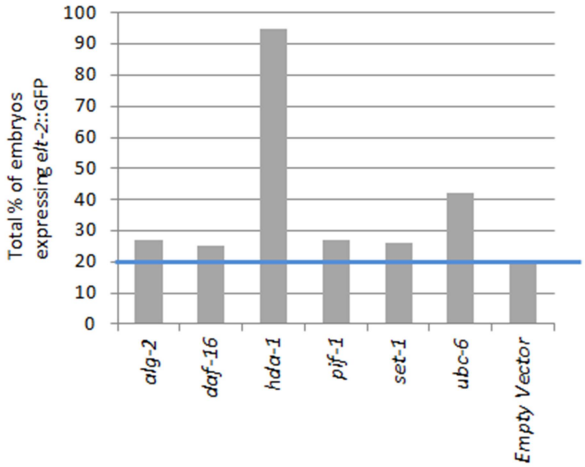
In this study we utilized various *end-1, 3* (MED-) mutants, where (MED-) denotes mutated MED-1 binding sites. Two mutant strains MS1815 and MS1799 have been used to perform the primary screen, and a third strain MS2330 was used to retest a random subset of our potential gene candidates. The progeny of these strains specify endoderm ~20% (MS1815) to 50% (MS1799, MS2330) of the time, where the percentage was calculated by total percentage of offspring that express the gut differentiation marker *elt-2*::GFP. Collectively these strains represent a background where we can perform a screen for genes that play a role in the embryo's ability to specify endoderm. The nature in which the strains were generated involves altering the interaction between specific regulatory events during early embryogenesis. Therefore, the strains allow us to identify possible suppressors of gene expression during that developmental stage.

Due to the mosaic nature of *C. elegans* development, cell specification occurs during an early time point of development. Therefore, we have selected to screen

through only maternal factors, since regulation of gene expression during that stage is activated by maternal factors deposited into the oocyte by the mother. Individual maternal genes where knocked down (strains taken from the VIDAL library) following the RNAi-by-feeding approach (Timmons, Fire 1998). *C. elegans* normally feed on bacteria, ingesting and grinding them in the pharynx and absorbing the bacterial content in the gut. Therefore, when fed on bacteria expressing dsRNA, the dsRNA will be absorbed and will subsequently activate the RNAi mediated gene silencing complex. This RNAi mechanism will deplete the maternal products from the gonad of *C. elegans*, thereby removing the protein products from the progeny.

In a pilot RNAi-by-feeding experiment approximately 10% of these clones were screened using the MS1815 strains. The MS1815 background is missing *end-3* (NULL) gene and has only a partially functional *end-1* gene that has mutated MED binding sites, an *elt-2*::GFP reporter, and an extrachromosomal rescue array that carries a copy of both *end-1,3* (+) and a *sur-5*::dsRed reporter marker, which expresses in every somatic nucleus starting in mid-embryogenesis. Only the wild-type-like rescued embryos will express both the rescuing array (marked by *sur-5*::dsRed) and the gut differentiation marker *elt-2*::GFP. For the analysis of the RNAi screen, only embryos that did not express the *sur-5*::dsRed rescue marker were scored for expression of *elt-2*::GFP, and our empty vector control bacterial strain showed that on average ~20% of non-rescued embryos expressed *elt-2*::GFP. We found several genes that, when knocked down, lead to an increase in the proportion of embryos expressing *elt-2*::GFP when compared to the

A



B

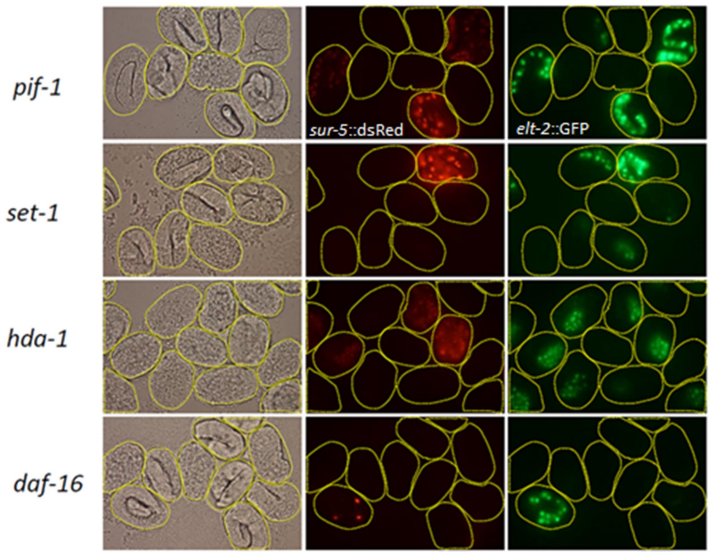


Figure 2-1 A) a pilot screen was conducted on ~300 bacterial clones to identify modifiers of gut specification. Reported is a subset of genes that showed an increase in the gut differentiation marker *elt-2*::GFP when knocked down, compared to the empty vector control. B) Embryos were collected from a seven-hour egg lay, and stored at 4°C overnight. Embryos were mounted on a standard agar pad and imaged. Embryos that express the *sur-5*::dsRed marker are rescue embryos and were omitted from the counts.

control. One of these gene, *hda-1* (Fig 2-1A, B), has been previously shown to cause an increase in gut specification when the endoderm network is compromised (Calvo et al., 2001; Raj et al., 2010). The pilot screen showed that our mutant strain is capable of detecting genes that alter gut specification, and validated that performing a high throughout screen using our strain was feasible.

The low overall percentage of embryos expressing *elt-2*::GFP from the MS1815 strain resulted in a biased selection for genes that resulted in a significant increase of *elt-2*::GFP when knocked down. Because the threshold for selection was ~20%, the strain was not sensitive enough to detect genes that decreased the expression of *elt-2*::GFP when knocked down. We decided to continue the screen using the strain MS1799, where the total percentage of embryos expressing *elt-2*::GFP hovers around 50%, which provided greater sensitivity to screen for genes that can increase or decrease expression of *elt-2*:GFP in a population of embryos. We note that this strain exhibits a tendency to undergo a phenotypic suppression after several months of laboratory propagation. In order to circumvent this problem, we regularly obtained stocks of the original isolates of these strains from frozen storage. Using the MS1799 strain, we conducted a primary reverse genetic screen utilizing RNAi-by-feeding and screened a total of 3,492 genes to identify factors that altered the percent of total embryos that express the gut differentiation marker *elt-2*::GFP when knocked down. 292 genes from our sample produced a sterile phenotype, which was expected since some maternal genes are known to affect the germline.

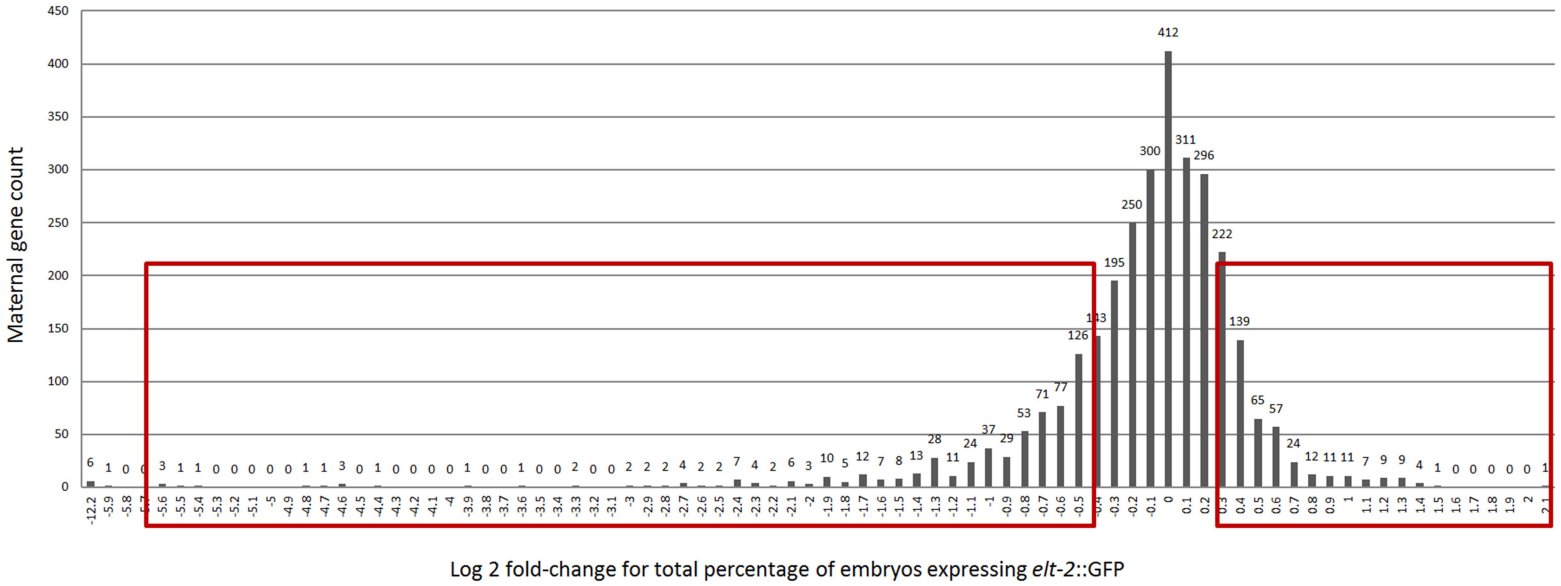


Figure 2-2 The total percentage of embryos expressing *elt-2*::GFP was standardize across all genes, by calculating the $\log_2$ fold-change value based on the percentage of *elt-2*::GFP expression yield from the control (empty vector) sample for that particular day. Arbitrary cut offs were selected at the $\log_2$ value $\leq$ -0.4 and $\geq$ 0.4, which translates to genes that have total percentage of embryos expressing *elt-2*::GFP at 30% and lower or 70% or greater.

For our control, the *end-1,3*(MED-) strain fed on the bacterial strain that contained an empty vector, ~50% of embryos expressed the *elt-2*::GFP reporter gene. We take that to mean that if ~50% of embryos expresses the *elt-2*::GFP marker, then there is no effect in buffering noise. This control was performed in parallel with each set of genes tested. Slight variations in the percent from week to week have been observed in the control; the change in expression was standardized by calculating the $\log_2$ fold - change for individual sets. The $\log_2$ value of genes whose *elt-2*::GFP percentage were the same as the empty vector control was 0 (~50%). This suggest that a $\log_2$ fold- change value of 1 would denote 100% of embryos expressing *elt-2*::GFP. To identify a list of potential candidate genes, we have arbitrary set a two-tailed selection of the $\log_2$ value $\leq$ -0.4 and $\geq$ 0.4, which translates to genes that have total percentage of embryos expressing *elt-2*::GFP at 30% and lower or 70% or greater (Fig 2-2).

As expected, a large proportion of genes (64%) that we screened did not have an effect on the total percent of embryos expressing *elt-2*::GFP. The pilot study and preliminary results suggest that our *end-1,3*(MED-) strains are suitable for a high throughput screen, and identified over 1,000 genes candidates of modifiers for gut specification. Of these, 214 putative modifiers led to an increase of overall percentage of embryos expressing the gut differentiation marker.

2.4.2 Go Term Analysis

To identify the types of modifiers that were identified from the primary RNAi screen to increase expression of the gut reporter, genes were grouped in order to identify common processes (Molecular Function, Cellular Component, Biological processes)

using DAVID (Huang da, Sherman, Lempicki 2009b; Huang da, Sherman, Lempicki 2009a) and PANTHER (Mi et al. 2013; Mi et al. 2017), on the 214 genes that lead to an overall increase of *elt-2*::GFP expression. Of the 214 genes, 211 were associated with a known function. The two software packages generated varying results. The DAVID analysis showed enrichment for genes that function in transcriptional regulation GO: 0045182, and the PANTHER analysis showed enrichment of genes that contain Catalytic Activity GO: 0003824 and Cellular Process GO: 0009987. However, both analyses displayed enrichment of genes that function in Apoptosis GO: 0006915, Metabolic Process GO: 0008152 and Binding GO: 0005488 (Fig. 2-3, 2-4 A, B, C). By combining the enrichment data from both programs, we are able to see a trend in the types of genes that were selected from the primary screen. The results suggest that the gut modifiers function to regulate the recruitment, binding, and catalytic activities of factors essential for transcriptional regulation of genes that play a role in metabolic pathways.

2.4.3 Validating Random subset of Gut modifiers using MS2330 mutant.

Following the functional analysis of our candidate genes, we tested the reproducibility of our results. It is known that there are biological variations even within isogenic populations, and environmental factors can also contribute to stochastic gene expression. First, we looked at the effect of RNAi knockdown on 17 genes over 4 separate trials on four consecutive days, to observe any day to day variations that might occur (Fig. 2-5). Overall, the results were reproducible in a large subset of the genes; however, we did observe some day to day variation between the samples. For the majority of genes, the results were reproducible for 3 out of the 4 trials. Due to

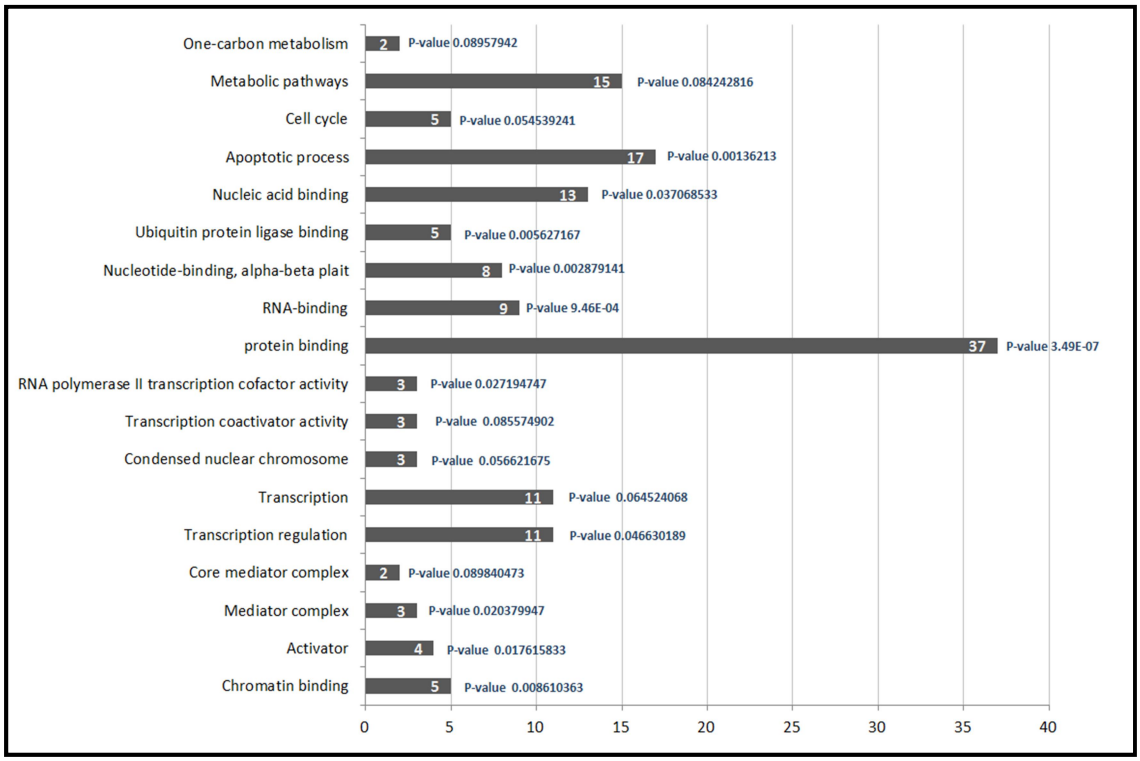
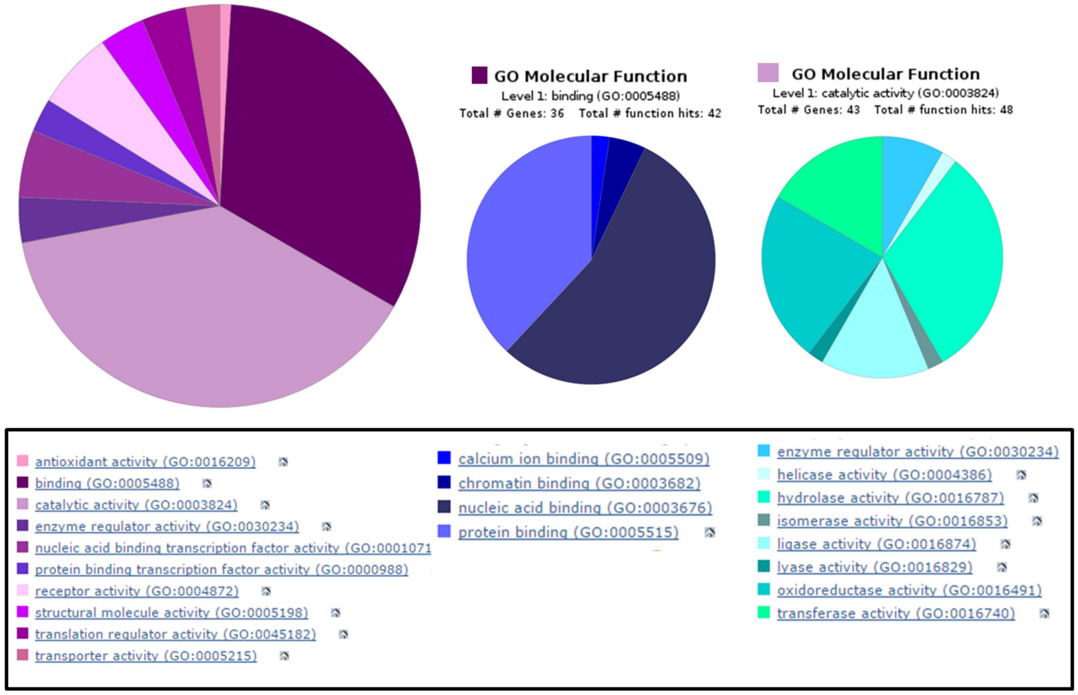
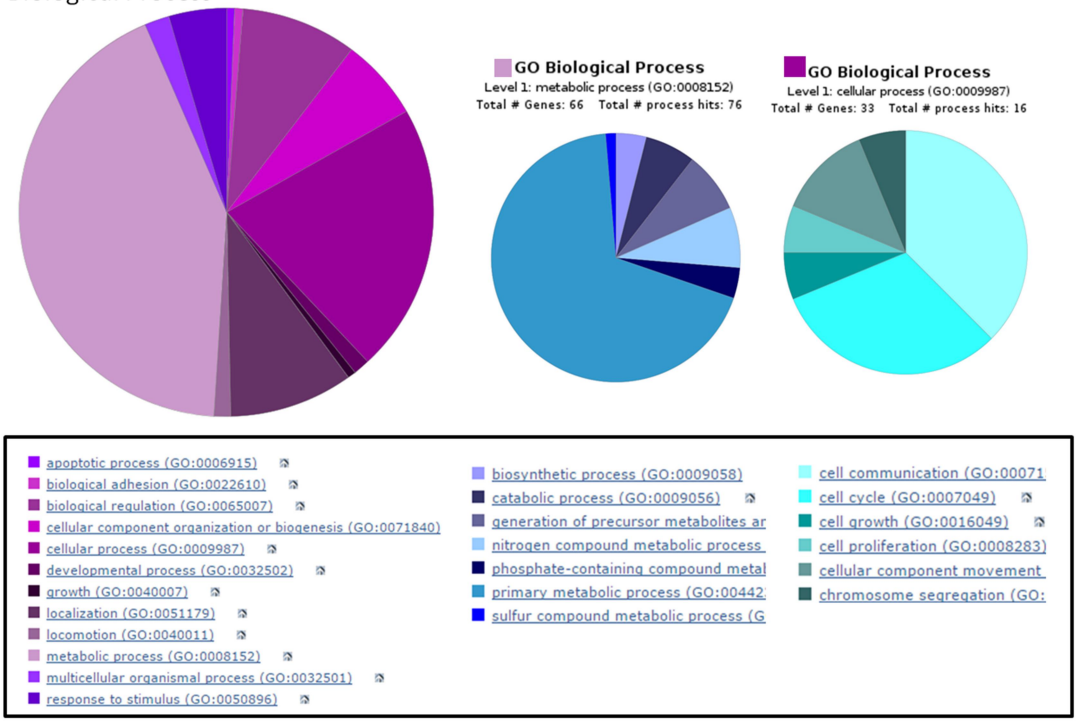


Figure 2-3 Categories identified using DAVID on 214 genes, to identify Molecular Function, Cellular Component, Biological processes. Enriched GO-terms, the number of genes associated to each GO Identifier and their p-values are reported above.

A Molecular Functions



B Biological Process



C

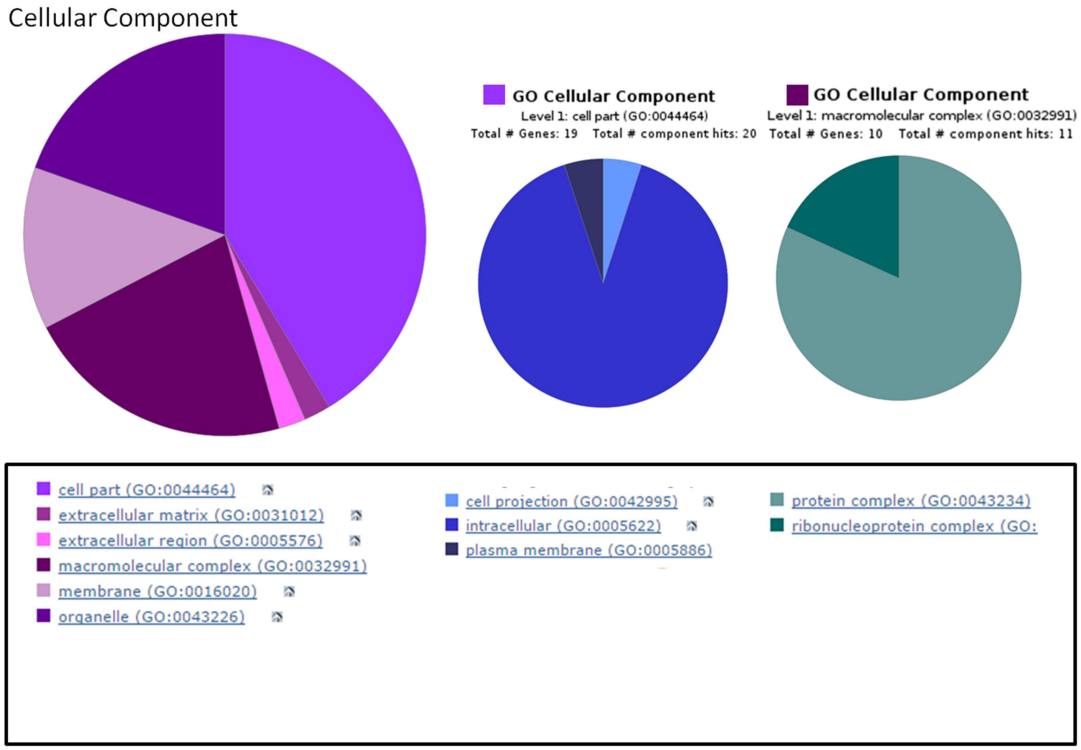


Figure 2-4 The PANTHER software was used to identify the A) Molecular function, B) Biological process, and C) Cellular components for 214 genes in form of pie charts. A, B, and C are represented in the same way, where the large pie chart on the right depicts the general GO identifiers that are enriched for their respective categories, and the smaller pie charts to the left depict the two most enriched GO identifiers and the more specific GO identifiers that belong to that category.

contamination on the feeder plates, some genes were scored out of 3 trials. The reproducible results from the knockdown of the *hda-1, pop-1*, and *rba-1* genes across 4 separate trials are consistent with previously published work (Shi, Mello 1998; Calvo et al. 2001; Maduro et al. 2005b; Owraghi et al. 2010; Raj et al. 2010). The knockdown of these genes has been reported to influence the specification of gut development. The data suggest that our screen can identify keys genes that modify gut specification within a noisy population and allow us to set the parameters to use for retesting a larger subset of genes.

We selected a random set of 117 genes that showed increased in overall GFP expression to evaluate reproducibility. Due to the suppression of phenotype seen in the MS1799 strain we transitioned to using the MS2330 strain for the validation screen. We found this strain to provide a more stable phenotype, with the total percentage of embryos expressing *elt-2*::GFP hovers around 40%. We have retested each of the 117 genes between 1-3 times, and have observed that ~17 % (20 genes) of the retested genes showed reproducible results upon validation (Table 2-1). Of the 20 genes that reproduced the phenotype, a GO-term analysis was conducted using WormMine. Due to the small number of genes, we were unable to perform GO-term enrichment using PANTHER or DAVID. Using WormMine, we were able to identify GO identifiers for each of the genes. The GO identifiers were manually sorted into functional groups, and the functional groups were similar to the functional groups identified from the 214 genes selected from the primary RNAi screen (Table 2-2). Although further screening needs to be conducted to provide a more confident readout, the current results we have from our

RNAi screen suggest that the modifiers of gut specification function in pathways in metabolism and global gene expression.

2.5 Conclusion

Using various *end-1,3*(MED) mutants strains we have conducted a reverse genetic screen to identify potential modifiers of gut specification. Our primary screen focused on identifying maternally provided gene products, due to the early specification of the endoderm lineage in *C. elegans*. This reverse genetic screen was able to detect changes in the expression of the gut differentiation marker, *elt-2*::GFP, in animals where the genes product of *hda-1*, *rba-1*, and *pop-1* were knocked down independently. The knockdown of these genes has been shown to have an effect on endoderm specification and our results confirm those findings (Shi, Mello 1998; Calvo et al. 2001; Maduro et al. 2005b; Owraghi et al. 2010; Raj et al. 2010). This suggests that our screen can allow us to utilize the endoderm GRN of *C. elegans* as a model, with which we can understand how it is buffered in living organisms.

From our primary screen we have identified 214 genes that when knocked down increase the total percentage of embryos expressing the gut differentiation marker. Through grouping by gene ontology, we found a large subset of these genes functioning in the regulation of gene expression, cell cycle, metabolism, and binding activities. This leads us to believe that modifiers of gut differentiation in a highly stochastic background may be buffered by genes that are necessary for the recruitment, binding, and initiation of transcription and translation of genes necessary for metabolic pathways and cell cycle progression.

Our validation screen attempted to collect data in triplicate and further screening stills needs to be conducted. Of the random 117 genes chosen to be retested, only a subset of 20 genes showed reproducible results. Of the genes that were retested, 33 of those genes were screen with our MS1815 strain. This particular screen had a low threshold of specification, where the controls displayed ~20% of embryos expressing *elt-2*:: GFP. Due to this low percentage, $\log_2$ fold-change thresholds are reached much more easily. The low threshold of the MS1815 strain and the variability observed between animals led to a large number of false positive results. By switching over to the MS1799 strain with a median threshold (~50%) we were able to get more consistent results between trials. However, this strain exhibits a tendency to undergo a phenotypic suppression after several months of laboratory propagation. Although we regularly obtained stocks of the original isolates of these strains from frozen storage, for the validation screen we decided to transition to the MS2330 strain, which had a lower tendency to undergo phenotypic suppression.

A proportion of the 20 genes identified from the validation screen functionally overlap in transcriptional/ translational regulation, cell cycle progression, metabolism, and apoptosis. We saw that knockdown of maternal genes involved in global gene regulation altered the proportion of embryos that specify gut in our compromised stains. This could be the results of removing genes products that normally function to repress excess gene expression, which allowed the embryos to conserve its resources. Our screen also identified genes that function in chromatin binding which can be interpreted as modification of histones, therefore the knockdown of these gene can result in open

chromatin structures, allowing transcription factors to interact with the promoter region of genes, and increase the proportion of embryos able to express gut specifying genes. While further screening still needs to be conducted, the results of our RNAi screen identified genes involved in metabolic pathways and gene expression to be regulators of gut specification.

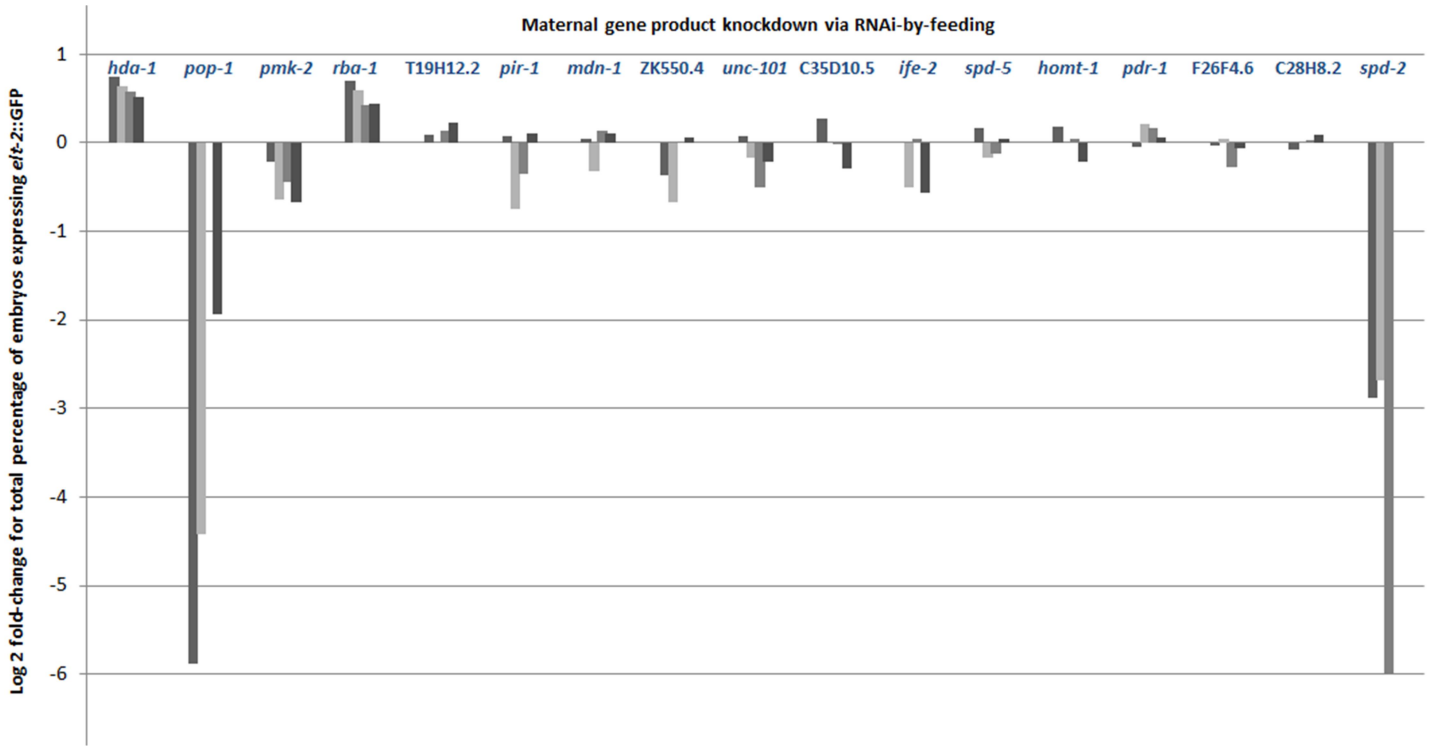


Figure 2-5 A subset of 17 genes were screen over 4 separate trials on four consecutive days, to observe any day to day variations. In particular, we selected genes such as *hda-1*, *pop-1*, and *rba-1* to test, based on their effect on gut specification when knocked down. Four separate experiments were conducted, but contamination of the feeder plates led to some genes being represented by only three data sets.

Gene ID	HHC 1	RA 1	RA 2	RA 3
	log2 FC	log2 FC	log2 FC	log2 FC
brc-1	0.5025	0.497476012	-	-
glh-1	0.641546	0.400650464	0.346710445	-
gst-1	1.104337	0.442964407	-	-
let-721	0.519374	0.377928192	0.472927461	-
mel-32	0.549339	0.236536604	0.567426289	0.504849477
ppk-2	0.549339	-0.438947309	0.454793487	-
psr-1	0.573185	0.145904877	0.063094301	0.52106915
ubc-16	0.925999	0.425112463	-	-
C56A3.6	0.408085	0.448006889	-	-
ekl-1	0.77259	0.481622751	-	-
F25H2.12	0.415037	0.532638711	-0.070470563	0.263385312
F26E4.2	0.450661	0.431552586	0.360794854	-
R09B3.2	0.847997	0.424219272	-	-
mrps-14	0.433897	0.06079578	0.304031123	0.501481848
T05H10.4	0.62293	1.09097366	0.994575713	-
Y43C5A.2	0.449803	0.136313791	0.313660479	0.464002438
sdha-2	0.443607	-0.906055944	0.550482644	0.288368888
F27C1.2	0.407424315	-0.392469742	0.421646956	-
brc-2	0.489038	0.304521752	0.230918616	0.560360528
unc-132	0.556393	0.413523901	-	-
arl-8	0.887525	-0.209051333	-	-
ccf-1	0.464668	0.115267263	0.052920314	-
ced-3	0.530515	-0.110975825	0.219737902	-
cit-1.1	0.678072	-0.096537609	-	-
cki-1	1.137504	-0.438121112	0.380565282	-
cyb-3	0.584963	-0.073172594	-	-
egl-18	0.443607	-0.04695728	-	-
eif-6	1.232661	-0.57572581	-0.462565869	0.059265949
gpa-4	0.519374	-0.364584497	-0.224573868	-
hil-5	0.925999	0.105208882	0.241023901	-
hpl-2	0.504473	0.020891685	-	-
mes-3	0.401363	-0.280900827	0.3671021	-
mif-3	0.584963	0.323604673	-0.029828579	0.127830416
ndx-4	0.485427	-0.37020832	-0.473233158	-0.059898063
ndx-7	0.415037	0.39502671	-0.117306115	0.386917637
phb-2	0.925999	-0.914433144	-0.09920746	-
pqn-68	0.415037	0.304912786	-	-
pqn-85	0.616671	-0.647463297	-	-

puf-8	1.115477	0.727920455	-	-
rab-10	1.070389	-0.398549377	-	-
rba-1	2.053111	0.08556119	-	-
rpl-5	0.807355	0.00737953	0.104325935	0.402802471
rpn-12	0.524662	-0.161683736	-	-
skr-15	0.925999	0.028956609	-	-
smu-1	0.443606651	0.352789059	0.073741545	-
ssq-1	0.584963	-0.181997834	-0.197647733	-
unc-37	0.481127	-0.381393954	-0.370860412	0.248091267
vps-26	0.7176	0.06029791	-	-
C01G10.10	0.702614	0.154290524	-	-
C04F12.5	0.584963	-0.142797823	-0.312048304	-0.286662663
C08B6.8	1.070389	-0.014666258	-	-
C14A4.11	0.542149	-0.117914203	-	-
har-1	1.070389	0.259358748	-	-
dhhc-7	1.035624	-0.612074346	-	-
C27D8.4	0.450661	0.263059242	-	-
dhfr-1	1.263034	0.331029219	-0.290318488	-
C38C10.2	0.475338	-0.244460884	0.028390432	-
C50B6.3	0.464668	0.079506673	-0.008007645	0.180453549
D2023.4	0.438884	0.23428621	-	-
F23B12.4	0.508147	0.214246098	-	-
F25H5.5	0.481127	-0.060606765	-	-
lsy-22	0.504043	-0.63606049	-0.2440298	-0.064969052
F35G2.1	0.485427	-0.143397303	0.186351733	-
F52A8.5	0.584963	0.002123172	-0.187891548	0.27918899
cdr-6	0.415037	0.190517763	-	-
K08F4.3	1.137504	-0.187891548	0.027282754	-
M04B2.4	0.584963	-0.323117828	0.35453276	-
gss-1	0.449803	0.320738362	-0.252501477	-
prdx-3	1.104337	-0.345173839	-	-
nrde-2	0.545968	-0.446227463	-	-
mrpl-51	0.400538	0.066512989	-	-
T15H9.2	0.707819	-0.865650771	-0.113684337	-
T22C1.3	1.047306	0.319416876	-0.255941608	0.200375514
Y11D7A.3	0.462343	-0.068356191	0.29370763	-
Y43F8B.2	0.517399	0.364402951	-	-
Y45F10A.3	0.647698	0.291447419	-0.00345005	-
Y52B11A.8	1	0.332690639	-	-
Y57A10A.10	0.470891	0.069664645	-	-

Y105E8A.21	0.575502	0.011893789	-	-
atfs-1	0.556393	0.39828815	-	-
ZK632.12	1.137504	-1.400475205	0.185568624	0.347923303
B0281.5	0.765535	-0.188989907	0.236200483	-
B0304.2	1.263034	0.375243909	0.125308227	-
C01B10.3	0.619728	0.070919623	-0.084900439	-
C01B10.9	0.530515	-0.125249956	0.300709228	-
C06E1.9	0.485427	-0.280134657	0.25788897	-
egal-1	0.481869	0.134347992	-	-
flu-2	0.596644	0.095671807	-0.087527236	-
C24D10.6	0.526069	0.107464697	-	-
C32E8.3	0.765535	0.271746575	-	-
C45G9.2	0.456378	-0.122745383	-	-
phat-3	1.350497	-0.072695228	-	-
F11G11.5	1.104337	0.095977806	-	-
F13A2.4	0.438884	0.069962284	-	-
bra-2	1	-0.626636499	-	-
gcst-1	0.619728	0.015941544	0.206271091	-
F25B4.4	1.169925	-0.302054994	-	-
F32D1.5	0.685127	0.278832649	-	-
F42A9.6	0.400538	-0.150628967	-	-
F45F2.11	0.524662	-0.036631271	-	-
mdh-1	1.201634	0.251684101	-	-
F55G1.9	0.433071	-0.662130361	-0.138994024	0.347268872
F56C9.6	0.527932	0.010676001	-	-
K07H8.3	0.678072	-0.983342656	-	-
K10C9.3	0.508147	0.128590014	-	-
fbxc-47	0.450661	0.139602092	0.282181916	-
fbxc-36	0.632268	-0.328012857	-	-
hrpf-1	0.584963	0.053599389	0.168149776	0.136650391
W08F4.3	1.405992	-0.012627608	-0.080768404	-
Y4C6B.4	0.678072	0.192645078	-0.046223017	-
eel-1	0.616671	-0.391774188	-	-
ZK112.5	0.584963	-0.058905231	-	-
npp-24	0.584963	-0.382038546	-	-
ubxn-4	0.449803	0.178133967	0.095025884	0.398549377
ZK546.2	1.321928	-1.161596642	-0.110016632	-
tag-322	0.574236	0.224775073	-	-
szy-4	0.659925	-0.142516678	-0.327887911	-

Table 2-1 A random selection of 117 genes was re-screened three times to test for reproducibility. The screen is currently incomplete. For each gene, the $\log_2$ value for the primary screen is represented in the second column, as well as current data collected, which is represented in columns 3-5. The genes shaded in green represent genes that show reproducibility so far.

Functions	Genes
Transcriptional/Translational regulation	*brc-1, glh-1, ppk-2, psr-1, ekl-1*, R09B3.2, *mrps-14, brc-2*
DNA repair/ Cell Cycle	*brc-1, glh-1, ekl-1, brc-2*
Cellular respiration	*psr-1, sdha-2*
Apoptosis/ Cell death	brc-1, gst-1, psr-1
Metabolic	*gst-1, mel-32, ppk-2*

Table 2-2 Gene names were inputted into the WormMine software on WormBarse.org, and a list of GO identifiers associated with each gene was provided. Genes were manually grouped base on similarity of function.

2.6 References

Bowerman, B, BW Draper, CC Mello, JR Priess. 1993. The maternal gene skn-1 encodes a protein that is distributed unequally in early C. elegans embryos. Cell 74:443-452.

Calvo, D, M Victor, F Gay, G Sui, MP Luke, P Dufourcq, G Wen, M Maduro, J Rothman, Y Shi. 2001. A POP-1 repressor complex restricts inappropriate cell type-specific gene transcription during Caenorhabditis elegans embryogenesis. EMBO J 20:7197-7208.

Chalancon, G, CN Ravarani, S Balaji, A Martinez-Arias, L Aravind, R Jothi, MM Babu. 2012. Interplay between gene expression noise and regulatory network architecture. Trends Genet 28:221-232.

Choi, H, G Broitman-Maduro, MF Maduro. 2017. Partially compromised specification causes stochastic effects on gut development in C. elegans. Dev Biol 427:49-60.

Davidson, EH. 2010. Emerging properties of animal gene regulatory networks. Nature 468:911-920.

Fraser, AG, RS Kamath, P Zipperlen, M Martinez-Campos, M Sohrmann, J Ahringer. 2000. Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. Nature 408:325-330.

Gursky, VV, SY Surkova, MG Samsonova. 2012. Mechanisms of developmental robustness. Biosystems:329-335.

Huang da, W, BT Sherman, RA Lempicki. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1-13.

Huang da, W, BT Sherman, RA Lempicki. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44-57.

Kormish, JD, J Gaudet, JD McGhee. 2010. Development of the C. elegans digestive tract. Curr Opin Genet Dev 20:346-354.

Macneil, LT, AJ Walhout. 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res 21:645-657.

Maduro, MF, RJ Hill, PJ Heid, ED Newman-Smith, J Zhu, JR Priess, JH Rothman. 2005a. Genetic redundancy in endoderm specification within the genus Caenorhabditis. Dev Biol 284:509-522.

Maduro, MF, JJ Kasmir, J Zhu, JH Rothman. 2005b. The Wnt effector POP-1 and the PAL-1/Caudal homeoprotein collaborate with SKN-1 to activate C. elegans endoderm development. Dev Biol 285:510-523.

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

McGhee, JD, T Fukushige, MW Krause, 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.

Mi, H, X Huang, A Muruganujan, H Tang, C Mills, D Kang, PD Thomas. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res 45:D183-D189.

Mi, H, A Muruganujan, JT Casagrande, PD Thomas. 2013. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc 8:1551-1566.

Owraghi, M, G Broitman-Maduro, T Luu, H Roberson, MF Maduro. 2010. Roles of the Wnt effector POP-1/TCF in the C. elegans endomesoderm specification gene network. Dev Biol 340:209-221.

Raj, A, SA Rifkin, E Andersen, A van Oudenaarden. 2010. Variability in gene expression underlies incomplete penetrance. Nature 463:913-918.

Raj, A, A van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135:216-226.

Rual, JF, J Ceron, J Koreth, et al. 2004. Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. Genome Res 14:2162-2168.

Shi, Y, C Mello. 1998. A CBP/p300 homolog specifies multiple differentiation pathways in Caenorhabditis elegans. Genes Dev 12:943-955.

Sommermann, EM, KR Strohmaier, MF Maduro, JH Rothman. 2010. Endoderm development in Caenorhabditis elegans: the synergistic action of ELT-2 and -7 mediates the specification-->differentiation transition. Dev Biol 347:154-166.

Stiernagle, T. 2006. Maintenance of C. elegans. WormBook:1-11.

Sulston, JE, E Schierenberg, JG White, JN Thomson. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol 100:64-119.

Timmons, L, A Fire. 1998. Specific interference by ingested dsRNA. Nature 395:854.

Updike, DL, SE Mango. 2007. Genetic suppressors of Caenorhabditis elegans pha-4/FoxA identify the predicted AAA helicase ruvb-1/RuvB. Genetics 177:819-833.

CHAPTER 3

Partially compromised specification causes stochastic effects on gut development in *Caenorhabditis elegans*

3.1 Abstract

The intestine of the model organism *Caenorhabditis elegans* (*C. elegans*) starts off as a single progenitor E cell, from which the entire gut is derived through a series of stereotyped cell divisions and morphogenetic events. The consequences of mis-regulation of key specification factors on genes upstream of cell specification have not been extensively investigated. Our lab has built an allelic series of mutants that is compromised in the specification of endoderm by preventing the activation of the gut specifying genes *end-1* and *end-3*. Using two reporters which mark all E descendant cells with an *end-3*::mCherry reporter, and an *elt-2*::GFP marker that identifies all cells that have adopted an endoderm fate, we have examined the fate and cell position of E descendants among hundreds of embryos. We find that when specification is partially compromised, the activation of the gut differentiation factor *elt-2* becomes delayed in these strains, but ultimately the protein levels of a translational ELT-2::GFP reporter are similar to wild type levels. Our data suggest that he fewer number of embryos specifying gut, the stronger the defects in the E lineage and delay in activation of *elt-2*.

3.2 Introduction

The formation of organs or "organogenesis" is a crucial part of the developmental process in metazoans. During this developmental time frame, several essential components and distinct processes must occur. This includes activation of gene regulatory networks in progenitor cells to drive tissues specification, correct timing of mitotic divisions, cell migration to the proper locations and morphogenesis to form normal organ shape and size, and lastly activation of genes necessary for tissue differentiation. These systems are exposed to various extrinsic and intrinsic factors that can create variation in gene expression during this process. However, robust organ formation persists even in the presence of such noise. Embryonic gene networks are comprised of feed-forward and autoregulation loops that enforce downstream gene expression and help maintain cell identity (Davidson 2010). Global and local mechanisms influence cell proliferation and ultimately regulate the size of organs (Hariharan 2015; Irvine, Harvey 2015; Patel, Camargo, Yimlamai 2017). Due to the robustness of the system, experimental perturbation of early steps of organogenesis might not lead to phenotypic effects on later organ development.

The nematode *C. elegans* are mosaic developers; they undergo cell specification during early stages of embryogenesis and are unable to replace missing somatic cells. The gut is generated from a single progenitor cell called E (Fig. 3-1A). In wild-type animals, the single E cell undergoes a few rounds of stereotyped cell division, to produce the 20 guts cells that make up the entire intestine at the end of embryogenesis. Occasionally,

some wild-type animals will hatch with 21-22 gut cells (Sulston et al. 1983) (Fig. 3-1B). The E cell lineage produces the entire gut organ in wild-type *C. elegans* with little variation between individuals (Asan, Raiders, Priess 2016). Since the E cell lineage produces only one tissue type, and that development occurs in a very predictable way between individual *C. elegans,* the nematode model is exceptionally suited for studying the relationship among cell specification, fate, and lineage (Sulston et al. 1983; Boeck et al. 2011; Maduro 2017).

The key components for gut specification are a pair of redundant paralogue genes, *end-1* and *end-3*. Deletion of both genes results in failure of the E cell lineage to specify endoderm (Maduro et al. 2005; Owraghi et al. 2010). Because *C. elegans* are mosaic developers, other cell lineages in the embryos are still able to develop normally even in the absence of the gut (Maduro 2009; Owraghi et al. 2010). Expression of the *end-1* and *end-3* genes only occurs for a short period of time and regulates the expression of *elt-2* and *elt-7*, which maintain their own expression through autoregulation. The activation of *elt-2* is essential in gut differentiation, by activating all downstream gut-specific genes (Fukushige, Hawkins, McGhee 1998; Sommermann et al. 2010). In the absence of *elt-7* there are no phenotypic effects, but the absence of *elt-2* results in abnormal gut development. *Elt-2* and *elt-7* have synergistic effects, such that the typical effects of the loss of *elt-2* are increased in the absence of *elt-7* (Fukushige, Hawkins, McGhee 1998; Sommermann et al. 2010). Due to the same transcription factor activity and the ability to bind to the same HGATAR consensus sequence, the four previously mentioned genes are

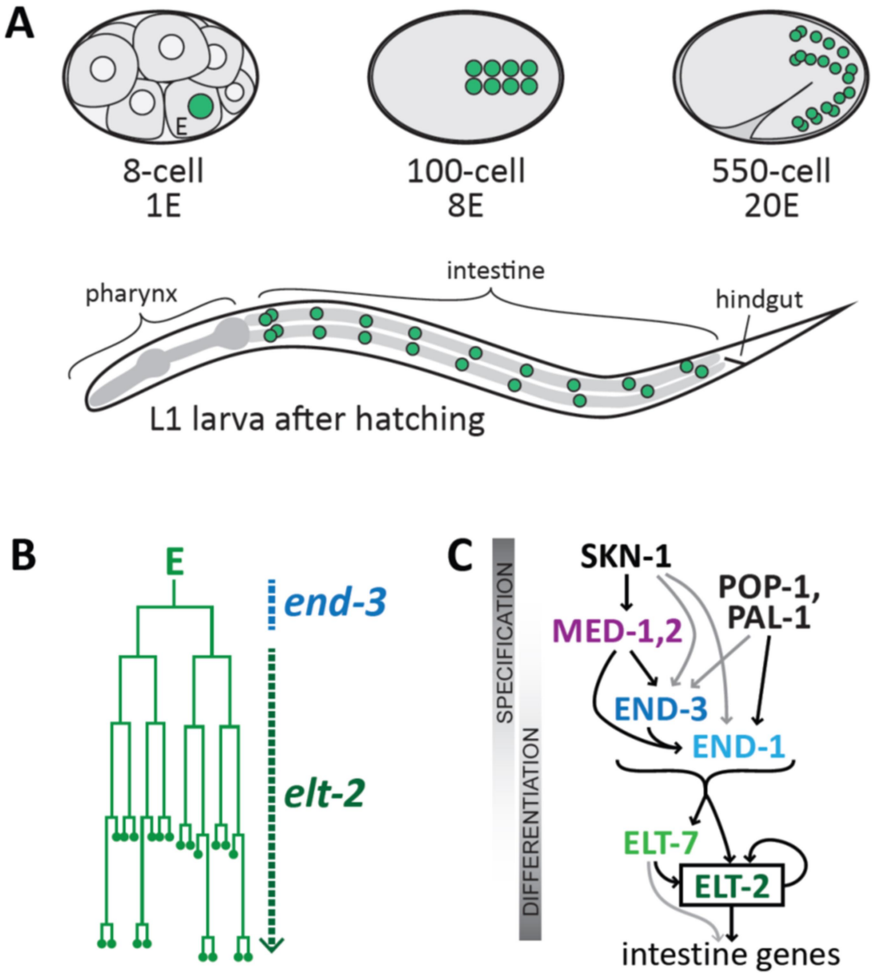


Figure 3-1. (A) The locations of the E lineage nuclei at various stages in normal development, modified from (Maduro et al. 2015). Overall embryogenesis takes ~12 h at 25 °C (Sulston et al., 1983). (B) Cell division pattern of the wild-type E progenitor cell, after (Sulston et al., 1983). The vertical axis is time and a horizontal line indicates a cell division. The approximate time of transcription of *end-3* and *elt-2* is shown to the right of the diagram. (C) Simplified pathway showing hierarchy of transcription factors, modified from (Maduro et al. 2015). The endoderm specification strains described in this work perturb the overall contributions made by the MED-1,2 and END-1,3 regulators in a way that does not affect other lineages. Embryos are ~50 µm long and the larva is ~200 µm long. Dorsal is at top and anterior is to the left

grouped together as the endodermal GATA factors' and can undergo gut specification when overexpressed ectopically. *Elt-2,* driven under the *end-1* promoter, is even capable of driving gut specification in the absence of the other three genes (Zhu et al. 1998; Maduro et al. 2005; Sommermann et al. 2010; Wiesenfahrt et al. 2015; Du, Tracy, Rifkin 2016). The endoderm gene regulatory network is filled with redundancy, and is made up of many parallel maternal and zygotic inputs upstream of the endoderm specifying genes, contributing to their timely activation. These upstream genes also play an essential role in other cell lineages, where deletion of those genes will cause arrested development with abnormal morphology (Bowerman, Eaton, Priess 1992; Lin, Thompson, Priess 1995; Hunter, Kenyon 1996).

It is crucial to understand how developing organisms manage stochastic gene expression. In many systems, variation in gene expression occurs due to both intrinsic and extrinsic factors (Blake et al. 2003; Colman-Lerner et al. 2005; Raj et al. 2010; Holloway et al. 2011). The redundant nature of the endoderm gene regulatory network, in terms of its gene duplication and multiple sources of activation for a single gene, makes gut specification a robust process (Maduro et al. 2007; Maduro et al. 2015). The layers of redundancy allow for researchers to build strains that are partially compromised in gut specification, resulting in mutants that have variable gene expression. In such genetic backgrounds, called "Hypermorphic Gut Specification" (HGS) strains, gut development becomes highly stochastic. Using an *elt-2* GFP transcriptional reporter, we are able to identify that the number of gut nuclei ranges from 0 to over 30 (Maduro et al. 2007; Maduro et al. 2015). These results suggest that gut specification is not an all-or-none

event at the level of the E-blastomere. We have proposed that this variable phenotype is the result of two concurrent effects: One effect is the increase in the number of gut cells resulting from the progenitor E cell, and the other is the stochastic adoption of gut fate among those descendants (Maduro et al. 2015).

HGS strains were imaged during the mid-embryo stage, to analyze their effects on the E lineage. We used two reporters in conjunction to allow us to identify the E lineage cells and the E lineage descendants that have also adopted a gut fate. We find that HGS strains do produce extra E descendant cells due to extra divisions within the E lineage. There is an inverse relationship between defects in the gut specification and the proportion of cells that commit to a gut fate; as the severity of defect increases, the number of gut cells decreases. Among the HGS animals that survive to adulthood, a similar number of gut nuclei is observed across all strains. This is consistent with the minimum number of gut cells necessary for survival to the first larval stage. The delayed expression of *elt-2* correlates with the severity of the HGS strain; the sicker the animal, the later *elt-2* expression appears. However, in the surviving HGS animals, the final levels of *elt-2* expression are essentially normal. The results suggest that timely activation of gut specification is necessary for correct patterning of E-lineage cell divisions and for the E descendants to the commit to the gut fate.

3.3 Materials and methods

3.3.1 Strains and worm handling

The wild-type control was N2. All strains were grown on E. coli OP50 and maintained at 20–22°C and observed at 23–25 °C except as noted. Mutations and transgenes were as follows. LG I: *cdc-25*.1(rr31), irSi10 [*end-3*(MED sites mutated) + Cb-*unc-119*], irSi13 [*end-3*(+) + Cb-*unc-119*(+)]. LG II: irSi7 [*end-1*(MED sites mutated), Cb-*unc-119*(+)], irSi9 [*end-1*(+), Cb-*unc-119*(+)], irSi12 [*end-1,3*(+), Cb-*unc-119*(+)]. LG III: *med-2*(cxTi9744). LG IV: *him-8*(e1489), *him-8*(me4),irIs98 [Cb-*unc-119*(+), *end-1*(MED sites mutated), *end-3*(MED sites mutated), irSi24 [*pept-1/opt-2*::mCherry::H2B, Cb-*unc-119*(+)], itIs37 [Cb-*unc-119*(+), *pie-1*::H2B ::mCherry]. LG V: *end-1*(ok558), *end-3*(ok1448), *end-3*(ir62), *end-3*(ir64), oxTi389 [Cb-*unc-119*(+), *eft-3*::H2B::dTomato], stIs10116 [Cb-*unc-119*(+), *his-72*::H2B::mCherry], zuIs70 [*end-1*::GFP::CAAX], irIs133 [*elt-2*::mCherry::H2B, *unc-119*::CFP, *rol-6*D]. LG X: *med-1* (ok804), rrIs1 [*unc-119*(+), *elt-2*::NLS::GFP::lacZ], wIs84 [*rol-6*D, *elt-2*::NLS::GFP::lacZ]. Unmapped: stIs10064 [Cb-*unc*-119(+), *end-3*::HIS-24::mCherry::let-858_3'UTR], gaIs290 [*elt-2*::ELT-2::TY1::EGFP::3xFLAG]. Extrachromosomal arrays: irEx697 [*elt-2*::mCherry::H2B, *unc-119*::CFP, *rol-6*D]. Mutations and transgenes were combined using standard crosses. The *cdc-25.1*(rr31) strain was grown at 23–25 °C. Because of the close proximity of zuIs70 to *end-3* (< 2 mapunits), we made a de novo null mutation (ir64) in *end-3* in a zuIs70 strain using CRISPR/Cas9-mediated mutagenesis (Arribere et al. 2014) with guide RNAs with targeting regions 5′-aaacacgtgaaatttagag-3′ and 5′-tcgggaaacgaaattgtgg-3′, which delete

Table 3-1 A set of Zygotic endoderm specification strains of varying

Strain	Genotype	Embryos making gut[1]	Viability[2]
E(100) / N2	wild type	100% (n>500) (control)	100% (n>200)
E(100) / MS1810[3]	*end-1(ok558) end-3(ok1448); Si[end-1(+)]; Si[end-3(+)]*	100% (>500) (control)	100% (n>200)
E(95) / RB1331	*end-3(ok1448)*	96% (221)	81% (140)
E(95) / MS2267	*end-3(ir62)*	95% (232)	n.d.
E(75) / MS1809[3]	*end-1(ok558) end-3(ok1448); Si[end-1(MED-)]; Si[end-3(MED-)]*	75% (459)	42% (120)
E(50) / MS1548	*end-1(ok558) end-3(ok1448); irIs98 [end-1(MED-), end-3(MED-)]*	48% (143)	39% (140)
E(40) / MS404[4]	*med-1(ok804); end-3(ok1448)*	42% (251)	23% (239)
E(30) / MS1763[4,5]	*end-1(ok558) end-3(ok1448); Si[end-1(MED-)]*	28% (98)	<5% (>100)
E(0) / MS1762[3,5]	*end-1(ok558) end-3(ok1448); Si[end-3(MED-)]*	0% (93)	0%
E(0) / MS1248[5,6]	*end-1(ok558) end-3(ok1448)*	0% (190)	0%

[1]Gut was scored by presence of any amount of birefringent gut granules.
[2]Percentage of eggs laid that survive to adulthood.
[3]Strain described in Maduro et al. (2015).
[4]Strain described in Maduro et al. (2007).
[5]These strains were maintained by an extrachromosomal array balancer carrying *end-3(+)* and a broadly-expressed *sur-5*::dsRed fluorescent reporter. Progeny embryos lacking the reporter were scored for the presence of gut.
[6]Strain described in Owraghi et al. (2010).

1.1 kbp of the *end-3* locus including its promoter and most of the END-3 coding region. The ir64 lesion is identical to that of ir62 which was made in an N2 background (strain MS2267). Mutations in *him-8* were used to generate males, and the resulting strains were confirmed to be lacking *him-8* by the absence of males over several generations. To facilitate recovery of some strains, mapped transgene insertions were used as counter-selection markers. Recovery of strains was confirmed by PCR, reporter expression and/or expected phenotype. Some strains may carry a background *unc-119* mutation whose phenotype is rescued by the presence of one or more integrated transgenes carrying *unc-119(+).*

3.3.2 Construction of specification-compromised strains

Endoderm specification strains are listed in Table 3-1. Construction of MS1762, MS1763, MS1809, and MS1810 was described previously (Maduro et al. 2015). MS1548 was produced as follows. Plasmid pMM869 was constructed, which carries Cb-*unc-119(+)* and the *end-1* and *end-3* genes with mutations in the MED binding sites as described (Maduro et al. 2015). pMM869 was integrated into the genome by micro-particle bombardment of an *unc-119*(ed4) strain (Praitis et al. 2001). The resulting integrated line was crossed into an *unc-119*(ed4); *end-1*(ok558) *end-3*(ok1448); Ex[*end-3*(+), *sur-5*::dsRed] strain and F2 progeny were obtained that failed to segregate *Unc* and which lacked the *end-3*(+) array. The resulting strain, *end-1*(ok558) *end-3*(ok1448); *unc-119*(ed4); irIs98 was saved as MS1548. We note that both the MS1548 and MS404 strains, and others derived from them, exhibit a tendency to undergo a phenotypic suppression after several months of laboratory propagation. We regularly obtained stocks

of the original isolates of these strains from frozen storage, especially for work involving quantification of traits that tracked with the severity of the gutless phenotype. We performed detailed analyses on only a subset of the HGS strains as effects were generally similar across all strains in preliminary assays.

3.3.3 Microscopy, imaging and data analysis

Conventional epifluorescence and differential interference microscopy were performed on an Olympus BX-51 microscope, imaged through an LMscope adapter (Micro Tech Lab, Graz, Austria) and Canon Rebel T1i Digital Camera using software supplied with the camera. Confocal Microscopy was performed on a Zeiss 510 LSM at the UC Riverside Microscopy Core. To obtain staged embryos for the analyses in Fig. 3-6 and 3-8, gravid hermaphrodites were dissected with a 25- gauge needle and 4-cell stage embryos were collected and placed on an unseeded 6-cm agar plate and allowed to develop for 5 h in a 20 °C incubator. For microscopic observation, embryos were mounted as described (Bao, Murray 2011). Images were combined, adjusted for color/contrast and cropped using FIJI (ImageJ) and Adobe Photoshop. To generate heat maps for Fig. 3-8, the summed Z-stack red channel (mCherry) images were converted to binary format with FIJI, where the red signal was converted to black, and the background to white. These images were compiled into a stack with the Images to Stack tool. Heat maps were generated with the heat map tool in FIJI on the newly built stacks. Plots were generated in ggplot2 in an R environment (http://ggplot2.org) or Microsoft Excel.

3.4 Results

3.4.1 An allelic series of strains affecting specification of (only) the gut

We have built a series of strains designed to interfere with specific MED-END association in the early steps of the endoderm gene regulatory network, which do not affect other cell lineages Fig. 3-1C. Of our series, two strains cannot specify gut, while our control strains adopt a gut fate 100% of the time. We categorize the strains where only a proportion of embryos can make gut as "Hypomorphic Gut Specification" (HGS) strains, which has a range of 28%- 96% of embryos making gut (Table 3-1). The HGS strains are made up of chromosomal mutants, as well as single insertion of a single copy transgene to a null mutant background. Chromosomal mutants were generated from either single or double null mutation of *end-3* (ok1448) or double mutant of *med-1*(ok804) and *end-3* (ok1448) respectively (Maduro et al. 2005; Maduro et al. 2007). The rest of the strains where generated from the *med-1*(804) and *end-3*(ok1448) double null mutant background, with single copy insertions of transgene *end-1* and/ or *end-3* with mutations in several MED- 1,2 binding sites (Maduro et al. 2015). To reference individual strains of our allelic series, we used the notation E(X) to indicate a specific strain, where (X) represents the percentage of embryos that make gut, rounded to the nearest 5%. Using this set of strains, we will explore how a slight change in the E lineage development associate with effects of gut development and specification.

3.4.2. Partial-specification strains display highly stochastic E lineage

First, we used an integrated nuclear-localized *elt-2*::GFP reporter (wIs84), to access the number of gut nuclei in individual animals across our HGS strains and to correlate the changes in the E lineage with the varying degree of partial gut specification in the animals. The control HGS strain, E(100) displayed 20.00 ± 0.5 SD *elt-2*::GFP expressing nuclei, consistent with wild-type *C. elegans* (Table 3-2; Fig. 3-2, 3-3). From all HGS strains that expressed at least one gut nucleus, we scored the amount of *elt-2*::GFP expressing nuclei present during late embryo stage, and saw that all HGS strains exhibit a wide range in number of gut nuclei produced. On average the number of nuclei across the strain varied, from 7.3 ± 5.3 for the more severe mutant strain E(30) to 19.7 ± 6.7 for the less severe strain E(95). The averages correlated well with overall percentage of embryos that made gut in the strains (R2 =0.84; note that the horizontal scale is not uniform). In analyzing all HGS strains, our data supports the idea that the mere presence of gut-like cells in the embryo, does not imply the gut will be normal. The data also suggests that the specification of gut cells is not an "all-or-none" event, otherwise every embryo that adopted a gut fate should have between 20-22 *elt-2*::GFP expressing nuclei at the late embryo stage.

A minimum number of gut cells is necessary for the survival of the animal, as embryos with low numbers of gut nuclei arrest during embryogenesis (Fukushige, Hawkins, McGhee 1998; Owraghi et al. 2010). We were interested in the variation of gut

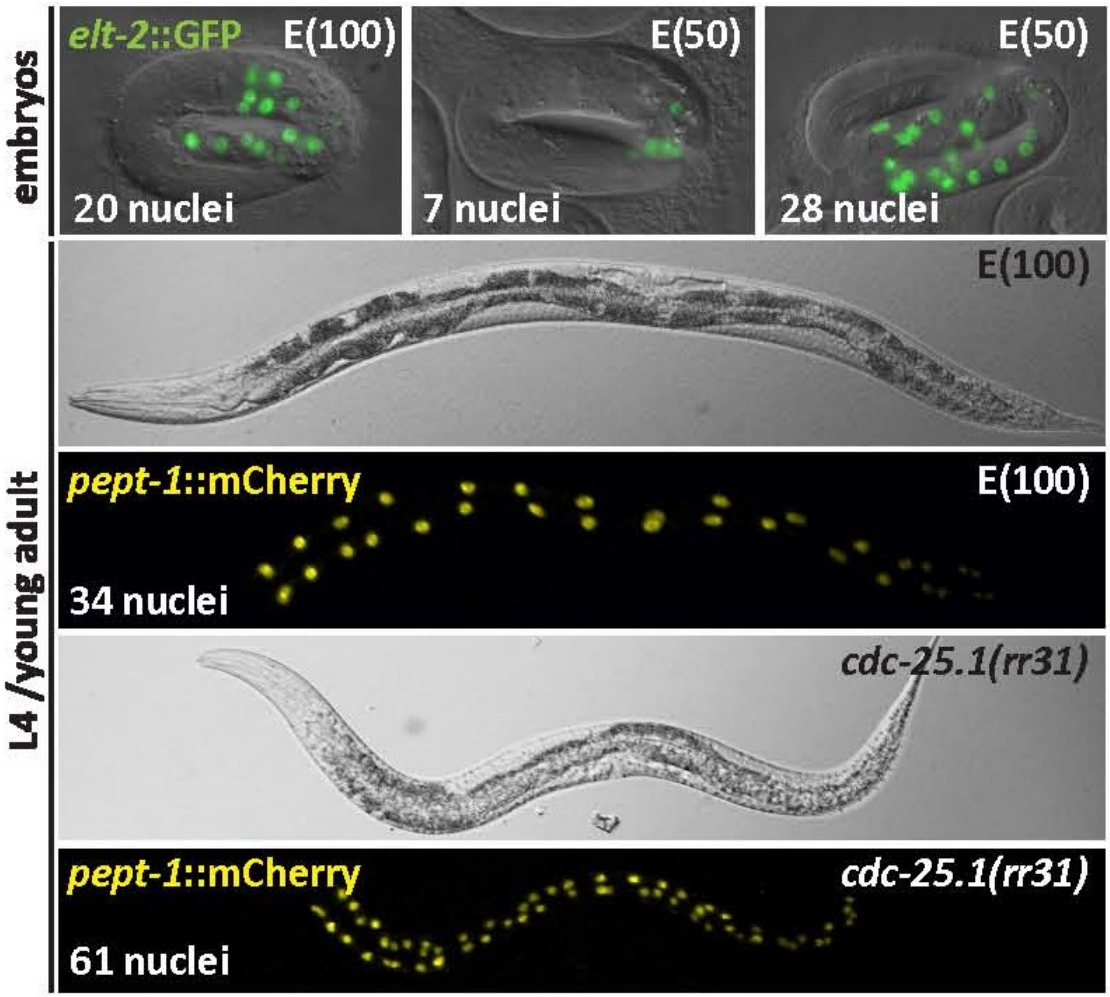


Figure 3-2. Examples of embryos (top row) and L4 animals (bottom two images). Gut nuclei were counted in living embryos homozygous for a nuclear-localized *elt-2*::GFP transcriptional reporter (wIs84), and in L4 animals with a single-copy nuclear-localized *pept-1*::mCherry reporter (irSi24), pseudocolored yellow. The embryo images are overlaid with a DIC image of the same embryo. An embryo is approximately 50 µm along its long axis, and the L4 intestines are approximately 750 µm long. Anterior is to the left.

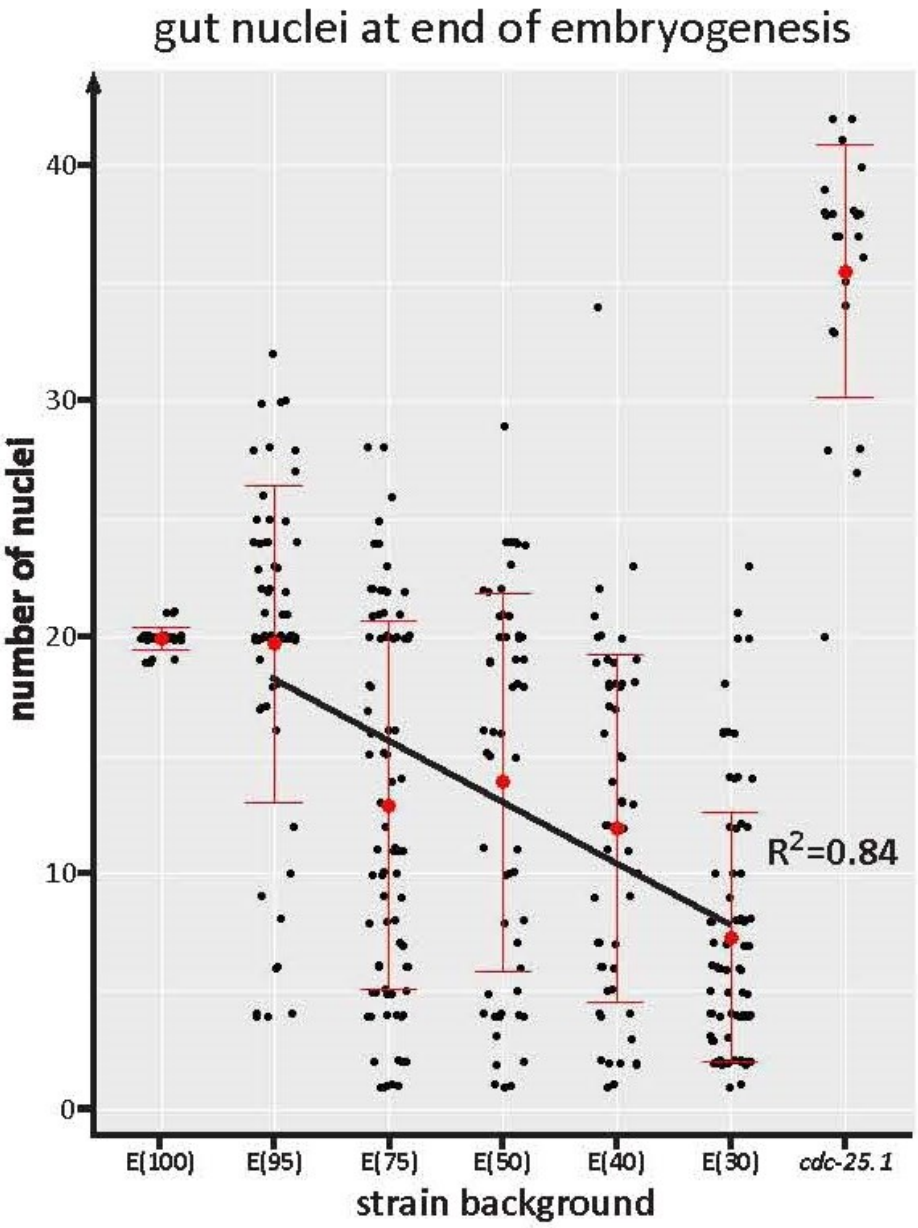


Figure 3-3. Perturbation of specification causes aberrant numbers of gut nuclei to form, in a quantity proportional to the severity of the defect in specification. We note that the proportion of gut specification across the X-axis is not linear as shown.

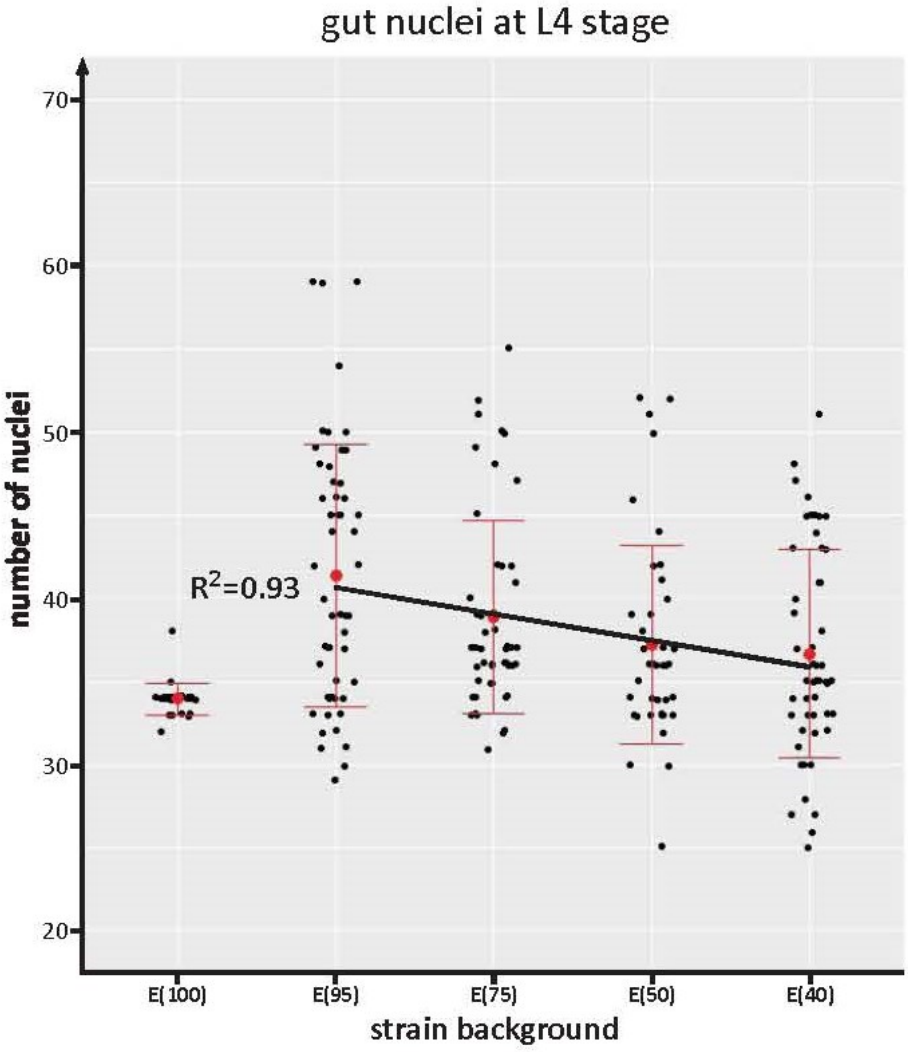


Figure 3-4. Surviving L4 animals display a similar range of aberrant gut nuclei that may be only mildly correlated with severity of specification. All four of the specification strains had gut nucleus numbers significantly different from the control ($p < 0.004$, t-test). Compared with end-3, E(75) was not significantly different ($p=0.07$) while the other two strains were ($p < 0.005$). Pairwise differences among E(75), E(50) and E(40) were not significant ($0.06 < p < 0.2$).

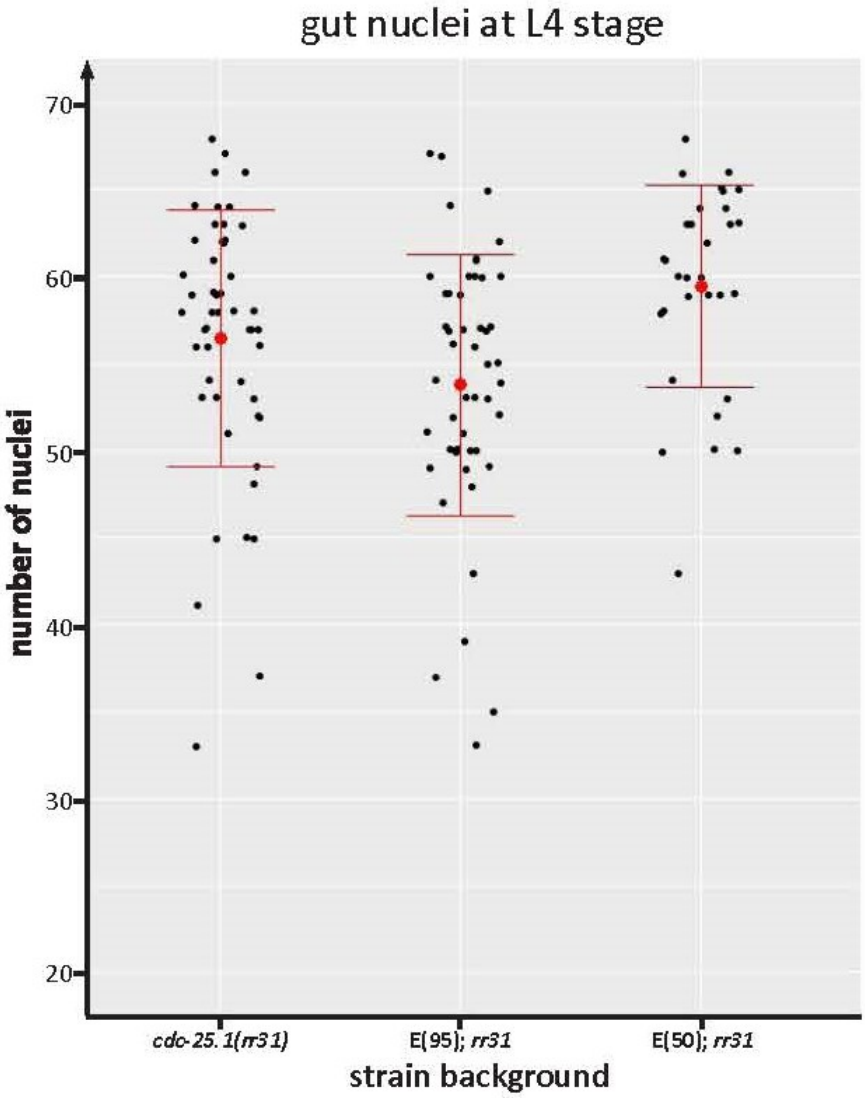


Figure 3-5. The gain-of-function *cdc-25.1*(rr31) results in an increase in the number of gut nuclei that is not significantly enhanced by the E(95) or E(50) genetic backgrounds (p=0.07 and p=0.05, t-test). The E(95); *cdc-25.1* and E(50); *cdc-25.1* strains were significantly different from each other (p=0.0003).

nuclei in surviving adults, to see how the animals cope with gut cells numbers the deviate from the norm. While ~20 gut nuclei are present in hatched larvae, during the L1 mothing stage 14 of the nuclei undergo bi-nucleation, bringing the total nuclei number in adult worms to 34 (Sulston, Horvitz 1977). Recent studies by (Lee et al. 2016), showed that *cdc-25.2* adult animals are viable with only 16 gut cells. The 16 gut cells that are present in the adult *cdc-25.2* mutants, are the results of inhibiting cell division in the last four divisions in E lineage development and the suppression of bi-nucleation during the L1 stage. Therefore, if HGS animals also fail to undergo bi-nucleation, we might expect to see fewer than 34 gut nuclei in the surviving adults. To access the number of intestinal nuclei in surviving HGS adults, we used a single-copy transgene reporter (irSi24), containing a nuclear localizing *pept-1*::mCherry fusion (Maduro et al. 2015). The results are as follows: control strains were as expected and displayed 33.9 ± 1.0 nuclei (Table 3-2; Fig. 3-2, and Fig. 3-4). The HGS strains showed variability in the amount of gut nuclei present and ranged from as few as 25 to as many as 59 nuclei (Fig. 3-4), and displayed a mean of 36.6 ± 6.3 in E(40) to 41.4 ± 7.9 in E(95). The data showed a correlation (R2 =0.93) between the number of gut nuclei and the severity of specification among the HGS strains. The increased cell division could be due to the E lineage partially adopting a different cell fate, where under certain circumstances the E lineage can adopt a C or MS like fate, which generates 47 and 80 cells, respectively (Sulston et al. 1983; Maduro et al. 2015).

By counting the amount of gut nuclei in both embryos and adults, we can calculate the likelihood of postembryonic bi-nucleation in most surviving HGS embryos.

Table 3-2 Number of gut nuclei among embryos making gut in strains affecting E lineage

	Strain[1]								
Stage	E(100)	E(95)	E(75)	E(50)	E(40)	E(30)	*cdc-25.1(rr31)*	E(95); *rr31*	E(50); *rr31*
end of embryogenesis[2]	20.0 ± 0.5 (34)	19.7 ± 6.7 (63)	12.9 ± 7.8 (85)	13.9 ± 8.0 (54)	11.9 ± 7.4 (54)	7.3 ± 5.3 (79)	35.5 ± 5.4 (23)	nd[3]	nd
L4[4]	33.9 ± 1.0 (29)	41.4 ± 7.9 (52)	38.9 ± 5.8 (50)	37.2 ± 6.0 (42)	36.6 ± 6.3 (51)	nd	56.5 ± 7.4 (51)	53.8 ± 7.5 (50)	59.5 ± 5.8 (31)

Numbers shown are mean ± SD with total number of embryos scored reported underneath.
[1]See Table 1 for full strain genotypes.
[2]Determined by scoring number of nuclei expressing *elt-2*::GFP reporter *wIs84*
[3]nd, not determined
[4]Determined by scoring number of nuclei expressing *pept-1*::mCherry reporter *irSi24*

By sorting the abundance of gut nuclei from highest to lowest for each HGS strains, and using the data that correspond to viable embryos in each strain, we were able to calculate the average number of gut nuclei in surviving embryos, 20.7 ± 4.1 for E(40) to 22.4 ± 3.6 for E(95). We then compared the data with the average number of gut nuclei at the L4 stage, reported above, to infer that ~16-19 cells undergo bi-nucleation in surviving HGS adults. This is higher the 14 cell bi-nucleation that occurs in wild-type animals, which suggest that in HGS embryos, some of the extra gut nuclei also can undergo bi-nucleation. There is also evidence for occurrence of such additional nuclear divisions in the *cdc-25.1*(rr31) strain, where on average of 21 more nuclei is present between late embryos and L4s (Table 3-2).

Our HGS strains also shed light on the minimum number of gut cell nuclei that are compatible with life, by looking at HGS animals that survived to the L4 stage with the least amount of gut nuclei (Fig. 3-4). For our strains the lowest number of gut nuclei in L4 animals are: 17 for E(95), 15 for E(75), 19 for E(50) and 18 for E(40). These numbers are comparable to the prior results generated by (Lee et al. 2016) suggesting that 16 embryonic gut nuclei are necessary for survival. However, it appears that the more severe HGS strains require slightly more gut nuclei to survive. Overall, these results confirm that among all HGS strains, the number of gut nuclei produced in embryos and adults correlates with the severity of the gut specification defect.

3.4.3. A marker the identifies E lineage descendants independent of the fate they adopt

Previous work done in our lab suggests that HGS strains experience hyperplasia within the E lineage, although some embryos contain significantly fewer than 20 gut nuclei, due to only a subset of E descendant cells adopting a gut fate (Maduro et al. 2015). In order to directly test this hypothesis, we attempted to trace the E lineage over time in HGS embryos carrying the wIs84 *elt-2*::GFP reporter, alongside a ubiquitously expressed nuclear-localized mCherry reporter (Murray et al. 2008) (Fig. 3-6). The control embryos showed onset of the *elt-2*::GFP reporter around the expected 8 cell stage. In addition, we also followed a small number of E(50) background embryos using 4D time lapse imaging and identified a single embryo in which only the posterior four of the 8E cells activated *elt-2*::GFP. This validates the hypothesis that the specification of the E lineage is not an "all-or-none" event and of all E descendants it is possible that it is possible for part of the E lineage to acquire *elt-2*::GFP expression. However, due to increased numbers of cells and the migration of the cells over time, we were unable to reliably follow all the descendants of E.

An alternative approach was taken to follow the descendant of the E cell over time. We used a reporter that was specific to only the E lineage, which allows for accurate marking of all E descendant cells regardless of the fates that the cells will later adopt. Using the *end-3*::mCherry marker, E descendant cells could be observed at any cell stage of embryogenesis, even when the location of the descendants and cell division patterns are drastically different from wild-type animals. The transgene reporter

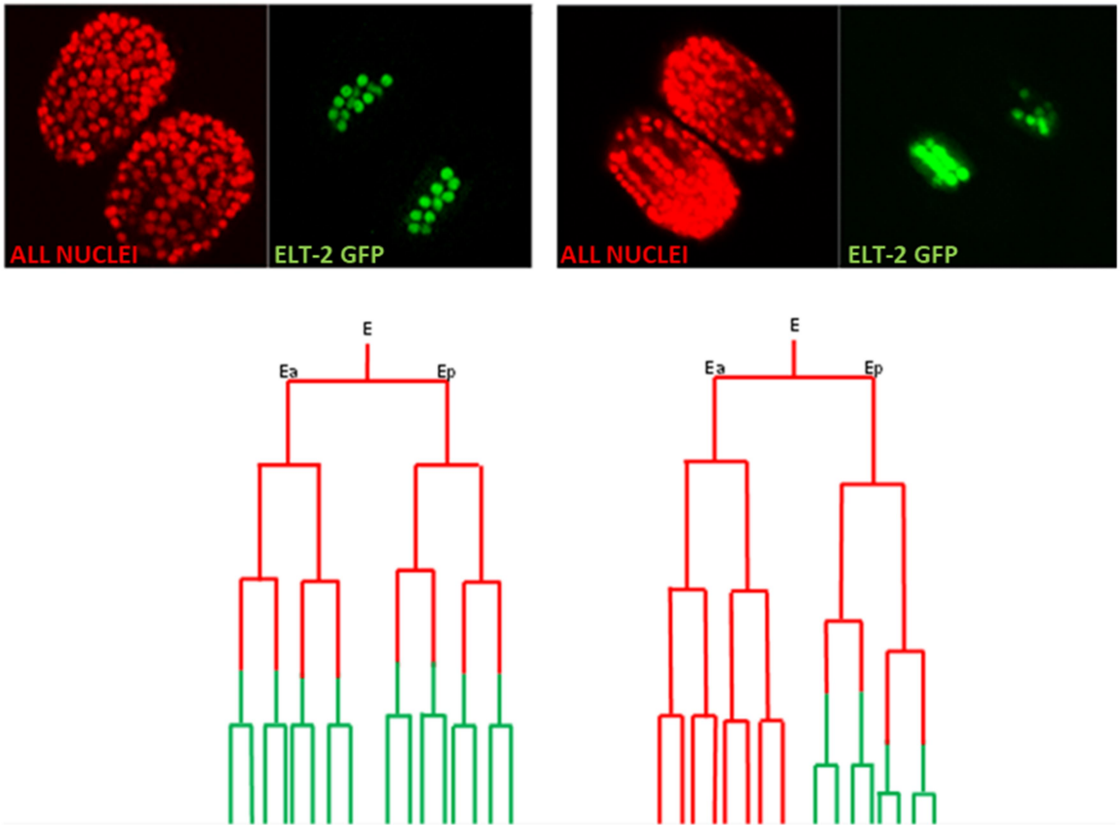


Figure 3-6. Following all of the descendants of the E cell, and determining whether these adopt an intestine fate or not, will allow us to identify patterns of mis-specification. Mutant embryos with a histone H2B::mCherry and *elt-2*::GFP markers were mounted at the 4-cell stage. 4D stacks were taken at 1µm intervals every 2 minutes until the embryo reaches the morphogenesis. The E lineage for each embryo was deciphered manually for each embryo until the 16 E cell stage.

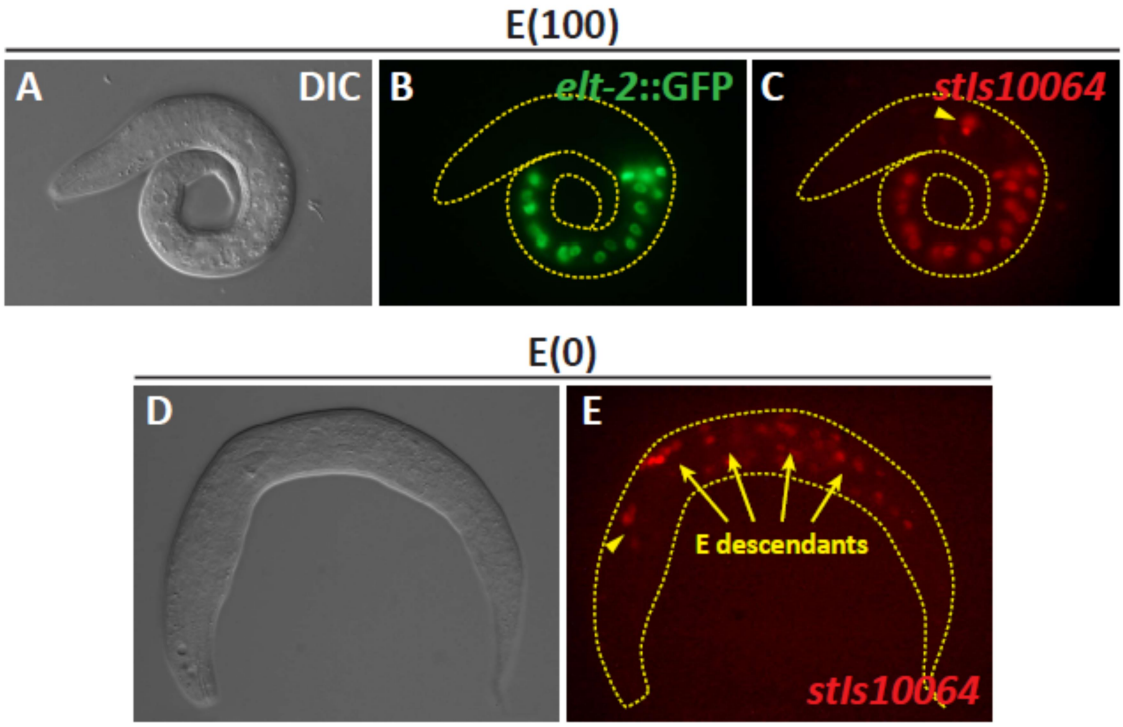


Figure 3-7. A reporter that marks the E lineage. The *end-3*::mCherry::H2B::*let-858*_3'UTR reporter stIs10064 (Murray et al., 2008) marks the descendants of E even when they do not make gut. (A-C) Control E(100) embryo. Panel (B) shows nuclei of intestinal cells. (C) The stIs10064 reporter is expressed in the same nuclei as *elt-2*::GFP, plus a small number of nuclei in the head (arrowhead) that are likely to be neurons based on their location. (D,E) Arrested larva of strain E(0), *end-1*(ok558) *end-3*(ok1448), with no intestinal cells (Owraghi et al., 2010). In such embryos, E adopts the fate of the C cell, and produces muscle and hypodermal cells (Maduro et al., 2005a; Owraghi et al., 2010). These ectopic C-like descendants are found in various positions throughout the larva (E). The ectopic expression in the head is also visible (arrowhead). Images are shown at the same scale. The larva in (D) is approximately 200 µm long.

for *end-3* should behave normally in the early E lineage in all HGS backgrounds, and since the reporter does not produce a functional END-3 protein, it should not interfere with specification. The *end-3* transgenes normally used in our lab only express from the 2E through 8E stages (Maduro et al. 2005), which is consistent with the transient expression of END-3 in wild-type animals. However, others have reported an *end-3*::mCherry transgene (stls10064) that maintained expression throughout embryogenesis (Murray et al. 2008). We confirmed the long lasting expression of the stIs10064 during multiple stages of the E lineage (Fig. 3-7 A,B,C). We confirmed that stIs10064 is capable of marking E descendants even when they do not adopt a gut fate. By crossing the stIs10064 reporter to an *end-1*(ok558) *end-3*(ok1448) double mutant background (Owraghi et al. 2010), essentially preventing gut specification, we still observed expression of stls10064 marker (Fig. 3-7 D,E). In addition to marking E descendant cells, we also saw additional expression of stIs10064 in non-intestinal cells. The expression was localized to 2-4 small cells near the nerve ring, suggesting that they are neuronal cells (Fig. 3-7 C, E).

3.4.4 Misspecification of gut leads to extra divisions and stochastic acquisition of gut fate within the E lineage

We crossed both *elt-2*::GFP transcriptional reporter and stIs10064 into the whole HGS allelic strain to simultaneously mark all E descendants and the cells that have committed to a gut fate. We chose E(100) as our control and E(75), E(50) background strains for further analysis. Image stacks were collected through confocal microscopy at

roughly 300-350 minutes after fertilization. The time that we chose represents the 8E-16E cell stage of gut development, during the time frame that *elt-2*::GFP should already be expressed in the gut cells and the gut primordium has not begun to extensively elongate (Sulston et al. 1983; Asan, Raiders, Priess 2016). The image stacks that we took of the HGS embryos capture both GFP and mCherry expression. The number of *elt-2*::GFP nuclei vs stla10064 nuclei were plotted to identify the proportion of E descendants that adopted a gut fate, which allowed us to further categorize the embryos based on the severity of E-fate cell transformation. The image stacks were also used to generate a heat map, to provide a reference to where E descendant cells were localizing to across the different categories (Table 3-3 and Fig. 3-8). As expected, the E(100) strain showed co-localization of *elt-2*::GFP and stIs10064 in the middle-posterior of the embryo, and 10–16 nuclei consistent with the time window of observation (Fig. 3-8A, panel a; Fig. 3-8B, panel a′; Fig. 3-9, bottom plot). Among both E(75) and E(50) backgrounds, we observed a wide range of embryos that showed as little as a single *elt-2*::GFP-expressing cell to as many as 24 (Fig. 3-8A, panels b–e; Fig. 3-9, middle and top plots). Evident in both HGS strains, the number of stls10064 marked EA descendant cells range from 6 to 41, and the number of *elt-2*::GFP nuclei ranged from 0 to 24. We saw that the more severe HGS strains had more E descendant cells on average. The E(100) embryos had an average of 12.4 ± 1.6 E descendants, with all expressed *elt-2*::GFP. The milder HGS strain E(75) showed an average of 16.7 ± 4.8 E descendants, while the more severe HGS E(50) strain showed an average of 18.4 ± 6.7. We notice a correlation between the fewer number of

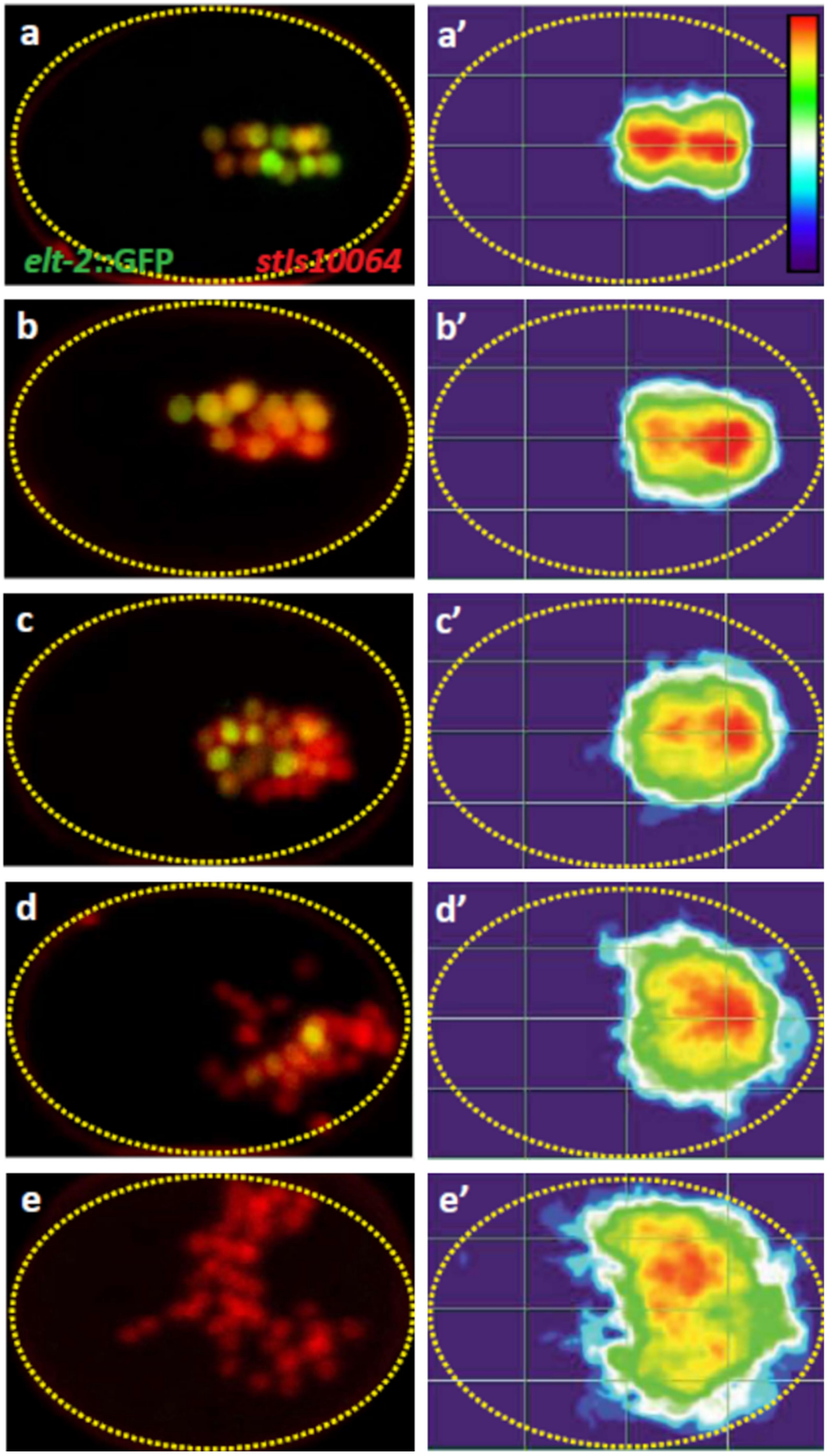

a
a'
elt-2::GFP
stIs10064
b
b'
c
c'
d
d'
e
e'

Figure 3-8. Behavior of E lineage descendants that adopt gut or non-gut fates in HGS strains at 300–350 min past fertilization. (A) Representative maximum-value projections of confocal micrographs showing embryos produced by HGS strains, arbitrarily divided into five classes. *elt-2*::GFP expression (gut fate) is green and stIs10064 expression (E lineage) is red. Panel a, wild type, in which all E descendants express *elt-2*::GFP. Panels b–e, progressively lower fractions of *elt-2*::GFP-expressing nuclei among E descendants, with no *elt-2* expression in panel e. (B) Heat maps showing location of E lineage descendants among multiple images with similar classes of percentage of *elt-2*::GFP-expressing nuclei, along a similar scale as the images in (A). The look-up table is shown in panel a′. Embryos are shown with anterior to the left. A *C. elegans* embryo is approximately 50 µm long.

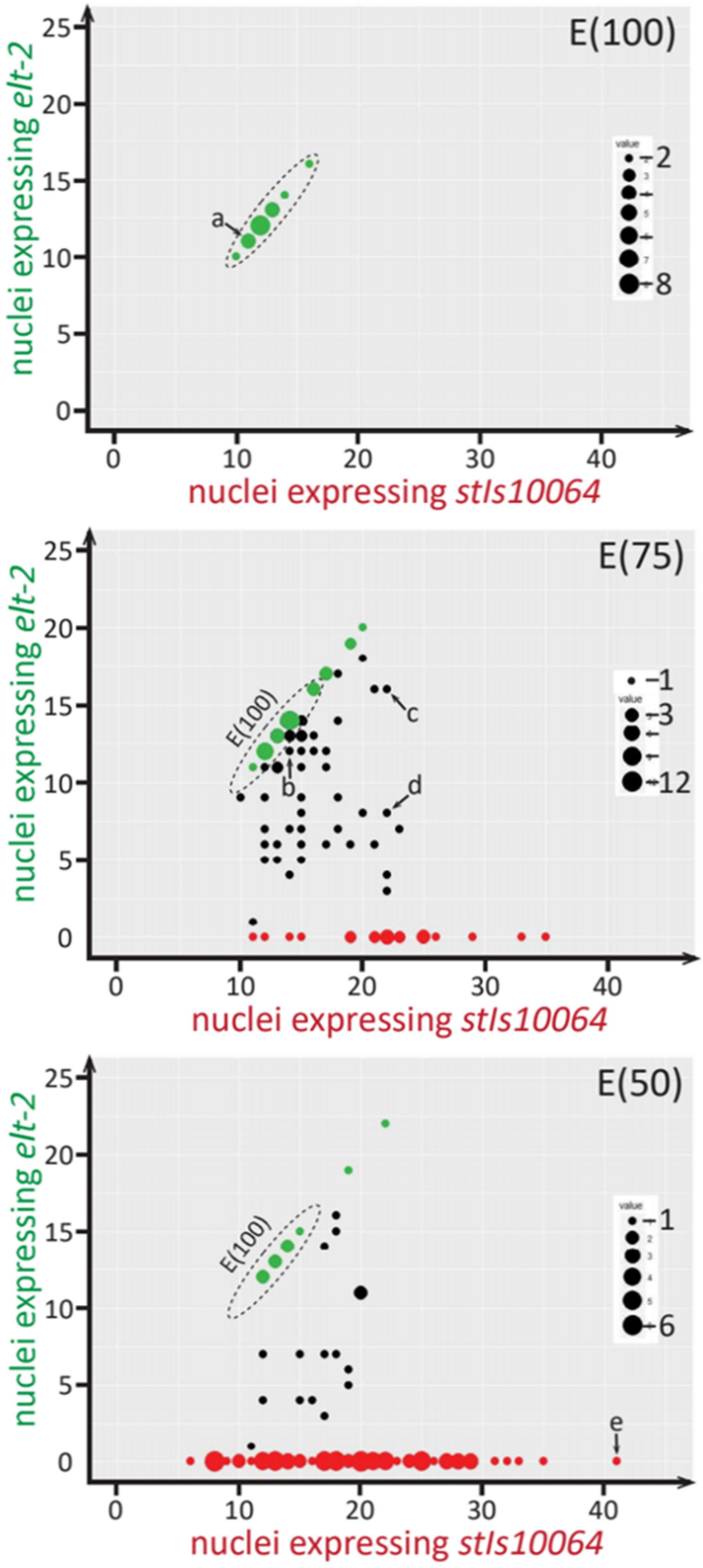

E(100)
nuclei expressing elt-2
nuclei expressing stIs10064
a
value
2
8
E(75)
E(100)
b
c
d
1
3
12
E(50)
e
1
6

Figure 3-9. Two-dimensional plots of number of *elt-2*::GFP-expressing nuclei vs. number of stIs10064-expressing nuclei (E lineage descendants). Green dots indicate embryos in which all E descendants expressed *elt-2::*GFP; red dots indicate embryos in which none of the E descendants expressed *elt-2*; and black dots indicate embryos where a portion of the E descendants expressed *elt-2*. The dot sizes represent numbers of observations with scale shown at right within each plot. Lowercase letters (a–e) indicate the data points corresponding to embryos that appear in the row of images in figure 3-8A. A dotted oval in the right two panels indicates the range of data obtained in the control.

Table 3-3 Summary of E lineage vs. *elt-2*::GFP expression data at 300-350 min past fertilization

	Control (n=22)	E(75) (n=100)	E(50) (n=100)
Average # of E descendants	12.4 ± 1.6	16.7 ± 4.8	18.4 ± 6.7
Average # of *elt-2*::GFP-expressing nuclei	12.4 ± 1.6	9.1 ± 6.0	2.6 ± 5.2
% of embryos in which all E descendants expressed *elt-2*::GFP	100%	33%	9%
% of embryos in which some (but not all) E descendants expressed *elt-2*::GFP	0%	46%	16%
% of embryos with no *elt-2*::GFP	0%	21%	75%

overall E descendants expressing the *elt-2*::GFP reporter and the greater likelihood of the E descendants to migrate away from the gut primordium, which in wild-type animals are located in the middle posterior of the embryo (Fig. 3-10, 3-11). The results validate the notion that partially compromised HGS strains do not follow an "all-or-none" mode of gut specification.

The following conclusion can be made about the HGS strains: First, in comparison to the control, the HGS strains make on average the same number of E descendant cells, if not more. Second, stochastic specification occurs at the level of individual E descendants, rather than the E lineage as a whole. This is evident in the 46% of E(75) embryos, and 16% of E(50) embryos that expressed the gut differentiation marker *elt-2*::GFP in a subset of E descendants. Third, there is correlation between the greater probabilities of a strain failing to specify gut with the increase in number of E descendants produced. The results display an inverse relationship between the greater numbers of E descendants produced with a lower proportion of E descendants adopting a gut fate. Lastly, as smaller subsets of E descendants adopt a gut fate, the farther away from the middle-posterior of the embryo these cells migrate. The relationship between the severity of gut specification defects and the effect on the E lineage suggests that E fate is adopted by E descendants through stochastic acquisition from a subset of E descendants in HGS embryos.

3.4.5 HGS strains shows delayed onset of elt-2, but normalized level at the end of development

In order for E descendant cells to undergo differentiation to become intestine cells, *elt-2* must be activated and the expression maintained through endodermal GATA factors and autoregulation from itself once its activated (Fukushige, Hawkins, McGhee 1998; Fukushige et al. 1999; Sommermann et al. 2010; Wiesenfahrt et al. 2015; Du, Tracy, Rifkin 2016). Since acquiring gut fate among E descendants occurs in a stochastic manner, this suggests that activation of *elt-2* is also stochastic and is stalled until a later time during gut development. Delays in an *elt-2* transcriptional reporter in an *end-3* mutant background have also been observed by others (Boeck et al. 2011). The more severe HGS strains show a smaller subset of E descendants adopting a gut fate. We hypothesize that *elt-2* expression also appears later in those strains. To test our hypothesis, we utilize the wIs84 reporter to evaluate the onset of *elt-2* expression. Upon analyzing embryos in several HGS background by tracing the onset of *elt-2*::GFP reporter in real time under fluorescence microscopy, we found that expression of *elt-2*::GFP was variable among the HGS strains. We saw a correlation between the severity of the specification defect across the strains and the delayed expression of *elt-2*(Fig. 3-10). E(100) embryos began to express the *elt-2* reporter gene on average of 2.9 ± 0.6 h after four-cell stage, while the HGS strains were delayed by 0.6 – 1.7 h with standard deviations ranging from 0.5 to 1.0 ($p < 10^{-4}$, Welch's T-test). We note that all embryos that activated *elt-2* maintain the expression throughout embryogenesis, consistent with the time frame that marks the commitment to a gut fate.

Because *elt-2* maintains its own expression upon initial activation, the final levels of *elt-2* should be similar across various HGS and control embryos that were able to active *elt-2* (Fukushige et al. 1999). We accessed the steady state level of *elt-2* expression across the HSG embryos using a different *elt-2*::GFP than before. The previous reporter derived from wIs84 was a transcriptional reporter, which cannot be used to examine the steady state level of the ELT-2 protein, due to the absence of most of the coding region (Fukushige, Hawkins, McGhee 1998). Instead, we opt to use gaIs290, a chromosomal insertion of a fosmid clone carrying the *elt-2* genomic region and surrounding genes with an in-frame GFP insertion into the coding region of ELT-2 (Sarov et al. 2006; Mann et al. 2016). The gaIs290 reporter was crossed into the HGS backgrounds and fluorescence images of 1.5-fold E (95) and E (40) embryos were captured and analyzed using ImageJ. We saw that there was no discernable difference in fluorescence between the E(100), E(95), and E(40) embryos (Fig. 3-11; $p > 0.2$ across all pairwise comparisons, Welch's t-test), and expression level was not skewed by the number of *elt-2* expression cells (Fig. 3-11). The elt-2 expression in E(95) also corroborated the finding of (Wiesenfahrt et al., 2015), who also detected relatively normal expression of elt-2 in a E(95) background.

The data strongly suggest that specification of gut fate is pushed back due to activation of elt-2 in individual E descendants being highly random, even though the delayed onset of *elt-2*::GFP does not affect the final levels of ELT-2 protein. This shows the necessity of timely activation of specification to orchestrate proper E cell cycle, even in mildly specification compromised strains, to prevent variation in numbers of E

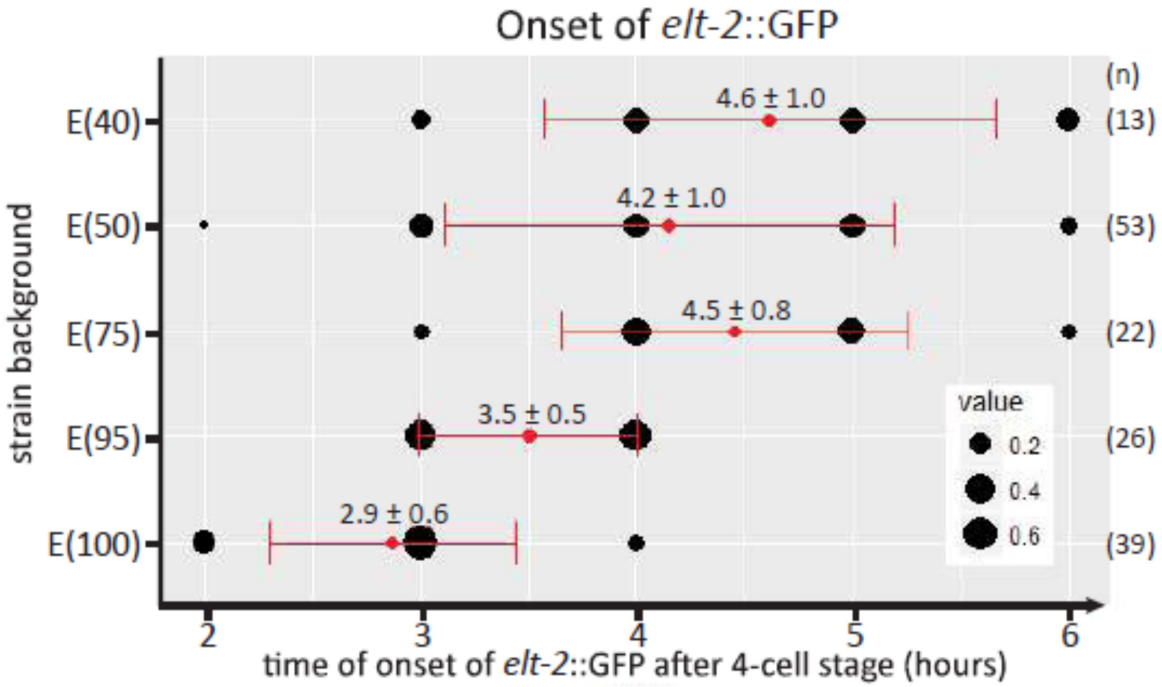


Figure 3-10. Time of onset of *elt-2::*GFP transcriptional reporter wIs84 in various backgrounds. Embryos were observed at 23–25 °C starting at the 4-cell stage (0 h) and onset of GFP expression determined by checking fluorescence every 60 min. Dot size corresponds to the proportion of embryos exhibiting onset at that time point after four-cell stage. Means ± SD are reported and indicated in red and the number of GFP(+) embryos scored is shown to the right of the graph. Embryos that did not activate the reporter were not included in the plot.

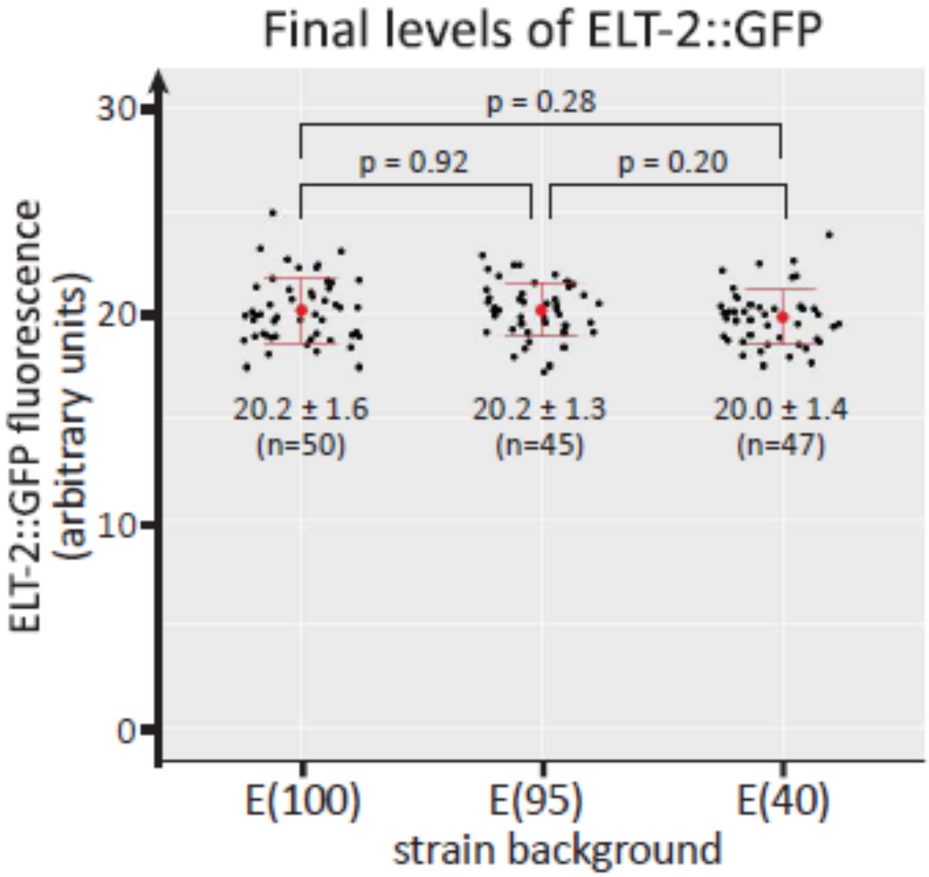


Figure 3-11. Expression levels of the gaIs290 translational ELT-2::GFP reporter. Fluorescence of individual nuclei was measured using ImageJ on digital images at 1.5-fold stage (identified by morphology) taken using identical settings across all embryos. Nuclei were chosen for analysis only if they were in focus (sharp outer edge and nucleolus).

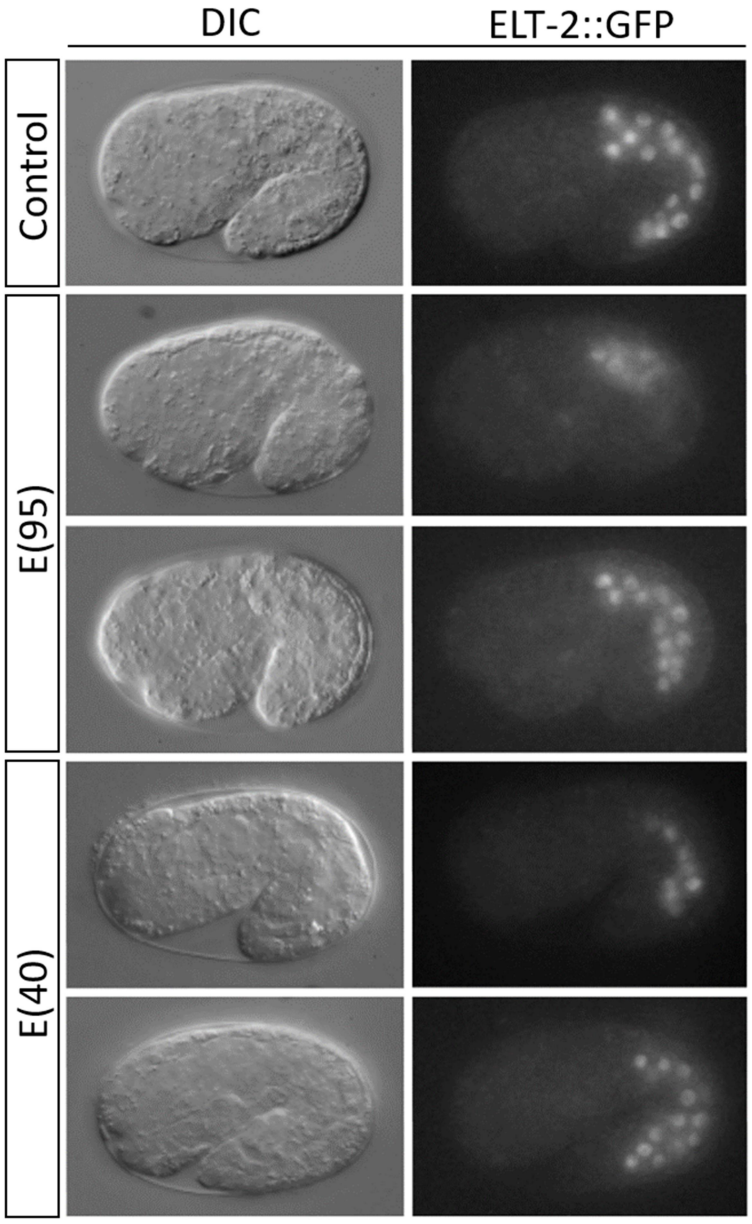


Figure 3-12. Representative single focal plane images of 1.5-fold stage embryos expressing gaIs290. In the E(95) and E(40) examples, embryos are shown that contained too few and an approximately normal number of nuclei. For these panels, fluorescence images were taken at identical settings across all embryos, made monochrome and adjusted for contrast using the same settings. Embryos are approximately 50 μm long.

descendants. Just as delayed *elt-2* expression in E descendants can push back specification of gut fate, the ectopic expression of *elt-2,* even in the absence of *end-1* and *end-3,* can generate a normal intestine (Wiesenfahrt et al. 2015). This corroborates the idea that the delay of ELT-2 is the major cause of the defects we observe in the HGS backgrounds. However, some HGS strains were reported to accumulate lipids during the adult stage (Maduro et al. 2015), suggesting that delayed gut specification has effects on function of the gut that cannot be compensated by ELT-2 in later stages.

3.5 Conclusion

These findings have advanced previous studies from our lab, by studying gut specification of individual E descendants rather than the E lineage as a whole in partially compromised animals (Maduro et al. 2005; Maduro et al. 2007; Maduro et al. 2015). Our results show the lack of robustness during endoderm specification affects cell division patterns, the ability of E descendants to adopt a gut fate, and gut morphogenesis, in a highly stochastic manner. The results demonstrate that the gut fate requires a robust and timely activation of both END-1 and END-3 for wild-type animals to acquire proper cell divisions and commitment to gut fate. When the activation of *end* genes is altered, the ability of E descendant cells to acquire a gut fate becomes stochastic and delayed. The observed results are similar to the phenotype derived from interference of Wnt/β-catenin asymmetry, which consequently also affects the direct activation of *end-1* and *end-3* (Robertson, Medina, Lin 2014) and a partial transformation of the E lineage observed in some *pop-1;end-1,3* embryos (Owraghi et al. 2010). The previous work done by others utilized mutant animals that are unable to completely morphogenesis, therefore it was not

possible to identify effects outside of the E lineage or to follow animals up to adulthood, as we have done here. In our HSG strains, the slight alteration of the specification network resulted in animals with more E descendants, which demonstrated that correct patterning is highly sensitive to partially compromised specification.

The results also shows a proportional response to delayed expression of *elt-2*::GFP transcription reporter and number of E descendant that acquired the gut fate. There is also an inverse relationship between a stronger specification defect and later adoption of gut fate, with a lower number of E descendant cells adopting a gut fate and cells positioning themselves further away from the gut primordium. The correlations that we observe are consistent with the probabilistic model of *elt-2* activation, derived from stochastic accumulation of the END transcription factors in HGS strains, and adds to the binary fate choice model that explains why ~ 20% of progeny embryo from a *skn-1* mutant mother are able to adopt a gut fate. In this model, gut specification can take place if the *end-1,3* transcripts reach a certain threshold. Our data shows that when the end genes are perturbed zygotically, the binary choice fate still applies to E descendants, but at the level of single cells. Future studies to model the effects of endoderm gene network perturbation would provide valuable insight.

Due to variable commitment to gut fate, the question arises: what do non- gut E descendant cells become? In HGS strains, non-gut E descendants migrate out of the gut primordium region, toward the exterior of the embryo, which indicates the cells have adopted adhesion and migratory behavior of other cell types. In the case where there is a

complete loss of *end-1* and *end-3*, a complete E fate transformation occurs and E will adopt the fate of the C cell, which differentiates to become muscle and hypodermis (Hunter, Kenyon 1996). Of our HGS strains we saw different degrees of E fate transformation. A proportion of embryos did not have an intestine and resembled an *end-1,3* double mutant, suggesting that when E descendant cells do not adopt an endoderm fate, they will acquire a C-like fate transformation. Non-E descendant cells were seen toward the posterior of the embryo (Fig. 3-8 A, B) which is indicative of a partial E to C fate transformation, since the cells now occupy a space where muscle and hypodermal descendants of C are typically found. Using a muscle specific reporter, *hlh-1*::GFP, we saw co-localization of the muscle specific marker and the E descendant marker as well as Hypodermis-lined cavity, validating the hypothesis that non-Gut E descendants undergoes partial C-like fate transformation (Fig. 3-13). It is important to note that, non-gut cells could contribute to the lethality observed in some HGS strains by interfering with morphogenesis leading to embryonic arrest.

Our work suggests that regulation of organ-level properties of the gut also occurs during the early stages of endoderm specification. Our results show that in surviving HGS adults there are steady-state levels of the differentiation factor ELT-2, which appear to be normal, regardless of the number of nuclei present. From this we can infer that ELT-2 does not influence the change in cell cycle, and the cycle cell change that results in the production of extra E descendants occurs before the expression of ELT-2. This is evident in the E(95) strain, where the slight delay of ELT-2 expression (~35 mins) resulted in a wide range of gut cell numbers among embryos. In addition to partial

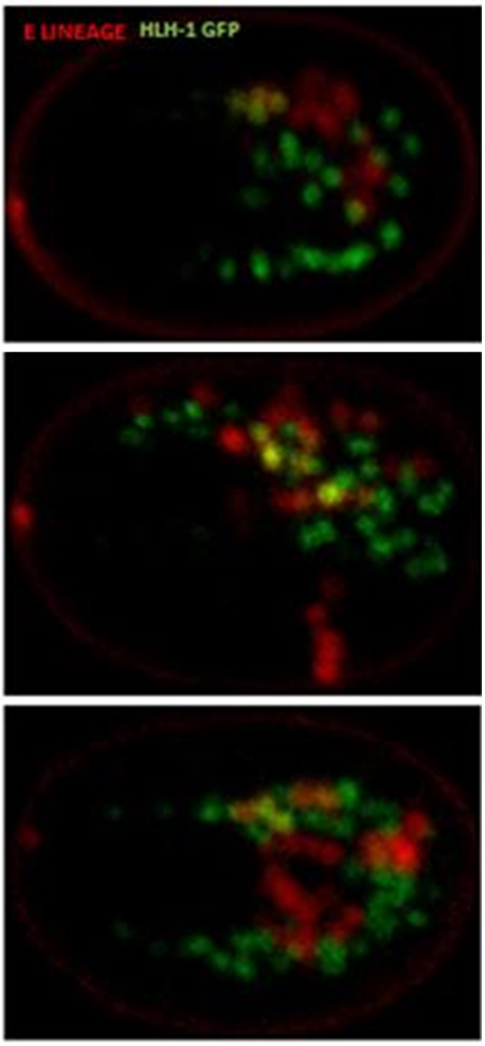


Figure 3-13. Non-E descendant cells were seen toward the posterior of the embryo which is indicative of a partial E to C fate transformation. Mutant embryos contain both the *end-3*::mCherry E cell lineage marker and *hlh-1*::GFP C fate reporter were imaged 300-350 minutes past fertilization Using a muscle specific marker, we saw co-localization of the muscle specific marker and the E descendant marker, which suggest that non-Gut E descendants undergo a partial C-like fate transformation.

transformation of E lineage, HGS strains are likely to have a change in gut function. Many pleiotropic effects are seen in E(75) and E(30) animals, such as delayed postembryonic development. This phenotype is consistent with that observed in a dietary restriction model (Palgunow, Klapper, Doring 2012; Maduro et al. 2015). Two possibilities arise from these observations: other targets within the endoderm network may establish normal patterning of division and metabolism, or delaying *elt-2* results in missing a sensitive developmental period during which normal *elt-2* target gene expression is necessary.

Our results support the idea for evolutionary bias of the specification gene network, favoring a highly robust mechanism during the initial steps of specification, to assure high fidelity of outcomes in early embryonic stages. Early gut specification relies on duplication of essential genes and parallel activation of key factors, but in terms of differentiation, ELT-2 is the main regulator of gut fate (McGhee et al. 2009; Sommermann et al. 2010; Maduro et al. 2015). Timely specification of the gut progenitor is crucial to generate enough gut cells for organ formation, as seen in our mutant strains, where stronger HGS strains generate fewer gut cells and have a reduced overall viability.

Our HGS strains display a range of gut nuclei at the young adult stage. The robustness of morphogenesis in the face of the supernumerary E cell lineage might reflect an ancestral mode of organogenesis, maintained from species with more variable cell divisions. The 20 cells seen in wild-type *C. elegans* may have been selected due to its sufficiency for gut morphogenesis and avoidance of the extra cost of energy dependent

cell divisions. For future studies, our HGS strains allow for the interrogation of organ level mechanisms of morphogenesis, which are robust to cell lineage variability.

3.6 References

Arribere, JA, RT Bell, BX Fu, KL Artiles, PS Hartman, AZ Fire. 2014. Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. Genetics 198:837-846.

Asan, A, SA Raiders, JR Priess. 2016. Morphogenesis of the C. elegans Intestine Involves Axon Guidance Genes. PLoS Genet 12:e1005950.

Bao, Z, JI Murray. 2011. Mounting Caenorhabditis elegans embryos for live imaging of embryogenesis. Cold Spring Harb Protoc 2011.

Blake, WJ, KA M, CR Cantor, JJ Collins. 2003. Noise in eukaryotic gene expression. Nature 422:633-637.

Boeck, ME, T Boyle, Z Bao, J Murray, B Mericle, R Waterston. 2011. Specific roles for the GATA transcription factors end-1 and end-3 during C. elegans E-lineage development. Dev Biol 358:345-355.

Bowerman, B, BA Eaton, JR Priess. 1992. skn-1, a maternally expressed gene required to specify the fate of ventral blastomeres in the early C. elegans embryo. Cell 68:1061-1075.

Colman-Lerner, A, A Gordon, E Serra, T Chin, O Resnekov, D Endy, CG Pesce, R Brent. 2005. Regulated cell-to-cell variation in a cell-fate decision system. Nature 437:699-706.

Davidson, EH. 2010. Emerging properties of animal gene regulatory networks. Nature 468:911-920.

Du, L, S Tracy, SA Rifkin. 2016. Mutagenesis of GATA motifs controlling the endoderm regulator elt-2 reveals distinct dominant and secondary cis-regulatory elements. Dev Biol 412:160-170.

Fukushige, T, MG Hawkins, JD McGhee. 1998. The GATA-factor elt-2 is essential for formation of the Caenorhabditis elegans intestine. Dev Biol 198:286-302.

Fukushige, T, MJ Hendzel, DP Bazett-Jones, JD McGhee. 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 96:11883-11888.

Hariharan, IK. 2015. Organ Size Control: Lessons from Drosophila. Dev Cell 34:255-265.

Holloway, DM, FJ Lopes, L da Fontoura Costa, BA Travencolo, N Golyandina, K Usevich, AV Spirov. 2011. Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in Drosophila segmentation. PLoS Comput Biol 7:e1001069.

Hunter, CP, C Kenyon. 1996. Spatial and temporal controls target pal-1 blastomere-specification activity to a single blastomere lineage in C. elegans embryos. Cell 87:217-226.

Irvine, KD, KF Harvey. 2015. Control of organ growth by patterning and hippo signaling in Drosophila. Cold Spring Harb Perspect Biol 7.

Lee, YU, M Son, J Kim, YH Shim, I Kawasaki. 2016. CDC-25.2, a C. elegans ortholog of cdc25, is essential for the progression of intestinal divisions. Cell Cycle 15:654-666.

Lin, R, S Thompson, JR Priess. 1995. pop-1 encodes an HMG box protein required for the specification of a mesoderm precursor in early C. elegans embryos. Cell 83:599-609.

Maduro, MF. 2009. Structure and evolution of the C. elegans embryonic endomesoderm network. Biochim Biophys Acta 1789:250-260.

Maduro, MF. 2017. Gut development in C. elegans. Semin Cell Dev Biol 66:3-11.

Maduro, MF, G Broitman-Maduro, H Choi, F Carranza, A Chia-Yi Wu, SA Rifkin. 2015. MED GATA factors promote robust development of the C. elegans endoderm. Dev Biol 404:66-79.

Maduro, MF, G Broitman-Maduro, I Mengarelli, JH Rothman. 2007. Maternal deployment of the embryonic SKN-1-->MED-1,2 cell specification pathway in C. elegans. Dev Biol 301:590-601.

Maduro, MF, RJ Hill, PJ Heid, ED Newman-Smith, J Zhu, JR Priess, JH Rothman. 2005. Genetic redundancy in endoderm specification within the genus Caenorhabditis. Dev Biol 284:509-522.

Mann, FG, EL Van Nostrand, AE Friedland, X Liu, SK Kim. 2016. Deactivation of the GATA Transcription Factor ELT-2 Is a Major Driver of Normal Aging in C. elegans. PLoS Genet 12:e1005956.

McGhee, JD, T Fukushige, MW Krause, 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.

Murray, JI, Z Bao, TJ Boyle, ME Boeck, BL Mericle, TJ Nicholas, Z Zhao, MJ Sandel, RH Waterston. 2008. Automated analysis of embryonic gene expression with cellular resolution in C. elegans. Nat Methods 5:703-709.

Owraghi, M, G Broitman-Maduro, T Luu, H Roberson, MF Maduro. 2010. Roles of the Wnt effector POP-1/TCF in the C. elegans endomesoderm specification gene network. Dev Biol 340:209-221.

Palgunow, D, M Klapper, F Doring. 2012. Dietary restriction during development enlarges intestinal and hypodermal lipid droplets in Caenorhabditis elegans. PLoS One 7:e46198.

Patel, SH, FD Camargo, D Yimlamai. 2017. Hippo Signaling in the Liver Regulates Organ Size, Cell Fate, and Carcinogenesis. Gastroenterology 152:533-545.

Praitis, V, E Casey, D Collar, J Austin. 2001. Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. Genetics 157:1217-1226.

Raj, A, SA Rifkin, E Andersen, A van Oudenaarden. 2010. Variability in gene expression underlies incomplete penetrance. Nature 463:913-918.

Robertson, SM, J Medina, R Lin. 2014. Uncoupling different characteristics of the C. elegans E lineage from differentiation of intestinal markers. PLoS One 9:e106309.

Sarov, M, S Schneider, A Pozniakovski, A Roguev, S Ernst, Y Zhang, AA Hyman, AF Stewart. 2006. A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. Nat Methods 3:839-844.

Sommermann, EM, KR Strohmaier, MF Maduro, JH Rothman. 2010. Endoderm development in Caenorhabditis elegans: the synergistic action of ELT-2 and -7 mediates the specification-->differentiation transition. Dev Biol 347:154-166.

Sulston, JE, HR Horvitz. 1977. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev Biol 56:110-156.

Sulston, JE, E Schierenberg, JG White, JN Thomson. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol 100:64-119.

Wiesenfahrt, T, JY Berg, EO Nishimura, AG Robinson, B Goszczynski, JD Lieb, JD McGhee. 2015. The Function and Regulation of the GATA Factor ELT-2 in the C. elegans Endoderm. Development.

Zhu, J, T Fukushige, JD McGhee, JH Rothman. 1998. Reprogramming of early embryonic blastomeres into endodermal progenitors by a Caenorhabditis elegans GATA factor. Genes Dev 12:3809-3814.

CHAPTER 4

Adult Survivors of Hypomorphic Specification have Defects in Stress Response and Metabolism

4.1 Abstract

In *Caenorhabditis elegans*, it has been shown that a threshold amount of *end-1* transcript is necessary for the specification of endoderm and to ultimately make an intestine (Raj et al. 2010). Our lab has an allelic series of mutants in which endoderm specification has been compromised through mutations of one or more MED binding sites in the promoter region of the endoderm specification genes, *end-1,3*(MED-), we also generated additional partial specification strains through the use of *end-3* single mutants and *med-1; end-3* double mutants (Maduro et al. 2015; Choi, Broitman-Maduro, Maduro 2017). Perturbation of an upstream factor in the endoderm gene regulatory network results in stochastic activation of the *end-1* and *end-3* genes. We observed that a subset of animals is capable of generating the necessary amount of *end-1,3* transcripts, undergo specification, and hatch out as larvae. We find that fully differentiated intestine retains a memory of their "near-miss" specification, which results in abnormal adult phenotypes. Of the surviving adult animals, a proportion of them express phenotypic traits such as sterility, defects in elongation and defects in movement as they mature into adults. These animals also contain excess lipids, which suggest that they are developing as if

undergoing food deprivation, even though they are feed *ad libitum*. Previous data supports the idea that our *end-1,3*(MED-) animals develop defects in adulthood as a result of delayed *elt-2* activation and late commitment to terminal gut fate (Choi, Broitman-Maduro, Maduro 2017). This also suggests that the adults that survive "near-miss" specification display later phenotypes that can be attributed to deficits in embryonic gene activation. We have decided to focus this study on the *med-1; end-3* double mutant strain based on the strength of defect and the reproducibility of other phenotypes in other assays. We have analyzed transcriptome data from gut cells collected from young adults of control and hypomorphic gut specification (HGS) defective animals to better understand how changes in gene expression during the early developmental process and delayed activation of important specifying factors can result in dysfunctional organ or tissue types, even though they superficially appear normal. Our data suggest that the HGS mutant worms have a heightened immune response due to the upregulation of innate immunity genes, and the increased expression of membrane raft genes correlates with the animal's excess lipid phenotype.

4.2 Introduction

In *Caenorhabditis elegans*, formation of the intestine is dependent on the activation of the endoderm specification genes *end-1* and *end-3* (Maduro et al. 2005; Owraghi et al. 2010), and their downstream target *elt-2*. Once activated, *elt-2* maintains its own expression through self-regulation and acts as the regulator of all intestine genes (Fukushige et al. 1999; McGhee 2007). The expression of *elt-2* is highly robust. The

activation of this gene is at the bottom of the endoderm gene regulatory cascade, which is made up of multiple input and redundant wiring. Animal development requires timely activation of essential genes. If there is an incomplete penetrance of gut specification that leads to delayed expression of *elt-2*, will it influence the function and integrity of the intestine?

To investigate this question, we needed a model of incomplete penetrance in gut specification. While the knockdown of *skn-1* and/or *med-1,2* eliminates gut specification some of the time, the removal of these gene products also affects other lineages, leading to embryonic arrest (Bowerman, Eaton, Priess 1992; Rocheleau et al. 1997; Thorpe et al. 1997; Rocheleau et al. 1999; Maduro et al. 2001). Our lab generated a series of allelic mutants that are able to specify gut at a range of 0%-100%. Mutants from the allelic series contain *end-1* and *end-3* genes with mutated MED-1,2 binding sites. The mutated GTATACTYYY binding site resides in the promoter of the *end* genes (Broitman-Maduro, Maduro, Rothman 2005; Lowry et al. 2009), which removes the inputs of MED-1,2 from the endoderm network without affecting other lineages. The *end-1* and *end-3* genes are functional and can be activated by other upstream targets such as *skn-1* and *pop-1*. We also generated additional partial specification strains through the use of *end-3* single mutants and *med-1; end-3* double mutants. The *med-1; end-3* double mutant will be referred to as MS404. These strains' ability to specify gut is dependent on the stochastic acquisitions of enough *end* transcripts to drive endoderm specification (Raj et al. 2010), resulting in a greatly reduced probability of gut specification that is consistent with late activation of *elt-2* and a stochastic commitment to endoderm fate.

This set of strains has been used to evaluate the effect of "near-miss" specification on organ differentiation. The current model for endoderm specification centers around the idea that a threshold level of *end-1* and *end-3* transcript are needed for the activation of *elt-2*, the essential gene for endoderm differentiation. After the initial activation of *elt-2*, it will auto regulate, and its protein product will be the central regulator for all intestinal genes. After activation of *elt-2* the cell will then "lock-down" its commitment to endoderm and its daughter cells will commit to the same fate (McGhee 2007; McGhee et al. 2009). However, it is not known if the integrity of the gut from animals that contained barely enough end transcripts to activate *elt-2* is normal. Previous data from the lab has shown that while the appearance of gut from *end-1,3*(MED-) strains is superficially normal, roughly 5-45% of adults showed various pleiotropic phenotypes. Defects in movement (Unc), body elongation (Dpy), sterility (Ste) and egg laying (Egl) were seen, which suggests functional defects in the gut.

We have decided to focus this study on the MS404 strain based on the strength of defect and the reproducibility of other phenotypes in other assays. These animals display a delayed *elt-2* activation and late commitment to terminal gut fate (Choi, Broitman-Maduro, Maduro 2017), which could result in delayed or missed activation of downstream target genes that are essential for a properly functioning intestine. We hypothesize that intestinal integrity of the surviving adults from the MS404 strain is compromised, which leads to starvation and pleiotropic phenotypes such as sterility (Seidel, Kimble 2011). It is possible that the adult phenotypes in surviving MS404 mutants are the results of differential gene expression, due to delayed *elt-2* expression

across different animals. We performed a tissue specific RNA-seq from isolated intestine tissues to get a reliable and sensitive read out of genes that are mis-regulated in surviving adults from the partially compromised endoderm specification strains.

4.3 Material and Methods

4.3.1 Worm Maintenance and Strains Used

All strains were grown on *E. coli* OP50 and maintained at 20–22°C. *C. elegans* animals were grown on *E. coli* OP50 and handled according to standard methods. The reference strain was Bristol N2, mutant strain was MS404: *med-1*(ok804); *end-3* (ok1448). Animals were synchronized by bleaching with a mild bleach solution (Stiernagle 2006) and left to develop at 20 °C until they reached the young adult stage. Animals were grown in the presence of excess food that was never depleted throughout the experiment. The GFP reporter constructs were injected into MS2382: *med-1* (ok804); *end-3*(ok1448), irEX710 [*end-3(+), unc-119*::mCherry].

4.3.2 Primers Used

CEL-seq libraries

CEL-seq 1	5’CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNAGACTCTTTTTTTTTTTTTTTTTTTTV-3’
CEL-seq 2	5’CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNAGCTAGTTTTTTTTTTTTTTTTTTTTV-3’
CEL-seq 3	5’CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNAGCTCATTTTTTTTTTTTTTTTTTTTV-3’

CEL-seq 4	5’CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNAGCTTCTTTTTTTTTTTTTTTTTTTTTTTTTV-3’
CEL-seq 5	5’CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNCATGAGTTTTTTTTTTTTTTTTTTTTTTTTTTV-3’
randomhexRT	5’GCCTTGGCACCCGAGAATTCCANNNNNN-3’
RP1	5’AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA-3’
RPI (2)	5’CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA-3’
RPI (4)	5’CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA-3’
RPI (5)	5’CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA-3’
RPI (6)	5’CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA-3’
RPI (7)	5′CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA-3’

qPCR validation

elt-2 Forward	5′-ctgcgactgccacatatgtgt -3’
elt-2 Reverse	5′-atcggcaggtcttaggccagt-3’
F57F4.4 Forward	5′-caacgtcaaggaagctgccgtaaacaccac-3’
F57F4.4 Reverse	5′-gttatgcgcgaaaagagcacatatgtacat-3’
gfi-1 Forward	5′- agagaatttttccatcatttctttcatgta-3’
gfi-1 Reverse	5′- tgaaagaacgctcaacaattaagatttagt-3’
Actin Forward	5′-aagtcctacgaacttcctgac-3’
Actin Reverse	5′- gagatccacatctgttggaag-3’

GFP reporter constructs

metr-1 forward	5′-ttggaaatgaaataagcttgcatgcctgcatgcaacaactaccgtattccttgcgcattc-3’
metr-1 reverse	5′-cataccttgggtcctttggccaatcccgggacataaattgtgggaaaatttatgattaa-3’
msra-1 hind	5′-gataaagcttctgaagtcatagtaaattat-3’
msra-1 Bamh	5′-ggcaggatccaaataagccatgtttcacct-3’
C17H12.8 forward	5′-ttggaaatgaaataagcttgcatgcctgcagaactaatgttcaatttgaaactttggtat-3’
C17H12.8 reverse	5′-ataccttgggtcctttggccaatcccgggtgatgacaaagtggaaagtaaattatcaga-3’
dod-19 forward	5′-ttggaaatgaaataagcttgcatgcctgcatgagtgagacttcagaatgatataaatga-3’
dod-19 reverse	5′-ataccttgggtcctttggccaatcccgggttcttagaatggtaaagtttcaaaagtga-3’
gpdh-1 hind	5′-aacaaagcttgtaagcgttttcaaatcatt-3’
gpdh-1 bamh	5′-ggaaggatccaagttctggcacggcataat-3’
ckb-2 hind	5′-tgtgaagcttatgtggatgctgataataag-3’
ckb-2 Bamh	5′-tcgcggatccttcacttgacttttatccac-3’
spp-8 pstI	5′-gtccctgcaggccaaaacaaattcagtcag-3’
spp-8 bamh	5′-gagatggatccgattgtgctccaagagcca-3’

4.3.3 Isolation of Nematode Intestine

Young adult animals were selected from NGM plates seeded with *Escherichia coli* (OP50). Animals were picked and placed in a watch glass with 1 ml of ice cold 1x PBS. Animals were washed by gently pipetting the PBS, allowing the animals to settle to the bottom of the watch glass, removing the bacteria filled PBS, and adding back fresh PBS. Animals were washed a minimum of 3 times, to insure minimal amount of bacterial transfer in our dissected tissues. Animals were then carefully decapitated using a 25G5/8 needle. Intestines were gently extracted and cut from the carcass which was subsequently disposed. Individually dissected intestinal tissue was transferred with a

glass micropipette onto the cap of a 0.5 ml LoBind tube (Eppendorf). Excess liquid was aspirated off, and the tube was frozen in liquid nitrogen. One intestine was collected per tube.

4.3.4 RNA Extractions for library prep

Using the MessageAmp II aRNA Amplification Kit (AM1751) from Ambion, modified RT primers were used to isolate and amplify RNA from 10 frozen gut samples, according to the CEL-Seq (Single cell RNA-Seq by multiplexed linear amplification) protocol published by (Hashimshony et al. 2012). After successfully generating amplified RNA (aRNA) through linear in-vitro transcription, 5 ng of aRNA was used to generate CDNA libraries following the CEL-Seq2 protocol (Hashimshony et al. 2016) with the following modifications:

To increase the amount of starting material and remove the biological bias observed between the animals in our mutant strain, we pooled 10 dissected intestines per RNA library. Individually frozen intestines were pooled by adding 2 µl of nuclease free H_2O on top of one frozen gut, located on the cap. The cap was heated to 70 °C by placing it on top of a glass slide that was resting on the surface of a PCR machine set to appropriate temperature. The samples incubated for ~ 1 minute, during this time the sample was constantly mixed via pipetting After the 2 minute incubation, the droplet was transferred to the next frozen intestine sample, and the process was repeated until all 10 samples were pooled. To compensate for H_2O evaporation during the process, 1 µl of H_2O was added as needed. After the final intestine was pooled, we briefly spun down the

sample and adjusted the volume to 1 µl by SpeedVac or the addition of H_2O. The 1 µl sample was transferred to a PCR tube, which we used to carry out the remainder of the protocol. We used the primers from (Hashimshony et al. 2012), which are modified RT primers that include an internal barcode and the sequence for a T7 promoter. To the 1 µl sample, 0.2 µl of RT primer was added and incubated at 70 °C for 5 minutes, followed by a quick chill on ice. For the first strand cDNA synthesis, the following reagents were added to the sample and primer mixture: 0.2µl of First Strand buffer, 0.4µl of dNTP, 0.13 µl of RNase Inhibitor, and 0.13 µl of ArrayScript, and incubated for 2 hours at 42°C. The second strand synthesis was done by adding the following reagents: 6.3µl of DDW, 1 µl of Second strand buffer, 0.4 µl of dNTP, 0.2 µl of DNA Pol, 0.13 µl of RNaseH, and incubating the sample at 16 °C. The cDNA sample was cleaned through column purification, and the eluted cDNA was adjusted to 6.4 µl using a SpeedVac. In order to generate aRNA to make our sequencing libraries, we ran a 13 hour in-vitro transcription reaction at 37°C, using the 6.4µl of cDNA contain the T7 promoter and 1.6µl of the following reagents: T7 enzyme, 10x T7 buffer, ATP, CTP, GTP, UTP. Following the cleanup, the amount and quality of the aRNA was checked using the Agilent 2100 Bioanalyzer RNA pico chip, before proceeding to generate cDNA libraries.

4.3.5 Preparation of Small RNA Libraries

For the RT reaction step, we incubated RNA and randomhexRT primer for 2 minutes at 70 °Cs, followed by a quick chill on ice, and then incubated the sample at 5 minutes at 25 °Cs. The reaction was carried out using 1µl First Strand buffer, 2µl dNTP,

0.5µl RNase Inhibitor, and 0.5µl ArrayScript, and incubating for 2 hour at 42 °C in a thermal cycler with heated lid. For the PCR amplification, the following reagents were added to the reverse transcription reaction: 11µl of Ultra Pure water, 25µl of 2X Phusion Master Mix, 2µl of Illumina Adapter Sequences "RPI", and one of the following (RPI4, RPI5, RPI6, and RPI7) to multiplex our libraries. The PCR was amplified in the thermal cycler with these conditions: 30 seconds at 98°C, with 12 cycles of 10 seconds at 98°C, 30 seconds at 60°C, 30 seconds at 72°C, and 10 minutes at 72°C. The PCR product was cleaned up using AMPure XP Beads. Eluted sample was checked for amount and quality using the Agilent 2100 Bioanalyzer DNA chip. At the time when the RNA libraries were generated, CEL-Seq2 had not been published. A version of the protocol was acquired through personal communication with Itai Yanai and Tamar Hashimshony.

4.3.6 Transcriptomic Sequencing and Analysis

Libraries were sequenced on the Illumina MiSeq V3, with the 26 x36x 6 cycle, a loading molarity of 8.76, and the addition of ~20% phiX spike-in, at University of California, Riverside's Genomic Core. Paired-end sequencing was performed, reading 26 bases for read 1, 36 bases for read 2, and the Illumina barcode when needed. Due to the nature of the RT primers that we used to generate aRNA, read 1 provided us with little information, since we did not use the internal barcode for multiplexing. Downstream analysis was performed with only read 2. We received 4-5 million reads per library; raw reads were mapped to the *C. elegans* (ce10) genome assemblies using tophat (version v2.1.0) and HTseq-count (0.6.1p2) for quantification. In the R software (version 3.2.1)

we normalized the counts using EDASeq (Risso et al. 2011) and ran the differential gene expression analysis between N2 and MS404 using EdgeR (Robinson, McCarthy, Smyth 2010). A secondary differential gene expression analysis was performed with DEseq 2 (Love, Huber, Anders 2014) to confirm our findings.

4.3.7 Whole Animal RNA Extraction, DNase treatment.

Gravid worms were washed off a large plate with M9, pipette into a 1.5mL Eppendorf tube, and centrifuge at low speed, and supernatant removed. Each sample was washed at least four times with 1 mL M9, to remove any remaining bacteria. Worms were resuspended in 1mL Trizol and homogenized. For every 1mL of sample, we added 200µl Chloroform, vortexed, and centrifuged for 3 minutes at 14,000rpm. We then transferred aqueous phase (isolated RNA) to a new tube, and cleaned isolated RNA using RNeasy kit, following manufactures protocol. The A260 value was checked using the NanoDrop to determine concentration. After completion of the RNA cleanup step, the sample was treated with TURBO DNase, following manufacture's protocol.

4.3.8 qPCR Validation

The cDNA used was synthesized from RNA isolated from whole animals using conventional methods. Primers used were specifically designed to flank two exons, to discourage amplification of non-specific sites. The reactions were carried out using 2X SYBER green in a final volume of 25 µl, using the Bio_RAD CFX96 in the following condition: 1) 95C for 3 minutes, 2) 95C for 10 sec, 3) 58C for 10 sec, 4) 72 for 30 sec, 5) repeat steps 2-4 39 times, 6)95C for 10 sec, 7) melt curve 65C to 95C.

4.3.9 GFP Reporter Constructs, Microinjection

Using standard methods, 1000bp promoter region sequence of genes were individually cloned into either pGB525 vector via Gibson assembly (Gibson et al. 2009), or into pPD95.67 vector using traditional cloning (Sambrook, Russell 2001) with the primer sets listed above. Plasmids were precipitated and introduced into MS2382 animals alongside the plasmids pMM809 (*unc-119::CFP*) and pRF4(*rol-6*) via microinjections. Animals that expressed the *rol-6* phenotype, mCherry selection marker, *unc-119*::CFP, and *elt-2*::GFP were selected for analysis.

4.3.10 Microscopy and Imaging

Conventional epifluorescence and differential interference microscopy were performed on an Olympus BX-51 microscope, imaged through an LMscope adapter (Micro Tech Lab, Graz, Austria) and Canon Rebel T1i Digital Camera using software supplied with the camera. Images were analyzed using ImageJ, and plots were generated in ggplot2 in an R environment (http://ggplot2.org).

4.4 Results

4.4.1 Annotation and Verification of Intestine Specific Sequencing Libraries

To investigate whether the adult defects observed in our MS404 strain are due to mis-regulation of endoderm genes during embryogenesis, we compared the gut transcriptome between our endoderm specification mutant MS404 and our control strain N2. The anterior portion of the intestine was dissected from N2 and MS404 young adult

animals and used to generated RNA-sequencing libraries. Three biological samples were generated for each strain, and libraries were sequenced on the Illumina MiSEQ platform, which generated roughly 4 million plus reads per sample. Raw reads where annotated using the *C. elegans* (ce10) genome assemblies and libraries were normalized using EDA-seq in R program (Risso et al. 2011). Biological replicates were plotted against each other and Pearson Correlation values were calculated, to check for reproducibility (Figure 4-1). The Pearson Correlation values among the biological replicates were all above R=0.93, which suggests that there are no major deviations between biological replicates. Our sequencing libraries detected the expression of 5724 genes. When comparing our data set with other previously published gut transcriptomic data sets, 5458 genes matched at least one of the of the following data sets from (Pauli et al. 2006; Spencer et al. 2011; Haenni et al. 2012; Lightfoot et al. 2016) (Figure 4-2). These published data sets were generated by RNA sequencing or microarray analysis. The intestine cells used to generate libraries were collected via fluorescence-activated nuclei sorting, tissue-specific promoters to mark cells for isolation by FACS, dissecting and pool 250 intestines, and mRNA tagging respectively. Our transcriptomic data differs in the amount of starting material used, where only 10 animals were pooled per library, in order to minimize the amount of biological variation within each set. Genes detected from our sequencing libraries highly correlated (~ 95%) with previously published *C. elegans* gut transcriptomic data sets, we are confident that our dataset is a true representation of the gut transcriptome in *C. elegans*.

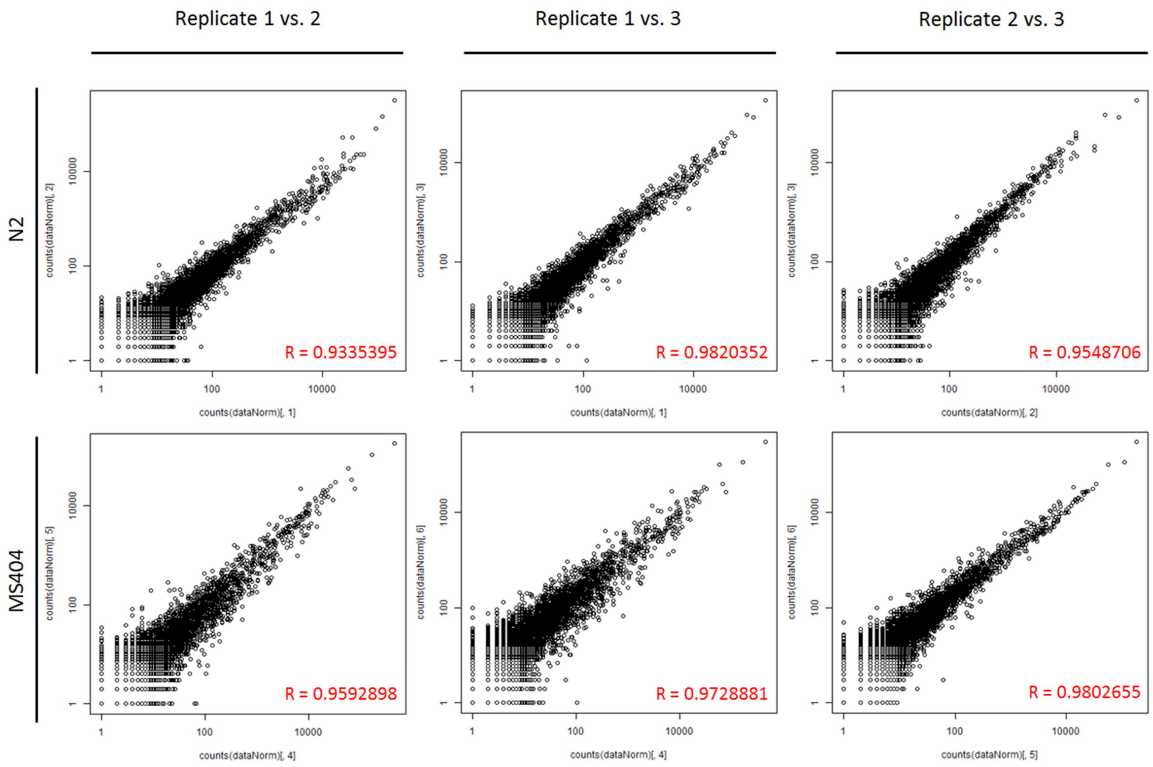


Figure 4-1 There are no major deviations between biological replicates. Pearson correlation values were calculated among the biological replicates to check for reproducibility of each sample in the R environment. All Pearson values were greater than 0.934.

4.4.2 Identification of Differential Gene Expression and Gene enrichment

After verifying the genes detected in our sequencing libraries are indeed gut specific, we used the edgeR program in R to compare the sequence reads from N2 and MS404 and identify differentially expressed genes using a False Discovery Rate (FDR < .1) as our cut off (Robinson, McCarthy, Smyth 2010). The analysis identified 63 differentially expressed genes, of which a large proportion of the genes displayed greater expression in the MS404 strain (Fig. 4-3, 4-4 Table 4-1). A secondary differential gene expression analysis was done using the DEseq2 package in R using an adj. p-value < .05 as our cut off (Appendix B), the results highly correlate with the edgeR findings.

Next, we wanted to see whether there was a correlation between the adult gut defects seen in MS404 and the levels of *elt-2* transcripts. *Elt-2* is the key endoderm differentiation gene, and is essential for activation of genes that have a wide range of functions, such as metabolism, intestine structure, and maintenance of the gut. Due to the delayed onset of *elt-2*:: GFP expression observed during embryogenesis seen in various gut defect mutants (Choi, Broitman-Maduro, Maduro 2017), we hypothesized that MS404 should contain a lower level of *elt-2* in comparison to wildtype animals. Surprisingly, *elt-2* was not found to be a differentially expressed gene. This shows that the defects seen in the adult animals are not due to improper activation of the key gut differentiation gene *elt-2* or the inability of *elt-2* to activate its downstream targets due to low transcript number. The results suggest that the alternative transcriptional state seen

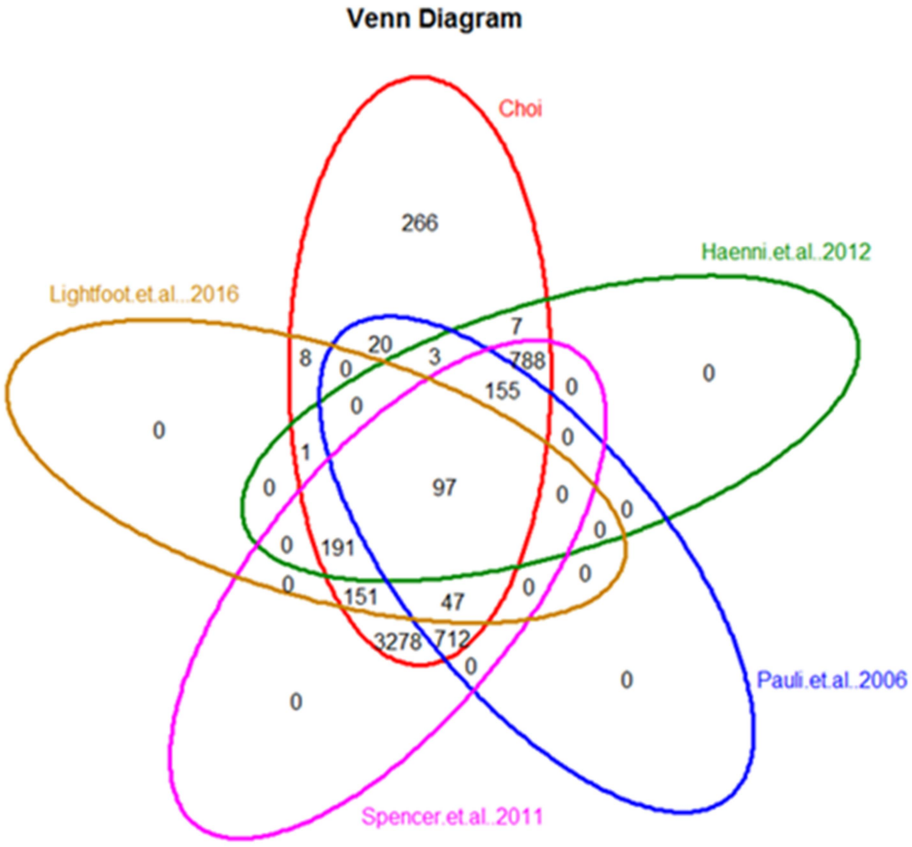


Figure 4-2 Venn diagram comparing five different intestine specific transcriptomic data sets. The number of genes expressed in our sequencing libraries (5,724) was compared to 4 previously published C. elegans gut transcriptome, with each transcriptome being generated by a different method. Our transcriptomic data setting had ~ 95% overlap with at least one other published data set, and only 266 were independently identified from our libraries. A 5-set Venn diagram plot was generated in R using ggplot 2.

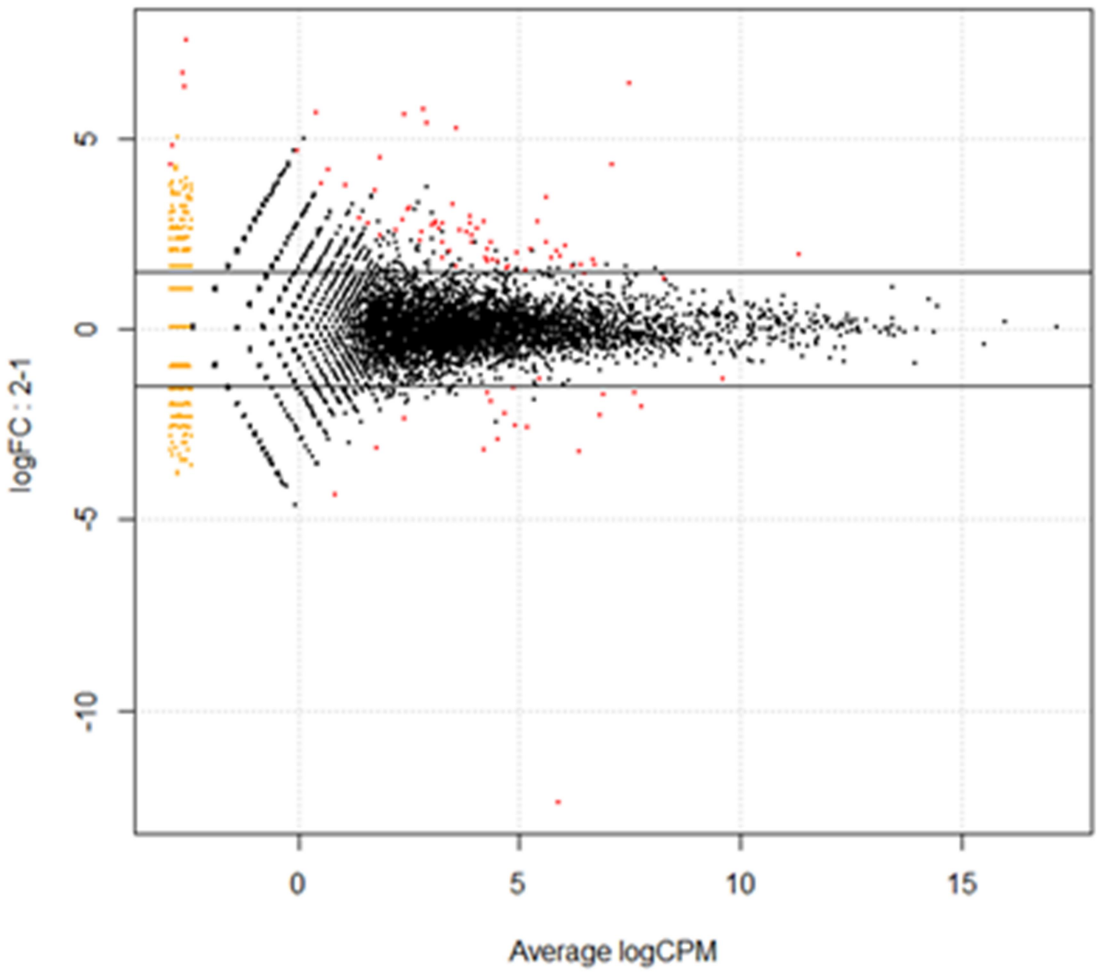


Figure 4-3 Smear plot generated from edgeR in the R environment shows the fold change (FC) against the average log concentration for each gene. The most differentially expressed genes are represented by red dots. The orange dots on the left side signify genes that were observed in only two of replicate samples.

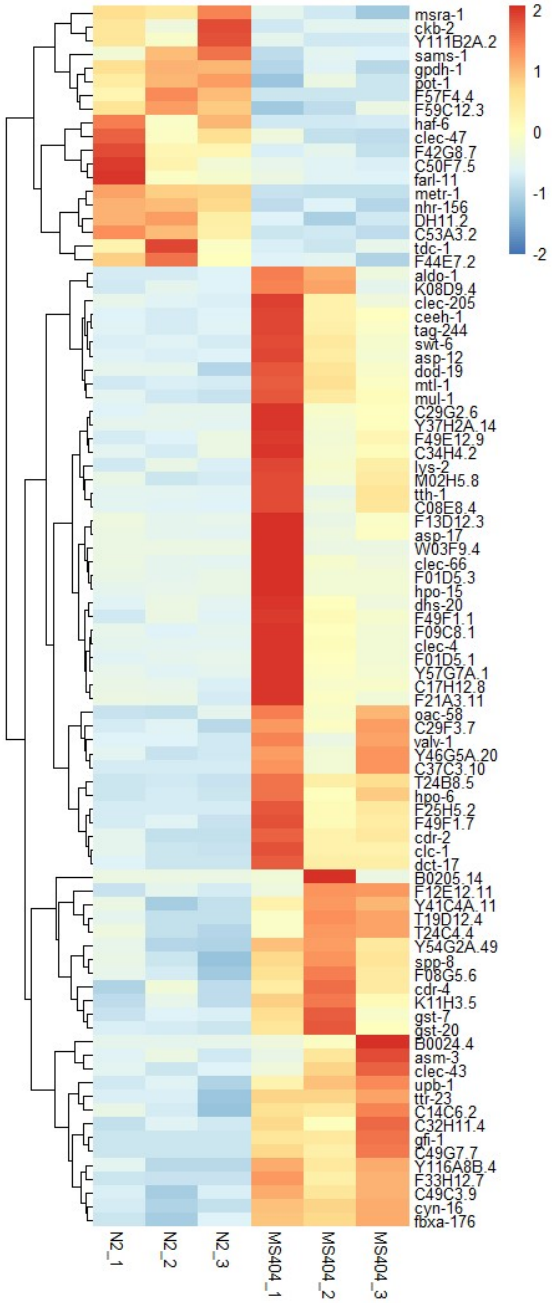


Figure 4-4 Heatmap generated from pheatmap in R using the normalized read counts. Genes are grouped based on expression profile. The X-axis represents the 3 biological replicate for N2 and MS404 and the Y-axis represents the differentially expressed genes identified with edgeR.

WormBase ID	Gene	N2_1	N2_2	N2_3	MS404_1	MS404_2	MS404_3	LogFC	FDR
WBGene00019017	F57F4.4	4695	9864	8158	4	0	0	-12.558	8E-59
WBGene00001581	gfi-1	18	46	37	2293	2186	3966	6.4111	4E-30
WBGene00011979	T24B8.5	50	72	38	1424	736	933	4.52	7E-13
WBGene00045457	F33H12.7	4	1	1	100	57	91	5.1111	7E-13
WBGene00003473	mtl-1	0	3	2	142	82	41	5.6698	4E-10
WBGene00016785	C49G7.7	0	0	0	24	18	36	5.8738	3E-07
WBGene00007440	C08E8.4	7	3	0	236	26	114	5.2114	3E-06
WBGene00001816	haf-6	134	47	108	9	10	12	-3.1405	4E-06
WBGene00016892	C53A3.2	192	162	112	28	30	18	-2.6682	4E-06
WBGene00021518	hpo-6	1	2	1	42	16	30	4.3642	7E-06
WBGene00020606	nhr-284	462	450	404	74	130	60	-2.4463	1E-05
WBGene00007097	B0024.4	1	0	3	12	29	149	5.164	5E-05
WBGene00010988	metr-1	856	736	722	194	183	172	-2.1187	5E-05
WBGene00007867	C32H11.4	15	44	26	186	114	282	2.7621	6E-05
WBGene00000512	ckb-2	396	146	779	12	64	64	-3.2971	0.0006
WBGene00001768	gst-20	12	10	8	50	88	35	2.4109	0.0012
WBGene00011474	aldo-1	0	0	1	24	20	4	5.1607	0.0016
WBGene00009824	gpdh-1	336	402	395	87	148	97	-1.8529	0.0017
WBGene00016845	C50F7.5	208	78	40	22	14	6	-2.9591	0.0017
WBGene00012910	Y46G5A.20	16	6	8	83	28	86	2.6285	0.0023
WBGene00044646	B0205.14	0	0	0	5	96	1	7.9554	0.0028
WBGene00018647	F49F1.7	6	8	3	59	24	30	2.6323	0.0033
WBGene00006649	tth-1	10	10	8	126	15	66	3.0195	0.0037
WBGene00011190	swt-6	50	62	49	380	207	122	2.2448	0.0043
WBGene00015216	valv-1	4	4	2	40	8	36	3.0494	0.0051
WBGene00017678	asp-12	43	36	38	308	150	86	2.2736	0.0063
WBGene00008494	F01D5.3	36	10	32	657	82	82	3.5119	0.0065
WBGene00003091	lys-2	1562	3196	1968	14032	4511	6958	2.3324	0.0068
WBGene00019917	clec-43	5	1	4	10	31	48	3.022	0.0068
WBGene00020579	T19D12.4	11	4	4	20	49	46	2.4799	0.0068
WBGene00008492	F01D5.1	8	13	16	172	46	32	2.8033	0.0074
WBGene00010782	K11H3.5	0	5	0	20	28	12	3.4499	0.0074
WBGene00019580	oac-58	4	3	10	49	18	40	2.4778	0.0077
WBGene00044492	Y54G2A.49	22	8	7	58	66	47	2.098	0.0077
WBGene00012583	clec-4	7	6	6	84	26	16	2.6649	0.0089
WBGene00018424	F44E7.2	98	134	59	26	30	5	-2.3295	0.0097
WBGene00013728	Y111B2A.2	12	6	24	2	0	0	-4.7926	0.0114
WBGene00011668	clec-47	192	68	120	48	9	6	-2.6035	0.0118
WBGene00021965	Y57G7A.1	14	8	13	140	37	30	2.6068	0.0118
WBGene00021002	W03F9.4	0	0	0	180	0	0	9.0512	0.0181

WormBase ID	Gene	N2_1	N2_2	N2_3	MS404_1	MS404_2	MS404_3	LogFC	FDR
WBGene00009125	F25H5.2	0	0	0	14	5	7	4.3906	0.0213
WBGene00018358	F42G8.7	29	12	12	2	4	0	-3.1535	0.022
WBGene00018643	F49F1.1	3	16	8	97	34	24	2.516	0.0228
WBGene00016502	C37C3.10	0	0	0	8	2	8	5.8107	0.0325
WBGene00000892	cyn-16	3	1	2	13	12	14	2.5102	0.0401
WBGene00008296	cdr-2	24	14	14	85	46	52	1.7379	0.0401
WBGene00008739	F13D12.3	12	3	4	140	8	26	3.2525	0.0442
WBGene00020760	T24C4.4	30	12	8	39	85	82	1.9685	0.0442
WBGene00016425	C34H4.2	16	9	24	142	34	48	2.2484	0.0472
WBGene00007807	C29F3.7	76	88	54	258	138	253	1.5824	0.048
WBGene00008017	fbxa-176	2	0	3	12	11	13	2.5962	0.048
WBGene00018393	msra-1	562	516	744	258	202	102	-1.6559	0.048
WBGene00018646	mul-1	5	3	2	30	16	12	2.3522	0.048
WBGene00012757	Y41C4A.11	16	4	8	28	44	40	1.9076	0.0481
WBGene00019329	ceeh-1	20	18	20	105	52	43	1.7679	0.0508
WBGene00015756	C14C6.2	82	56	10	178	165	252	2.0226	0.0557
WBGene00009397	clec-66	16	10	8	134	28	20	2.4849	0.0592
WBGene00022644	dod-19	122	120	42	486	294	186	1.8663	0.0592
WBGene00000983	dhs-20	26	44	28	265	86	53	2.122	0.064
WBGene00015933	C17H12.8	64	63	36	381	108	106	2.0046	0.0709
WBGene00019744	M02H5.8	135	78	90	490	173	272	1.7101	0.0712
WBGene00000522	clc-1	114	85	82	354	200	198	1.4918	0.0752
WBGene00004993	spp-8	456	354	226	806	966	754	1.3917	0.0963

Table 4-1 A list of differentially expressed genes identified using edgeR. Genes displayed contain a False discovery rate (FDR) of less than 0.1. Row one and two represent the WormBase ID and gene name. The following 6 rows are the normalized read counts, and the last two rows are the fold-change and FDR.

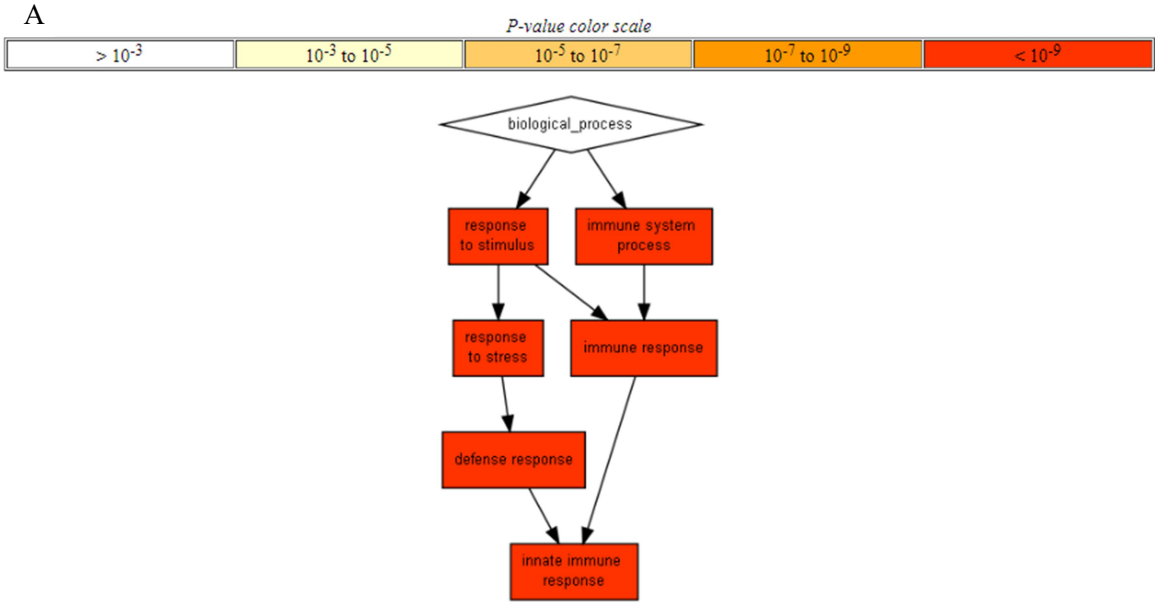
in MS404 could be regulated by gene(s) that function in a pathway that is parallel to *elt-2.*

To better understand the type of genes identified from the differential gene expression analysis, we performed a GO-term analysis for enrichment in Molecular function, Cellular component, and Biological processes was conducted using the software GOrilla (Eden et al. 2009). By grouping the gene ontology identifier terms, we were able to identify enriched functions in genes that are upregulated in MS404 (50 genes), but we were unable to make any correlation with the 13 genes that were down regulated in mutants. The upregulated genes in MS404 show an enrichment of genes that play a role in innate immune response and membrane rafts (Fig 4-5 A, B). The results from the differential gene expression analysis showed that *elt-2* transcripts levels are not different among N2 and MS404 young adult animals. Of the list of differentially expressed gene, the majority of the genes are upregulated in MS404 animals, especially genes that function in innate immune response. Lastly, MS404 animals have increased expression of genes that are associated with membrane rafts, which could be associated with the excess fat storage (lipid droplets) observed in the animals (G. Broitman-Maduro and M. Maduro unpublished data).

4.4.3 Validation using qPCR and GFP Reporter Constructs

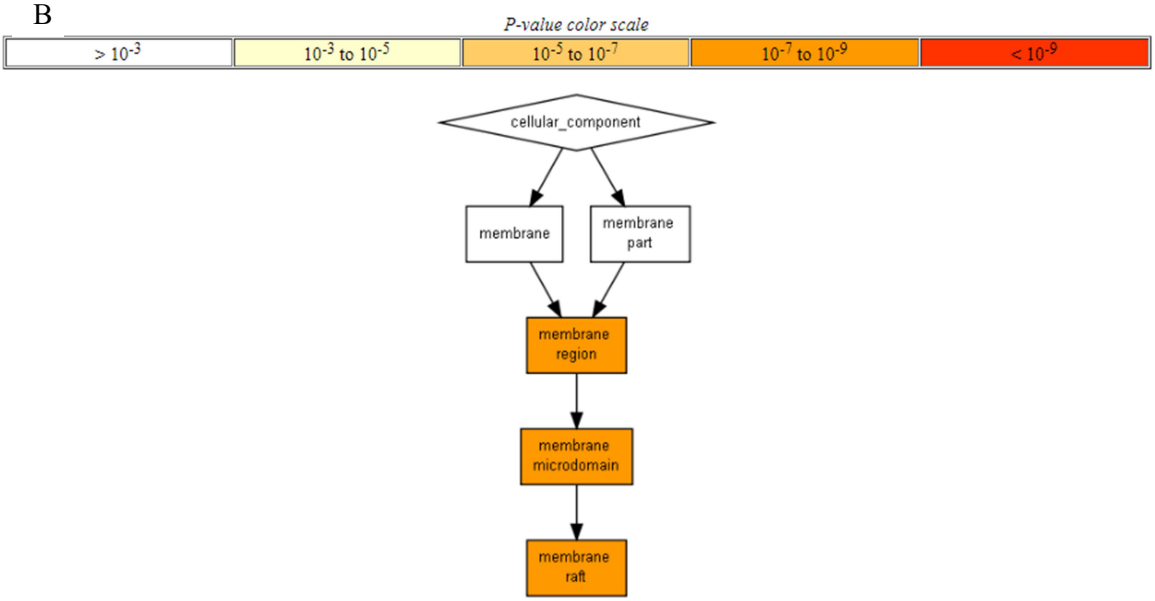
Although our biological replicates appear consistent across the separate trials, other quantification methods were used to validate a subset of our differential expressed genes. Due to the intricate nature of dissecting intestine tissue and the small sample

A



GO term	Description	P-value	FDR q-value	Genes
GO:0006955	immune response	5.23E-18	2.21E-14	T19D12.4 , B0024.4, C29F3.7 , *clc-1*, F01D5.1, C34H4.2, C32H11.4 ,C17H12.8, *lys-2*, *clec-66*, T24C4.4, F01D5.3, C08E8.4 T24B8.5, *hpo-6*, F49F1.1, *dod-19*, *valv-1*
GO:0045087	innate immune response	5.23E-18	1.11E-14	T19D12.4, B0024.4, C29F3.7, *clc-1*, F01D5.1 C34H4.2, *lys-2*, C17H12.8, C32H11.4, *clec-66*, T24C4.4, F01D5.3, C08E8.4, T24B8.5, *hpo-6*, F49F1.1, *dod-19*, *valv-1*
GO:0002376	immune system process	1.06E-17	1.49E-14	T19D12.4, B0024.4, C29F3.7, *clc-1*, F01D5.1 C34H4.2, *lys-2*, C17H12.8, C32H11.4, *clec-66*, T24C4.4, F01D5.3, C08E8.4, T24B8.5, *hpo-6*, F49F1.1, *dod-19*, *valv-1*
GO:0006952	defense response	8.12E-16	8.59E-13	T19D12.4, B0024.4, F01D5.1, C29F3.7, *clc-1*, C34H4.2, *lys-2*, C32H11.4, C17H12.8, *clec-66*, T24C4.4, F01D5.3, C08E8.4, T24B8.5, *hpo-6*, F49F1.1, *dod-19*, *valv-1*
GO:0006950	response to stress	7.78E-12	6.58E-9	T19D12.4, B0024.4, C29F3.7, F01D5.1, *clc-1*, C34H4.2, *mtl-1*, C17H12.8, C32H11.4, *lys-2*, *clec-66*, *cdr-2*, T24C4.4, F01D5.3, C08E8.4, T24B8.5, *hpo-6*, F49F1.1, *dod-19*, *valv-1*
GO:0050896	response to stimulus	9.79E-11	6.9E-8	T19D12.4, B0024.4, C29F3.7, F01D5.1, *clc-1*, C34H4.2, *mtl-1*, C17H12.8, *lys-2*, C32H11.4, *clec-66*, *cdr-2*, T24C4.4, F01D5.3, C08E8.4, T24B8.5, *hpo-6*, F49F1.1, *dod-19*, *valv-1*

B



GO term	Description	P-value	FDR q-value	Genes
GO:0098857	membrane microdomain	3.16E-9	2.67E-6	T19D12.4, B0024.4, C29F3.7, *hpo-6*, *gfi-1*, C34H4.2, C17H12.8, *dod-19*
GO:0045121	membrane raft	3.16E-9	1.34E-6	T19D12.4, B0024.4, C29F3.7, *hpo-6*, *gfi-1*, C34H4.2, C17H12.8, *dod-19*
GO:0098589	membrane region	5.91E-9	1.67E-6	T19D12.4, B0024.4, C29F3.7, *hpo-6*, *gfi-1*, C34H4.2, C17H12.8, *dod-19*

Figure 4-5 Functional analysis was perform using GOrilla on 50 genes, to identify Biological Process and Cellular Component for the 50 genes that were upregulated in MS404. Flow charts and tables generated are an output of the software. The color of the flow chart correlates with p-value of the enriched values. Enriched GO-terms, the number of genes associated to each GO Identifier and their values are reported above. A) Represents genes enriched in Biological Process, functioning mainly in immune response. B) Represents genes enriched in Cellular Components, associating with membrane rafts.

yield, qPCR analysis was done using cDNA that was generated from RNA extracted from whole animals. Some genes identified from the differential gene expression analysis are expressed in other C. elegans tissues, such as neurons and hypodermis. Therefore, quantification by qPCR was limited to only gut specific genes. The qPCR analysis on the genes *gfi-1*, F57F4.4, and *elt-2* correlated with the trends we saw from our CEL-seq data (Figure 4-4). The expression of *elt-2* between N2 and MS404 was not statistically significant (Student T-test, p-value 0.366418343) The expression of *gfi-1* and F57F4.4 in N2 and MS404 were statistically significant (P-value 2.58649E-14 and 5.40476E-11 respectively) (Fig 4-6A), and the directionality of the $\log_2$-fold change mirrors the results from RNA-seq data (Fig. 4-6B).

Transcriptional reporter constructs were generated by cloning in the 1000bp region upstream of the transcription start site for selected DEG genes into a plasmid backbone containing the sequence for GFP. Plasmids were injected into MS2382 animals, which is an *med-1; end-3* double mutant strain that carries a rescue array, consisting of wild-type *end-3* allele and an *unc-119*::mCherry reporter plasmid. After microinjection of the transcriptional reporter construct into MS2382, MS404 + rescue animals carrying the rescue array will express the mCherry rescue marker and the GFP reporter, while MS404 animals will only express the GFP reporter. Animals were picked at the young adult stage, imaged, and quantified using FIJI (Fig. 4-7 A, B, C, D, E). Quantification of GFP expression level was done by tracing the perimeter of the intestine and calculating the average pixel intensity using “Measure” tool within the FIJI (ImageJ) software (Fig. 4-8, 4-9). Data from individual animals were plotted for both strains across

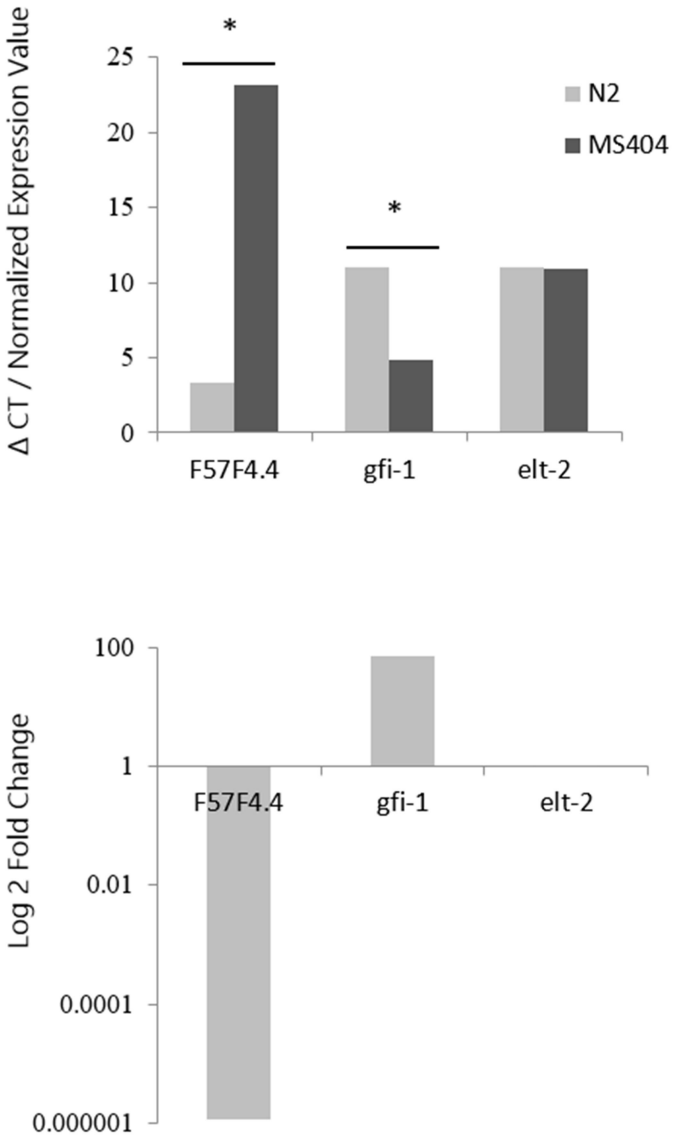
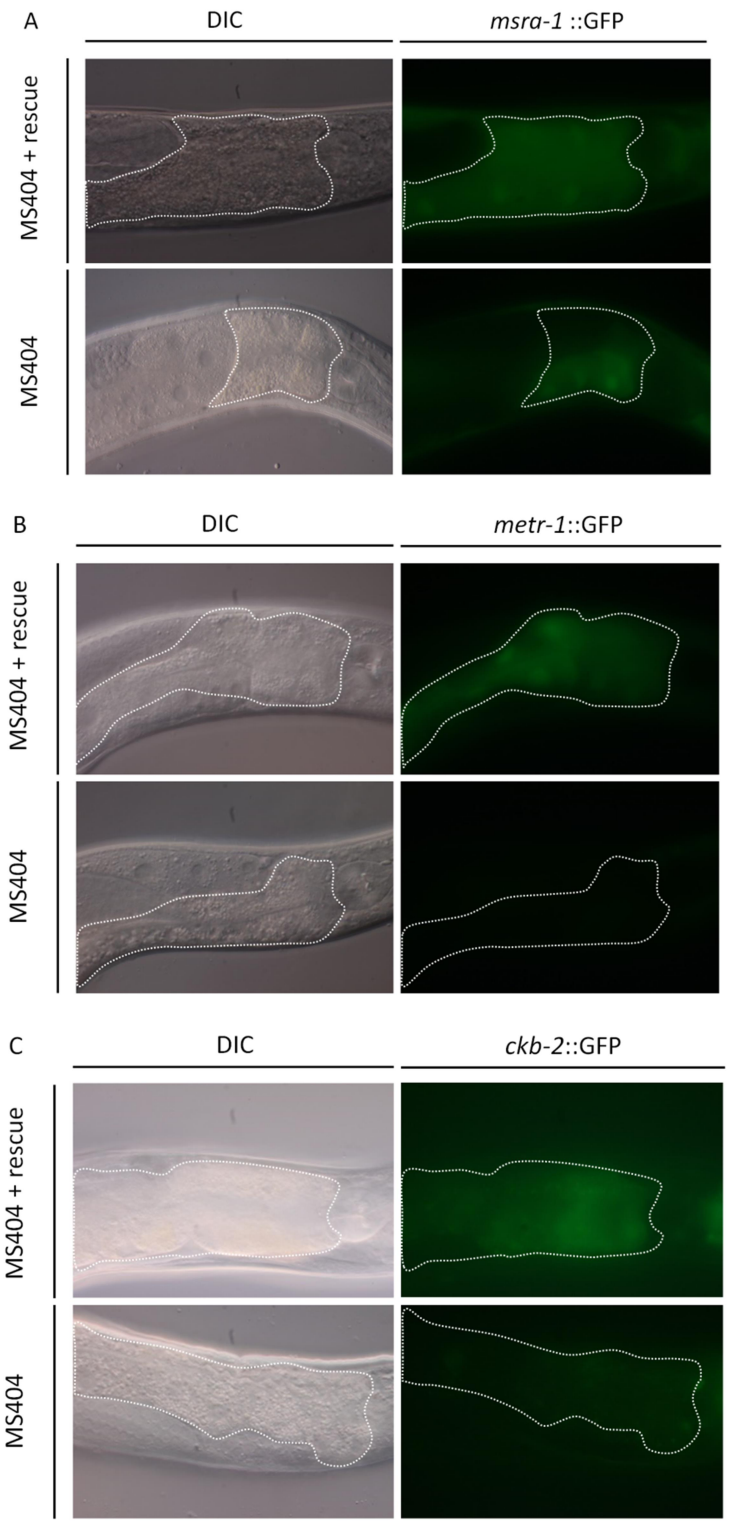


Figure 4-6 qPCR analysis validated differential gene expression of *elt-2, F57F4.4, gfi-1.* Actin gene acted as our house keeping gene and values obtained were used for normalization. Each sample was done in triplicates. A) The ΔCT value/ normalized expression value was calculated. Student t-test showed no significant difference between the expression of *elt-2* in MS404 and N2. However, in both F57F4.4 and *gfi-1*, there was a significant difference (p= 5.4E-11 and p= 2.6E-14 respectively). B) The $\log_2$ fold change was calculated for each gene, comparing the expression of N2 to MS404. The trend of the fold change is the same as the RNA-seq data.


A
DIC
msra-1 ::GFP
MS404 + rescue
MS404
B
DIC
metr-1::GFP
MS404 + rescue
MS404
C
DIC
ckb-2::GFP
MS404 + rescue
MS404

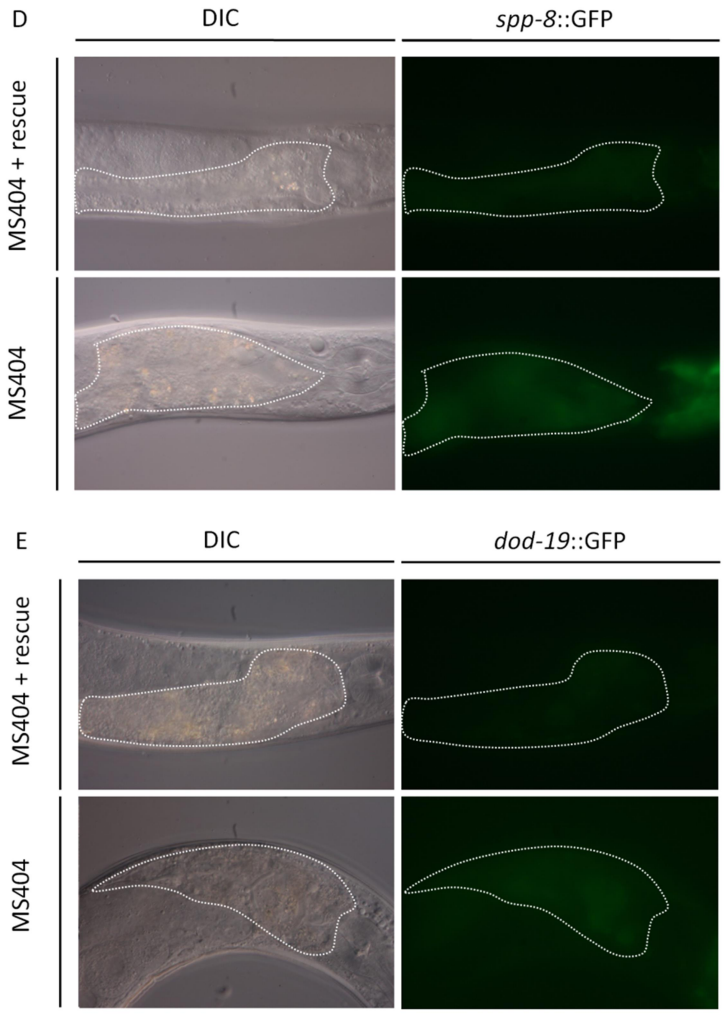


Figure 4-7 A, B, C, D, E GFP reporter constructs were generated to validate a subset of differentially expressed genes. The GFP reporter construct was injected into L4 animals along with the selection markers *rol-6* and *unc-119::CFP*. Young adult animals were mounted on a standard 5% agar pad and image. The same camera setting was used to image the same GFP reporter construct.

five different reporter constructs. The average GFP intensity for each strain is reported in Table 4-2. A student T-Test was calculated between the two strains, and the p-values are 7.3E-05, 7.0E-06, 8.2E-08, 0.0014, and 2.5E-09, respectively. The GFP expression value obtained for each construct was also consistent with the tissue specific RNA- seq data. Upon identification of 63 differentially expressed genes among N2 and MS404, we quantified the gene expression for several of these genes using qPCR and GFP reporter constructs. The data from both methods support the differentially expression genes identified using edgeR as truly being differentially expression between N2 and MS404.

4.5 Conclusion

Using our endoderm specification mutants, we assessed whether delayed expression of the gut differentiation gene *elt-2* resulted in adult pleiotropic defects seen in surviving HGS adults. Although the intestine of HGS surviving adults look normal, we hypothesized that the integrity of the gut is compromised, which leads to starvation and pleiotropic phenotypes such as sterility. Proper development requires the timely activation of essential genes, but our MS404 animals display a delayed onset of *elt-2* (Choi, Broitman-Maduro, Maduro 2017). Due to the delayed expression of the *elt-2*, which is responsible for the activation of all intestine specific genes, key genes necessary for a normally functioning gut might only be partially activated or never activated. The manifestation of pleiotropic traits can be attributed to differential gene expression, as the consequence of stochastic *elt-2* expression. The differential gene expression analysis identified a transcription factor, NHR-156. This transcription factor showed a reduced

Reporter construct	Strain	Mean GFP expression
msra-1::GFP	MS404 + rescue	14.629 ± 1.337
msra-1::GFP	MS404	7.892 ± 0.599
metr-1::GFP	MS404 + rescue	15.335 ± 0.919
metr-1::GFP	MS404	6.80 ± 1.34
ckb-2::GFP	MS404 + rescue	19.5 ± 1.25
ckb-2::GFP	MS404	10.2 ± 0.322
spp-8::GFP	MS404 + rescue	10.5 ± 0.238
spp-8::GFP	MS404	12.8 ± 0.607
dod-19::GFP	MS404 + rescue	7.67 ± 0.230
dod-19::GFP	MS404	11.2 ± 0.383

Table 4-2 Mean expression of GFP reporter constructs for MS404 + rescue and MS404 animals. Mean value is reporter with ± SEM.

expression in MS404 (log_2-fold change of -2.44). Reduced expression of a transcription factor supports the notion that delayed expression of *elt-2* in the endoderm specification mutants leads to missed or partial activation of an essential downstream target gene (*nhr-156*), which leads to a missed activation of genes in the proper spatiotemporal context and attributes to a dysfunctional intestine.

The sequencing data identified 63 differentially expressed genes, of which 50 of the genes are upregulated in MS404 animals. Functional categorization of molecular mechanism and cellular components showed that of the genes that are upregulated in MS404 animal are enriched with genes that associate with innate immune response and membrane raft genes. The *C. elegans* intestine does more than merely digest food, it is also the site of innate immunity. Although the animals are fed an ad libitum diet of standard *E. coli* OP50, they display characteristics as if they have undergone a long period of starvation or were fed an unfavorable diet. The differential expression of genes might have lead the mutant animals to A) not be able to properly absorb the nutrients or B) trick the animals that they are being fed a harmful diet. If the animals were not able to absorb the nutrients, then they would be malnourished and may display signs of starvation, such as egg-laying defect and storage of excess lipids. The enrichment of genes associated with membrane rafts correlates with the extra lipid droplets that the animals carry. The constant activation of innate immune response genes requires a large allocation of cellular energy, which puts a great amount of stress on the animals. The constant stress on the animal could be attributed to the shorten lifespan of MS404 in comparison with N2 (K. Dollan and M. Maduro unpublished results).

Our transcriptomic data support the pleiotropic defects observed in our MS404 animals, and uncovered a list of differentially expressed genes between N2 and a gut specification mutant. The study has shed light on functional differences in genes that are expressed in each strain, but more work still needs to be done to utilize the transcriptomic dataset to its full potential. Some key genes that should be followed up on are the transcription factor *nhr-156*, and the genes *gfi-1* and F57F4.4. These two genes are significantly differentially expressed, with *gfi-1* being highly upregulated in MS404 (log_2-fold change of 6.5), and F57F4.4 being highly expressed in N2 (log_2-fold change of -12). These genes share ~95% similarity in their protein coding sequence, and both genes are associated with membrane rafts. Future experiments will aim to decipher why these animals are expressing different types of membrane raft genes, and whether it is due to the excess lipid droplets that MS404 animals store or the excess lipid a result of the differential gene expression.

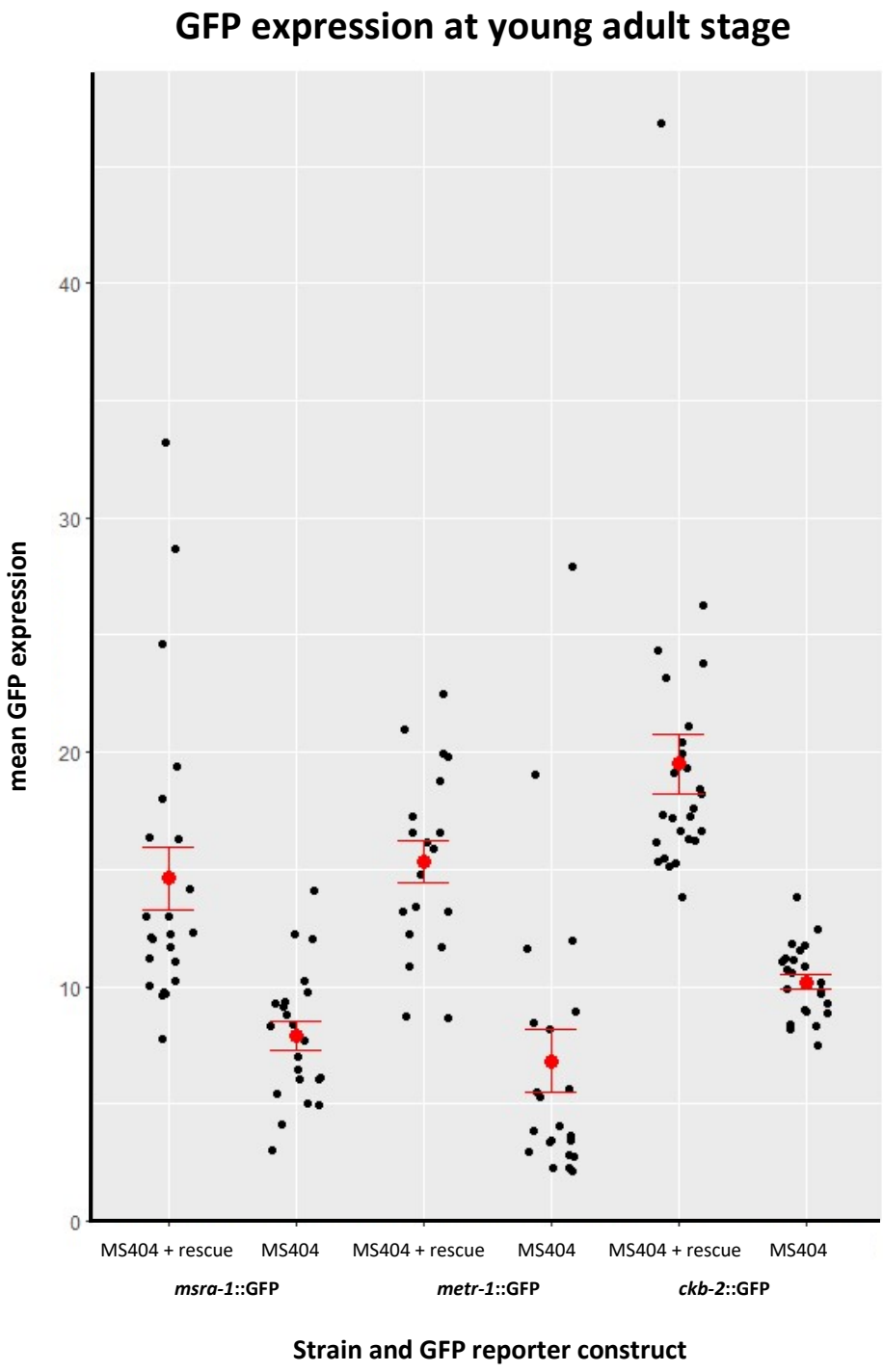


Figure 4-8 The expression of GFP was significantly enhanced in the MS404 + rescue strains for all three sets (p= 7.3E-05, p= 7.0E-06, and p= 8.2E-08, student t-test). The data above are grouped by the GFP reporter construct (*msra-1, metr-1,* and *ckb-2*).

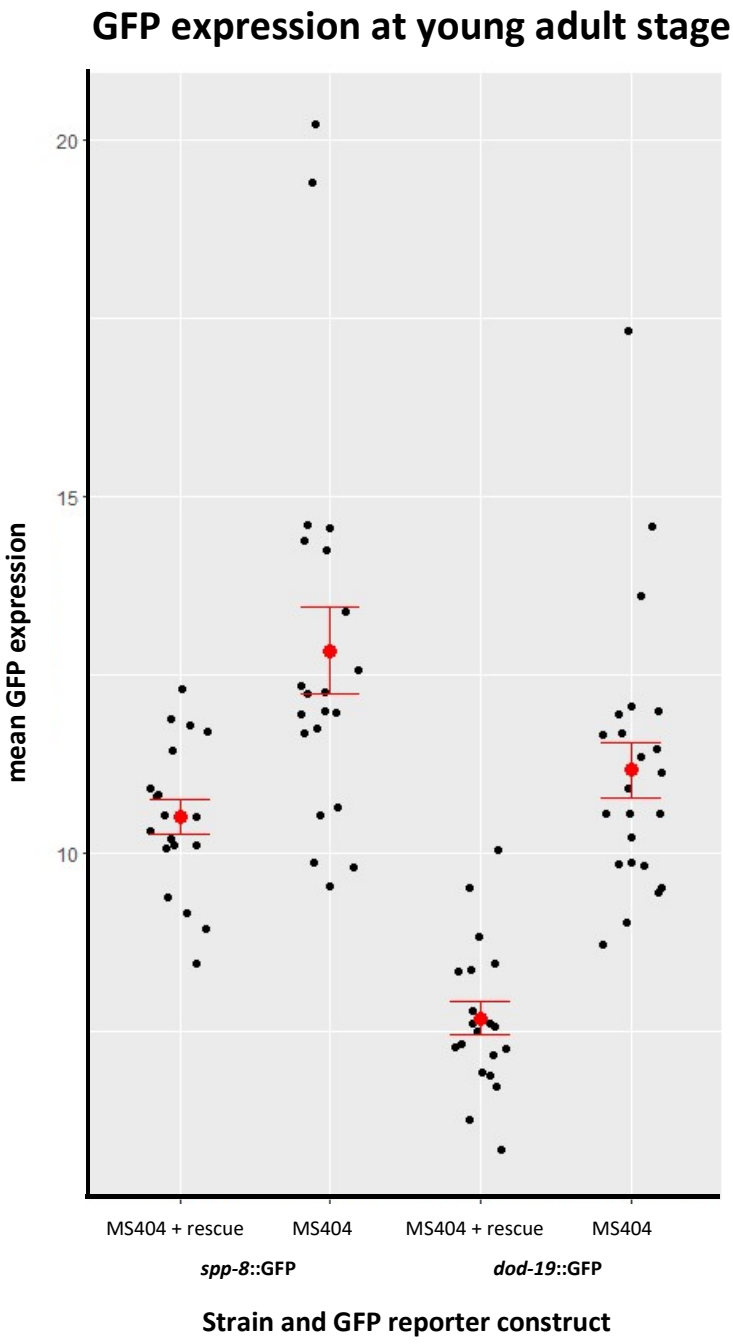


Figure 4-9 The expression of GFP was significantly enhanced in the MS404 strains for both sets (p= 0.0014 and p= 2.5E-09, student t-test). The data above are grouped by the GFP reporter construct (*spp-8* and *dod-19*).

4.6 References

Bowerman, B, BA Eaton, JR Priess. 1992. skn-1, a maternally expressed gene required to specify the fate of ventral blastomeres in the early C. elegans embryo. Cell 68:1061-1075.

Broitman-Maduro, G, MF Maduro, JH Rothman. 2005. The noncanonical binding site of the MED-1 GATA factor defines differentially regulated target genes in the C. elegans mesendoderm. Dev Cell 8:427-433.

Choi, H, G Broitman-Maduro, MF Maduro. 2017. Partially compromised specification causes stochastic effects on gut development in C. elegans. Dev Biol 427:49-60.

Eden, E, R Navon, I Steinfeld, D Lipson, Z Yakhini. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10:48.

Fukushige, T, MJ Hendzel, DP Bazett-Jones, JD McGhee. 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 96:11883-11888.

Gibson, DG, L Young, RY Chuang, JC Venter, CA Hutchison, 3rd, HO Smith. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343-345.

Haenni, S, Z Ji, M Hoque, et al. 2012. Analysis of C. elegans intestinal gene expression and polyadenylation by fluorescence-activated nuclei sorting and 3'-end-seq. Nucleic Acids Res 40:6304-6318.

Hashimshony, T, N Senderovich, G Avital, et al. 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol 17:77.

Hashimshony, T, F Wagner, N Sher, I Yanai. 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep 2:666-673.

Lightfoot, JW, VM Chauhan, JW Aylott, C Rodelsperger. 2016. Comparative transcriptomics of the nematode gut identifies global shifts in feeding mode and pathogen susceptibility. BMC Res Notes 9:142.

Love, MI, W Huber, S Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

Lowry, JA, R Gamsjaeger, SY Thong, W Hung, AH Kwan, G Broitman-Maduro, JM Matthews, M Maduro, JP Mackay. 2009. Structural analysis of MED-1 reveals

unexpected diversity in the mechanism of DNA recognition by GATA-type zinc finger domains. J Biol Chem 284:5827-5835.

Maduro, MF, G Broitman-Maduro, H Choi, F Carranza, A Chia-Yi Wu, SA Rifkin. 2015. MED GATA factors promote robust development of the C. elegans endoderm. Dev Biol 404:66-79.

Maduro, MF, RJ Hill, PJ Heid, ED Newman-Smith, J Zhu, JR Priess, JH Rothman. 2005. Genetic redundancy in endoderm specification within the genus Caenorhabditis. Dev Biol 284:509-522.

Maduro, MF, MD Meneghini, B Bowerman, G Broitman-Maduro, JH Rothman. 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 7:475-485.

McGhee, JD. 2007. The C. elegans intestine. WormBook:1-36.

McGhee, JD, T Fukushige, MW Krause, 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.

Owraghi, M, G Broitman-Maduro, T Luu, H Roberson, MF Maduro. 2010. Roles of the Wnt effector POP-1/TCF in the C. elegans endomesoderm specification gene network. Dev Biol 340:209-221.

Pauli, F, Y Liu, YA Kim, PJ Chen, SK Kim. 2006. Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in C. elegans. Development 133:287-295.

Raj, A, SA Rifkin, E Andersen, A van Oudenaarden. 2010. Variability in gene expression underlies incomplete penetrance. Nature 463:913-918.

Risso, D, K Schwartz, G Sherlock, S Dudoit. 2011. GC-content normalization for RNA-Seq data. BMC Bioinformatics 12:480.

Robinson, MD, DJ McCarthy, GK Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.

Rocheleau, CE, WD Downs, R Lin, C Wittmann, Y Bei, YH Cha, M Ali, JR Priess, CC Mello. 1997. Wnt signaling and an APC-related gene specify endoderm in early C. elegans embryos. Cell 90:707-716.

Rocheleau, CE, J Yasuda, TH Shin, R Lin, H Sawa, H Okano, JR Priess, RJ Davis, CC Mello. 1999. WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in C. elegans. Cell 97:717-726.

Sambrook, J, DW Russell. 2001. Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Seidel, HS, J Kimble. 2011. The oogenic germline starvation response in C. elegans. PLoS One 6:e28074.

Spencer, WC, G Zeller, JD Watson, et al. 2011. A spatial and temporal map of C. elegans gene expression. Genome Res 21:325-341.

Stiernagle, T. 2006. Maintenance of C. elegans. WormBook:1-11.

Thorpe, CJ, A Schlesinger, JC Carter, B Bowerman. 1997. Wnt signaling polarizes an early C. elegans blastomere to distinguish endoderm from mesoderm. Cell 90:695-705.

CHAPTER 5

Conclusion

5.1 Conclusion

Gene regulatory networks are an essential component of the developmental process. The timely activation of these networks sets forward a cascade of gene expression that ultimately leads to differentiation of cells and formation of tissues and organs. Due to the importance of the spatiotemporal activation of these networks they have evolved to be highly robust, through duplication of core genes and redundant wiring of multiple activators (Kitano 2004; Gibson 2009). The layers of robustness built into gene regulatory networks are used to buffer the stochastic nature of gene expression, which is caused by intrinsic and extrinsic factors. Although it is known that the robustness of gene regulatory networks is present to buffer stochastic noise, the exact mechanisms by which they do so are unclear (Wagner 1996; Keller 2002; Siegal, Bergman 2002; Bergman, Siegal 2003). To better understand this phenomenon, we have utilized the endoderm gene regulatory network of *Caenorhabditis elegans* to ask a series of questions pertaining to the robustness of gut development during different stages of the animal's life. In this series of experiments, we have described how stochastic gene expression is buffered, how stochastic gene expression affects cell differentiation, and revealed the effect on the integrity and function of tissues where cells were just able to undergo differentiation.

In using a series of allelic mutant strains where we have compromised the endoderm gene regulatory network, we have generated *end-1,3*(MED-) and *med-1; end-3* double mutant (MS404) strains that are highly stochastic and susceptible to various sources of noise (Maduro et al. 2005a; Maduro et al. 2007; Maduro et al. 2015; Choi, Broitman-Maduro, Maduro 2017). First, we have used these *end-1,3*(MED-) strains to study the way animals buffer stochastic gene expression during early embryogenesis. Because specification of the endoderm occurs early during development we have focused our study primarily on maternal factors, since these factors are known to be early acting. Using RNAi -by-feeding, we have screened over 3,000 genes to identify modifiers of gut specification, by screening for proportions of embryos expressing the gut differentiation marker *elt-2*::GFP in the absence of particular maternal products when compared to the control. Thus far, we have identified several diverse maternal pathways, (including transcriptional regulation, metabolism, and apoptosis) that influence the ability of progeny embryos to make gut, suggesting that embryonic gene expression is highly sensitive to many sources of variation.

Previous studies suggested that gut specification was an "all-or-none" phenomenon, where either the entire E lineage will specify a gut fate or fail to do so and adopt a different fate. Instead, the specification of E lineage cells is dependent on the acquisition of a threshold level of *end-1* and *end-3* transcripts (Raj et al. 2010). In our *end-1,3*(MED-) animals, we have observed animals that specify a range of gut cell number, and saw that in surviving adult animals there is a minimum number of gut cells necessary for survival. We conclude that the specification of the entire E lineage is not a

binary “all-or-none” event, but rather, follows a stratified model of specification, where each E lineage descendant cell has the potential to be a gut cell based on the individual cell’s ability to acquire the threshold level of transcripts necessary for gut specification. By using a two-reporter method, where individual E lineage descendant cells were labeled with an mCherry reporter and differentiated gut cells were labeled with a GFP marker, and imaging hundreds of embryos, we concluded that specification of endoderm is not “all-or-none” at the level of the whole lineage, but rather at the level of the individual E lineage descendant cells. Our results show the lack of robustness during endoderm specification affects cell division patterns, the ability of E descendants to adopt a gut fate, and gut morphogenesis, in a highly stochastic manner. We also noted that differentiation of endoderm cells occurs later in our *end-1,3*(MED-) animals in comparison to wild-typed animals, and E descendant cells that do not adopt an endoderm fate will adopt a C-like fate (Maduro et al. 2005b; Owraghi et al. 2010). Of the adult *end-1,3*(MED-) animals where E lineage descendants do adopt an endoderm fate, we saw steady-state levels of the differentiation factor ELT-2, which appears to be normal regardless of the number of gut nuclei present.

Our *end-1,3*(MED-) strains vary in the number of gut cells produced. This is the result of the varying level of *end-1,3* transcripts that individual gut descendant cells acquire during development. Of our *end-1,3*(MED-) animals that barely acquire enough transcripts to specify a gut fate and hatch out as larvae, a percentage of the surviving adults display pleiotropic phenotypes and have excess lipid storage. This raises the question: Does the intestine of animals that barely generated enough transcripts to specify

an endoderm fate and activate *elt-2* gene expression function normally? Could these adult pleiotropic phenotypes be the consequence of differential gene expression due to delayed expression of *elt-2*, a gene that is responsible for the activation of all intestinal genes? Our intestine-specific transcriptomic analysis showed that our MS404 animals expressed upregulation in genes that are involved in innate immune response as well as membrane rafts. The animals were fed the standard *E. coli* OP50 diet, and the increased innate immune response genes in our MS404 animals suggests that the function of the intestine has been compromised, therefore triggering immune response genes in the absence of a pathogen or toxin. The constant activation of innate immune response genes requires a large allocation of cellular energy, which puts a great amount of stress on the animals. The constant stress on the animal could cause the short lifespan of the MS404 animals in comparison with N2 (K. Dollan and M. Maduro unpublished results). The enrichment of genes associated with membrane rafts correlates with the extra lipid droplets that the animals carry. Although the intestine of our MS404 animal looks physically normal, it might not function normally, thereby preventing the animal from properly absorbing nutrients. This would result in malnourishment and may display signs of starvation, such as storage of excess lipids (Palgunow, Klapper, Doring 2012; Maduro et al. 2015). Our transcriptomic data also showed that levels of *elt-2* in wild-type and MS404 animals were not significantly different at the young adult stage, which suggests that the defects seen in the adult animals were not due to improper activation of the key gut differential gene *elt-2*. These results suggest that the alternative transcriptional state could be regulated by gene(s) that function in a pathway that is parallel to *elt-2*.

Taken together, our results demonstrate that robustness of gut development in *C. elegans* occurs at several different life stages. It requires the maternal pathways such as global regulation of gene expression to buffer noise during the early stages of embryogenesis. The specification and formation of the Endoderm lineage is not an "all—or-none" event, but rather follows a stratified mode of specification. The specification of E descendant cells is dependent on the stochastic acquisition of threshold level of *end-1,3* transcripts. And, the phenotypes that develop during adulthood in a proportion of surviving MS404 adults can either lead to or be the result of increased expression of genes that functions in immune response and formation of lipid rafts.

Future studies will need to be conducted to study the exact mechanism in which certain maternal pathways can buffer the stochastic noise during early embryogenesis. From the intestinal transcriptomic data, there are a few genes that warrant further study. Genes such as *nhr-156*, which encodes a transcription factor and is involved in lipid storage should be further studied (Reece-Hoyes et al. 2007). This gene was downregulated in the MS404 strains, which suggests that it might play an essential role in activating genes that are important for the function and integrity of the intestine. Other genes that should be considered are the pair of genes, *gfi-1* and F57F4.4, which share ~95% commonalty in their protein coding region (Wormbase.org). One is highly expressed in wild-type animals (F57F4.4) while the other is highly expressed in the specification MS404 animals (*gfi-1*). Both genes are associated with membrane rafts, so it would be interesting to see why one type of membrane raft is expressed in wild-type animals and another in *end-1,3*(MED-) animals.

Lastly, we noticed a gradient GFP expression from anterior to posterior for some of our GFP reporters. We also noticed a gradient GFP expression in MS404 animals, but not in wild-type. The gradient of GFP expression in only MS404 animals suggest that differences in gene expressions are localized to a sub-region of the intestine. The whole intestine is derived from a single progenitor cell, and each cell is required to activate the same differential factor *elt-2*, which ultimately drives the expression of all intestine specific genes. Although the intestine is made up of all the same intestinal cells, they are functionally different. The anterior cells contribute immune responses and the breakdown of food source (Kim et al. 2002; Kim et al. 2004; Mango 2007), while the posterior half metabolizes and defecates (Liu, Thomas 1994; Ghafouri, McGhee 2007). Data from MS404 animals suggest that they partially upregulate/ downregulate gene expression in a sub-region of the intestine. By studying the types of genes expressed in the anterior and posterior half of the intestine in wild type animals, we can better understand the normal expression profile in these regions of the intestine, and can test whether the compromised functionally of the gut is due to intestinal cells receiving the incorrect positional information, leading to their acquisition of the wrong functional fate.

5.2 References

Bergman, A, ML Siegal. 2003. Evolutionary capacitance as a general feature of complex gene networks. Nature 424:549-552.

Choi, H, G Broitman-Maduro, MF Maduro. 2017. Partially compromised specification causes stochastic effects on gut development in C. elegans. Dev Biol 427:49-60.

Ghafouri, S, JD McGhee. 2007. Bacterial residence time in the intestine of Caenorhabditis elegans. Nematology 9:87-91.

Gibson, G. 2009. Decanalization and the origin of complex disease. Nat Rev Genet 10:134-140.

Keller, EF. 2002. Developmental robustness. Ann N Y Acad Sci 981:189-201.

Kim, DH, R Feinbaum, G Alloing, et al. 2002. A conserved p38 MAP kinase pathway in Caenorhabditis elegans innate immunity. Science 297:623-626.

Kim, DH, NT Liberati, T Mizuno, H Inoue, N Hisamoto, K Matsumoto, FM Ausubel. 2004. Integration of Caenorhabditis elegans MAPK pathways mediating immunity and stress resistance by MEK-1 MAPK kinase and VHP-1 MAPK phosphatase. Proc Natl Acad Sci U S A 101:10990-10994.

Kitano, H. 2004. Biological robustness. Nat Rev Genet 5:826-837.

Liu, DW, JH Thomas. 1994. Regulation of a periodic motor program in C. elegans. J Neurosci 14:1953-1962.

Maduro, MF, G Broitman-Maduro, H Choi, F Carranza, A Chia-Yi Wu, SA Rifkin. 2015. MED GATA factors promote robust development of the C. elegans endoderm. Dev Biol 404:66-79.

Maduro, MF, G Broitman-Maduro, I Mengarelli, JH Rothman. 2007. Maternal deployment of the embryonic SKN-1-->MED-1,2 cell specification pathway in C. elegans. Dev Biol 301:590-601.

Maduro, MF, RJ Hill, PJ Heid, ED Newman-Smith, J Zhu, JR Priess, JH Rothman. 2005a. Genetic redundancy in endoderm specification within the genus Caenorhabditis. Dev Biol 284:509-522.

Maduro, MF, JJ Kasmir, J Zhu, JH Rothman. 2005b. The Wnt effector POP-1 and the PAL-1/Caudal homeoprotein collaborate with SKN-1 to activate C. elegans endoderm development. Dev Biol 285:510-523.

Mango, SE. 2007. The C. elegans pharynx: a model for organogenesis. WormBook:1-26.

Owraghi, M, G Broitman-Maduro, T Luu, H Roberson, MF Maduro. 2010. Roles of the Wnt effector POP-1/TCF in the C. elegans endomesoderm specification gene network. Dev Biol 340:209-221.

Palgunow, D, M Klapper, F Doring. 2012. Dietary restriction during development enlarges intestinal and hypodermal lipid droplets in Caenorhabditis elegans. PLoS One 7:e46198.

Raj, A, SA Rifkin, E Andersen, A van Oudenaarden. 2010. Variability in gene expression underlies incomplete penetrance. Nature 463:913-918.

Reece-Hoyes, JS, J Shingles, D Dupuy, CA Grove, AJ Walhout, M Vidal, IA Hope. 2007. Insight into transcription factor gene duplication from Caenorhabditis elegans Promoterome-driven expression patterns. BMC Genomics 8:27.

Siegal, ML, A Bergman. 2002. Waddington's canalization revisited: developmental stability and evolution. Proc Natl Acad Sci U S A 99:10528-10532.

Wagner, A. 1996. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. Biol Cybern 74:557-567.

Appendix A. Differential gene expression analysis on other specification mutants

Libraries for (MS1810, N2, MS1809, MS404, and MS1548) were sequences in triplicates and processed following the methods reported in Chapter 4. Differential expression genes with a FDR of less than .1 are reported below.

MS404 vs MS1548	ID	logFC	logCPM	PValue	FDR
	WBGene00019017	10.49912	10.71855	2.25E-39	1.12E-35
	WBGene00001581	-6.48922	7.869656	3.8E-28	9.51E-25
	WBGene00017498	-4.54399	6.602312	1.32E-09	2.2E-06
	WBGene00001725	9.778625	4.170203	7.72E-09	9.65E-06
	WBGene00015344	-4.9386	3.809509	3.91E-07	0.00039
	WBGene00020606	2.89073	8.126632	2.67E-06	0.002223
	WBGene00011046	4.102572	5.363409	7.36E-06	0.00485
	WBGene00018393	2.82594	9.095176	7.77E-06	0.00485
	WBGene00012631	2.813427	5.168807	9.12E-06	0.005066
	WBGene00017349	3.217785	5.536438	2.14E-05	0.010426
	WBGene00003163	-3.15814	4.181543	2.3E-05	0.010426
	WBGene00021002	-7.3805	3.125504	3.75E-05	0.015627
	WBGene00018643	-3.48225	3.130308	6.01E-05	0.023112
	WBGene00009772	-6.15069	1.776541	0.000146	0.052236
	WBGene00012634	3.965846	4.511578	0.000158	0.052764
	WBGene00008621	-2.06807	3.930216	0.000242	0.075699
	WBGene00005141	2.779412	3.398327	0.0003	0.088047
	WBGene00006466	2.820456	3.873331	0.000336	0.093351
MS404 vs MS1809	ID	logFC	logCPM	PValue	FDR
	WBGene00019017	9.794847	10.71855	7.38E-35	3.69E-31
	WBGene00001581	-8.25496	7.869656	4.63E-26	1.16E-22
	WBGene00017498	-9.474	6.602312	7.17E-14	1.19E-10
	WBGene00003473	-6.48882	3.916828	9.76E-10	1.22E-06
	WBGene00020909	-4.11861	6.379305	2.9E-08	2.89E-05
	WBGene00020606	3.454603	8.126632	1.25E-07	0.000104
	WBGene00011190	-4.72751	6.391242	1.6E-07	0.000114
	WBGene00013650	-2.89436	6.36957	6.76E-07	0.000422
	WBGene00018643	-6.4898	3.130308	1.25E-06	0.000693
	WBGene00007963	-3.4502	5.266101	1.53E-06	0.000765

MS404 vs MS1809	WBGene00012089	9.01571	3.22634	1.95E-06	0.000823
	WBGene00008621	-3.36632	3.930216	1.98E-06	0.000823
	WBGene00006418	-7.88913	4.041526	2.4E-06	0.000922
	WBGene00000983	-4.0508	4.908285	2.81E-06	0.001002
	WBGene00021965	-4.7436	4.282567	3.46E-06	0.001152
	WBGene00194674	-7.91957	3.09734	9.78E-06	0.003024
	WBGene00000407	-2.55029	5.587532	1.03E-05	0.003024
	WBGene00016875	-3.55901	4.393396	1.1E-05	0.003045
	WBGene00012631	2.836636	5.168807	1.48E-05	0.003883
	WBGene00018393	2.79314	9.095176	2.01E-05	0.005014
	WBGene00011333	2.899605	5.535431	2.16E-05	0.005014
	WBGene00021492	-6.95173	3.271604	2.35E-05	0.005014
	WBGene00006986	5.305366	3.045299	2.36E-05	0.005014
	WBGene00007365	-2.46951	9.050518	2.41E-05	0.005014
	WBGene00194742	-2.61756	5.408498	2.95E-05	0.005905
	WBGene00011979	-4.15024	7.269957	3.74E-05	0.007181
	WBGene00016845	5.087291	6.463568	3.88E-05	0.007185
	WBGene00007130	-2.36495	5.921371	6.07E-05	0.010826
	WBGene00010440	-2.22341	9.618826	7.42E-05	0.012784
	WBGene00004989	-2.5158	8.686785	8.46E-05	0.014087
	WBGene00007867	-2.61073	4.73691	0.000104	0.016765
	WBGene00001226	3.033031	6.181435	0.000117	0.018237
	WBGene00008068	-2.83767	5.754734	0.00014	0.020701
	WBGene00018036	-1.83759	8.153702	0.000141	0.020701
	WBGene00001816	2.893436	4.745815	0.000166	0.023197
	WBGene00003474	-1.80965	9.805017	0.000167	0.023197
	WBGene00044045	-2.48304	5.845126	0.000175	0.023274
	WBGene00021491	-5.19019	4.118392	0.000177	0.023274
	WBGene00022862	2.834882	3.294451	0.000182	0.023274
	WBGene00012412	-3.533	3.764316	0.000188	0.023495
	WBGene00019060	-2.15278	6.221458	0.000209	0.02546
	WBGene00004411	3.451683	3.521849	0.000223	0.026489
	WBGene00017238	-2.42767	6.320762	0.000233	0.027117
	WBGene00009824	-2.33202	6.350844	0.00024	0.027298
	WBGene00011205	-2.48709	5.052869	0.000249	0.027622
	WBGene00195093	2.601271	4.295119	0.000264	0.028378
	WBGene00016037	7.37252	1.794157	0.000267	0.028378
	WBGene00009898	-2.58015	8.142088	0.000285	0.029162
	WBGene00019654	2.307274	5.885119	0.000287	0.029162

MS404 vs MS1809	WBGene00009396	-2.04749	9.119089	0.000292	0.029162
	WBGene00000219	-1.96796	11.44983	0.00032	0.031305
	WBGene00014251	-2.14552	5.843476	0.000343	0.032181
	WBGene00003397	4.001603	3.83376	0.00035	0.032181
	WBGene00012727	6.655084	1.623868	0.000353	0.032181
	WBGene00016061	-2.77179	5.201152	0.000354	0.032181
	WBGene00013730	2.197168	4.180491	0.000384	0.034291
	WBGene00044492	-3.49729	4.323285	0.000399	0.034652
	WBGene00000984	-2.08823	4.991492	0.000402	0.034652
	WBGene00015813	2.375543	3.628578	0.000423	0.035816
	WBGene00012730	2.449056	3.690285	0.000443	0.036431
	WBGene00014000	-1.91066	6.870091	0.000445	0.036431
	WBGene00022412	2.287769	3.807662	0.000505	0.040682
	WBGene00006626	2.604512	6.004925	0.000513	0.040719
	WBGene00012747	-1.91569	6.090097	0.00056	0.043296
	WBGene00013576	-2.56591	3.846609	0.000563	0.043296
	WBGene00015105	3.116265	3.291549	0.000579	0.043833
	WBGene00007430	-3.44829	3.3826	0.000598	0.044605
	WBGene00006649	-3.82391	3.610791	0.000634	0.04628
	WBGene00015186	-1.98818	6.724491	0.000639	0.04628
	WBGene00015802	-1.97958	6.780664	0.000656	0.046853
	WBGene00020446	-1.84785	5.771449	0.000731	0.050767
	WBGene00016406	-2.45849	3.656666	0.000736	0.050767
	WBGene00020259	2.62778	3.027714	0.000748	0.050767
	WBGene00010272	2.63178	3.571584	0.000752	0.050767
	WBGene00017131	-2.34843	5.229899	0.000819	0.054564
	WBGene00008167	-2.30894	4.873619	0.000854	0.056132
	WBGene00000975	-2.57632	3.796448	0.000875	0.05641
	WBGene00009521	3.125212	3.910922	0.000891	0.05641
	WBGene00012634	3.533802	4.511578	0.000897	0.05641
	WBGene00007362	-1.86758	6.964895	0.000909	0.05641
	WBGene00007435	3.386971	1.937094	0.000914	0.05641
	WBGene00022200	2.245773	4.239144	0.000934	0.056895
	WBGene00003392	2.62986	3.622998	0.000958	0.057698
	WBGene00016913	6.586625	1.443505	0.000982	0.05798
	WBGene00008494	-3.51279	5.689561	0.000995	0.05798
	WBGene00006562	-4.39003	3.657355	0.000998	0.05798
	WBGene00019762	3.181877	3.742009	0.001131	0.064555
	WBGene00045013	-2.04903	4.668684	0.00114	0.064555

MS404 vs MS1809	WBGene00016048	-1.87567	8.04219	0.00115	0.064555
	WBGene00006621	-3.37726	4.234114	0.001197	0.066412
	WBGene00022497	-1.85361	11.94667	0.001209	0.066412
	WBGene00004990	-1.89191	10.95251	0.001223	0.066439
	WBGene00000971	-2.16115	5.000579	0.00129	0.069294
	WBGene00018221	7.069115	1.734669	0.001309	0.069591
	WBGene00017123	-1.72638	9.333115	0.001354	0.071203
	WBGene00019744	-2.66987	6.413298	0.00137	0.071311
	WBGene00001768	-2.8207	3.609545	0.001391	0.071311
	WBGene00020099	3.288374	1.90613	0.001399	0.071311
	WBGene00019215	-3.15516	2.786416	0.001433	0.072334
	WBGene00022645	2.654151	9.346673	0.00154	0.076946
	WBGene00011894	-2.37879	4.790331	0.001606	0.079434
	WBGene00006484	-6.48339	1.881029	0.001653	0.080984
	WBGene00000986	-3.02229	3.009349	0.001732	0.083385
	WBGene00019978	-1.64031	7.671011	0.001736	0.083385
	WBGene00011308	2.051165	3.7253	0.001752	0.083385
	WBGene00001086	2.040585	6.371627	0.001777	0.083703
	WBGene00019914	-1.76627	8.85832	0.001792	0.083703
	WBGene00015933	-2.23604	6.292513	0.001859	0.086026
	WBGene00010049	-2.2674	10.12707	0.001914	0.087137
	WBGene00044737	-2.0227	7.77739	0.001918	0.087137
	WBGene00000218	-1.60653	10.93889	0.001947	0.087658
	WBGene00003163	-2.63551	4.181543	0.001977	0.088185
	WBGene00004955	2.316706	3.733873	0.002063	0.091245
	WBGene00011507	2.405295	4.320019	0.002236	0.097145
	WBGene00019208	-2.45515	4.993372	0.002241	0.097145
	WBGene00001352	2.805781	3.691516	0.002255	0.097145
	WBGene00021002	-6.21417	3.125504	0.00231	0.097931
	WBGene00002269	-1.56846	12.33177	0.002328	0.097931
	WBGene00012289	5.773944	1.222546	0.002334	0.097931
	WBGene00195085	2.475465	3.639572	0.002352	0.097931
	WBGene00004947	2.417182	3.341912	0.002387	0.098276
	WBGene00000214	-1.48276	12.96962	0.002411	0.098276
	WBGene00001393	-2.03742	9.210412	0.002419	0.098276
	WBGene00021613	2.649917	3.034635	0.002468	0.098556
	WBGene00021261	5.863263	1.380123	0.002497	0.098556
	WBGene00020812	-1.70921	6.209352	0.002501	0.098556
	WBGene00012194	-1.98474	4.046088	0.002505	0.098556

	WBGene00010425	1.722023	7.175216	0.00255	0.098802
	WBGene00010924	-2.02626	3.812276	0.002551	0.098802
	WBGene00010782	-5.05697	1.941946	0.002592	0.099422
	WBGene00021292	1.9962	3.65001	0.002606	0.099422
MS1548 vs MS1809	ID	logFC	logCPM	PValue	FDR
	WBGene00015344	5.476789	3.809509	2.17E-09	1.08E-05
	WBGene00010440	-2.99994	9.618826	7.42E-08	0.000184
	WBGene00001725	-6.77825	4.170203	1.11E-07	0.000184
	WBGene00008386	3.514413	3.249822	3.05E-07	0.000381
	WBGene00011190	-4.32402	6.391242	6.83E-07	0.000683
	WBGene00011844	7.297151	1.795498	9.43E-07	0.000784
	WBGene00006418	-7.94517	4.041526	1.1E-06	0.000784
	WBGene00013650	-2.61528	6.36957	2.29E-06	0.001429
	WBGene00012866	-3.28842	4.540106	2.73E-06	0.001515
	WBGene00003473	-4.98126	3.916828	3.61E-06	0.001804
	WBGene00001029	2.985294	3.9615	4.42E-06	0.002007
	WBGene00020909	-3.15686	6.379305	5.21E-06	0.002169
	WBGene00021613	3.650886	3.034635	6.8E-06	0.002613
	WBGene00008599	5.547152	1.856353	1.48E-05	0.004916
	WBGene00001777	-3.79445	3.624625	1.49E-05	0.004916
	WBGene00004167	2.412547	4.870553	1.72E-05	0.004916
	WBGene00022073	6.612894	1.491445	1.72E-05	0.004916
	WBGene00010279	2.492419	4.861259	1.97E-05	0.004916
	WBGene00012885	-2.58372	8.37331	2.03E-05	0.004916
	WBGene00194742	-2.52518	5.408498	2.06E-05	0.004916
	WBGene00044492	-3.9951	4.323285	2.26E-05	0.004916
	WBGene00008726	-5.26408	3.348877	2.27E-05	0.004916
	WBGene00021965	-4.06602	4.282567	2.34E-05	0.004916
	WBGene00022200	2.470913	4.239144	2.36E-05	0.004916
	WBGene00007463	2.516775	7.902543	2.7E-05	0.005394
	WBGene00013730	2.217991	4.180491	3.2E-05	0.006156
	WBGene00044045	-2.58609	5.845126	4.18E-05	0.007555
	WBGene00018898	2.883426	3.908642	4.23E-05	0.007555
	WBGene00012730	2.400129	3.690285	4.62E-05	0.007962
	WBGene00006626	2.587893	6.004925	5.05E-05	0.008406
	WBGene00001393	-2.66056	9.210412	5.73E-05	0.009144
	WBGene00017238	-2.56883	6.320762	5.86E-05	0.009144

MS1548 vs MS1809	WBGene00000156	7.215801	1.611921	6.98E-05	0.010249
	WBGene00006562	-4.77136	3.657355	6.99E-05	0.010249
	WBGene00010002	6.41189	1.403625	7.32E-05	0.010249
	WBGene00017855	2.853095	4.235345	7.46E-05	0.010249
	WBGene00010015	4.367331	1.897958	7.59E-05	0.010249
	WBGene00017498	-4.95298	6.602312	7.93E-05	0.010429
	WBGene00044737	-2.50056	7.77739	8.51E-05	0.010728
	WBGene00012469	6.14625	1.605801	8.59E-05	0.010728
	WBGene00013463	-2.12756	11.47924	9.62E-05	0.011353
	WBGene00011894	-2.84107	4.790331	9.75E-05	0.011353
	WBGene00022772	-2.9491	3.438787	9.77E-05	0.011353
	WBGene00018015	-2.77786	3.68855	0.000102	0.011567
	WBGene00001394	-2.08631	10.16063	0.000104	0.011567
	WBGene00020099	3.136171	1.90613	0.000108	0.011567
	WBGene00019893	2.507461	3.252313	0.000109	0.011567
	WBGene00016002	2.44504	4.29914	0.000114	0.011861
	WBGene00017245	2.804838	4.092588	0.000127	0.01298
	WBGene00019589	2.740336	4.541311	0.000134	0.0131
	WBGene00007070	2.200388	3.234434	0.000137	0.0131
	WBGene00021492	-6.39764	3.271604	0.000139	0.0131
	WBGene00013340	2.579471	4.234819	0.000139	0.0131
	WBGene00014215	2.093775	3.887799	0.000154	0.01422
	WBGene00019654	2.051894	5.885119	0.000157	0.014234
	WBGene00006780	2.141239	3.819125	0.00016	0.014239
	WBGene00017797	2.389064	3.883795	0.000165	0.014324
	WBGene00021351	2.97277	3.134964	0.000166	0.014324
	WBGene00011676	2.918192	3.859868	0.00017	0.014393
	WBGene00019564	-4.04763	4.771018	0.000178	0.014831
	WBGene00006039	-2.21631	4.44669	0.000187	0.015299
	WBGene00013576	-2.5065	3.846609	0.000197	0.01591
	WBGene00195093	2.194823	4.295119	0.000211	0.016575
	WBGene00017726	-2.11679	7.924107	0.000212	0.016575
	WBGene00012348	2.255705	5.310934	0.00022	0.016894
	WBGene00018036	-1.68252	8.153702	0.000223	0.016894
	WBGene00007963	-2.49194	5.266101	0.000232	0.016894
	WBGene00007352	1.871831	4.915021	0.000235	0.016894
	WBGene00007365	-2.01835	9.050518	0.000239	0.016894
	WBGene00018030	4.611938	1.326008	0.000239	0.016894
	WBGene00009824	-2.18494	6.350844	0.00024	0.016894

MS1548 vs MS1809	WBGene00012780	-2.52617	4.456938	0.000247	0.017124
	WBGene00015362	-2.43174	4.373151	0.000258	0.017673
	WBGene00000659	6.377048	1.063205	0.000264	0.017851
	WBGene00016037	4.722319	1.794157	0.000268	0.017877
	WBGene00021857	1.784741	6.466102	0.000276	0.018137
	WBGene00009140	1.967195	3.594745	0.000311	0.020174
	WBGene00001752	-2.15044	6.422625	0.000321	0.02059
	WBGene00011507	2.390474	4.320019	0.000329	0.020619
	WBGene00000535	-7.94926	2.828116	0.00033	0.020619
	WBGene00020056	7.203016	1.525449	0.000345	0.021273
	WBGene00000818	-2.19202	3.862108	0.000353	0.021454
	WBGene00194674	-6.69178	3.09734	0.000356	0.021454
	WBGene00016785	-4.83176	4.99414	0.000364	0.021554
	WBGene00009521	2.769107	3.910922	0.000367	0.021554
	WBGene00000414	2.069571	7.530743	0.000387	0.022491
	WBGene00012929	2.670953	2.851502	0.000395	0.022681
	WBGene00009012	1.672516	5.8951	0.00041	0.023071
	WBGene00005025	-6.25357	1.842942	0.000411	0.023071
	WBGene00006617	1.639362	5.755722	0.000425	0.023615
	WBGene00004760	2.167458	4.032878	0.000453	0.024437
	WBGene00009772	5.889867	1.776541	0.000455	0.024437
	WBGene00007588	2.733502	3.073184	0.000455	0.024437
	WBGene00000975	-2.56992	3.796448	0.000462	0.024558
	WBGene00011046	-2.96999	5.363409	0.000478	0.02502
	WBGene00004955	2.195863	3.733873	0.000481	0.02502
	WBGene00007513	-3.28836	2.959185	0.000491	0.0253
	WBGene00019762	2.701502	3.742009	0.0005	0.025447
	WBGene00015236	4.934396	1.772469	0.000504	0.025447
	WBGene00018953	2.308048	3.685351	0.000512	0.02557
	WBGene00006936	1.802639	3.600898	0.000517	0.02557
	WBGene00021952	2.114281	6.705356	0.000528	0.025878
	WBGene00008068	-2.4422	5.754734	0.000549	0.026635
	WBGene00017880	-2.11866	3.89037	0.000569	0.027333
	WBGene00015116	-2.06675	5.806641	0.000577	0.027464
	WBGene00018804	-2.36927	4.158413	0.000597	0.028146
	WBGene00022326	-1.8879	6.798169	0.000614	0.028684
	WBGene00009065	4.491207	1.484296	0.000626	0.028978
	WBGene00018703	2.301257	4.367757	0.000636	0.029154
	WBGene00021640	3.686428	1.589734	0.000649	0.029463

MS1548 vs MS1809	WBGene00019592	2.289693	3.814281	0.000659	0.029663
	WBGene00017443	-2.69224	3.024243	0.000717	0.031991
	WBGene00006497	1.888134	3.759106	0.000762	0.03369
	WBGene00001081	1.938566	6.406096	0.000782	0.034285
	WBGene00015507	2.43204	2.89324	0.0008	0.034495
	WBGene00002637	1.713598	4.515543	0.000803	0.034495
	WBGene00017493	-6.07883	1.35772	0.000808	0.034495
	WBGene00004411	2.488332	3.521849	0.000832	0.034967
	WBGene00009628	1.891174	4.392299	0.000833	0.034967
	WBGene00004128	1.710492	3.988055	0.000856	0.03549
	WBGene00002783	3.050287	3.547186	0.00086	0.03549
	WBGene00010425	1.651049	7.175216	0.000866	0.03549
	WBGene00001226	2.160082	6.181435	0.000891	0.036075
	WBGene00001182	5.778248	1.143092	0.000895	0.036075
	WBGene00010845	2.361829	3.146079	0.000921	0.036801
	WBGene00022597	-2.71924	4.574926	0.000974	0.038632
	WBGene00011308	1.794271	3.7253	0.001055	0.041505
	WBGene00021337	-3.86688	3.201067	0.001069	0.041718
	WBGene00001324	3.682085	2.69367	0.001079	0.04179
	WBGene00022378	1.910939	3.090308	0.001095	0.04179
	WBGene00016676	1.974472	3.074544	0.001096	0.04179
	WBGene00022104	4.056339	1.479873	0.001104	0.04179
	WBGene00021292	1.787081	3.65001	0.001123	0.042094
	WBGene00015813	1.759227	3.628578	0.001136	0.042094
	WBGene00007180	-2.24918	6.985219	0.001137	0.042094
	WBGene00045212	-1.79544	4.579004	0.00117	0.042989
	WBGene00022575	-6.03663	1.863534	0.001179	0.042989
	WBGene00001007	-1.60969	5.857158	0.001191	0.043137
	WBGene00195085	2.165471	3.639572	0.001218	0.043786
	WBGene00008105	-4.40734	1.961348	0.001301	0.046043
	WBGene00006392	2.285717	3.588645	0.001305	0.046043
	WBGene00022862	1.868558	3.294451	0.001313	0.046043
	WBGene00010013	2.18883	4.527671	0.001318	0.046043
	WBGene00004993	-1.64444	8.432059	0.001371	0.047586
	WBGene00003020	2.621408	3.450103	0.001403	0.048359
	WBGene00018710	1.744134	5.825583	0.001427	0.048851
	WBGene00044650	4.220115	1.431551	0.0015	0.050991
	WBGene00009137	2.730777	1.832123	0.001537	0.05173
	WBGene00043097	-1.54993	6.324827	0.001542	0.05173

MS1548 vs MS1809	WBGene00001149	2.063546	3.326998	0.001571	0.052335
	WBGene00004759	2.505998	3.01359	0.001584	0.052429
	WBGene00001352	2.326235	3.691516	0.001604	0.052687
	WBGene00008646	2.077141	3.423169	0.001613	0.052687
	WBGene00006595	1.781963	5.038767	0.001683	0.05332
	WBGene00000931	1.792029	3.195543	0.001697	0.05332
	WBGene00003091	-2.07632	11.8815	0.001698	0.05332
	WBGene00009224	4.692382	1.474865	0.001705	0.05332
	WBGene00016888	1.598599	5.864064	0.001714	0.05332
	WBGene00002218	4.276161	1.320936	0.001714	0.05332
	WBGene00008006	2.05727	3.63204	0.001719	0.05332
	WBGene00019287	6.090239	1.06889	0.001722	0.05332
	WBGene00009254	2.190309	3.223022	0.001729	0.05332
	WBGene00004053	1.948946	3.775697	0.001745	0.053496
	WBGene00011670	-1.76399	4.978944	0.00176	0.053626
	WBGene00007258	2.452401	1.91079	0.001805	0.054439
	WBGene00009920	1.806777	5.948716	0.001808	0.054439
	WBGene00022114	1.669419	5.316332	0.001852	0.054888
	WBGene00001817	4.10874	1.515062	0.001853	0.054888
	WBGene00015933	-2.11898	6.292513	0.001856	0.054888
	WBGene00019124	1.627142	4.993982	0.001886	0.055438
	WBGene00006876	1.910231	4.424188	0.001901	0.055551
	WBGene00020878	5.14118	1.575215	0.001932	0.056126
	WBGene00020068	2.244656	2.845719	0.001972	0.05664
	WBGene00009172	4.197671	1.18485	0.001984	0.05664
	WBGene00003600	4.459878	1.752987	0.001993	0.05664
	WBGene00020360	1.793351	5.586595	0.001995	0.05664
	WBGene00010097	1.780821	4.762994	0.002054	0.057991
	WBGene00019318	4.492212	1.182347	0.002066	0.058008
	WBGene00016913	3.477339	1.443505	0.002133	0.059265
	WBGene00045247	6.251407	1.60783	0.002135	0.059265
	WBGene00045013	-1.83785	4.668684	0.002162	0.059699
	WBGene00015694	-3.5193	1.813516	0.002174	0.059703
	WBGene00000968	-2.87159	2.83881	0.002257	0.060785
	WBGene00022042	1.393095	9.782248	0.002257	0.060785
	WBGene00022337	-2.38639	3.263507	0.002262	0.060785
	WBGene00010272	1.893381	3.571584	0.002263	0.060785
	WBGene00012809	4.455297	1.25636	0.00229	0.061182
	WBGene00021697	2.002569	3.308221	0.002333	0.061885

MS1548 vs MS1809	WBGene00022015	2.232211	2.825098	0.002345	0.061885
	WBGene00016875	-2.3498	4.393396	0.002353	0.061885
	WBGene00001086	1.694253	6.371627	0.0024	0.062664
	WBGene00001042	2.014486	3.069237	0.002408	0.062664
	WBGene00008444	1.970203	5.566001	0.002456	0.063505
	WBGene00016391	1.697468	3.594228	0.002465	0.063505
	WBGene00022279	1.880436	4.695986	0.002494	0.063917
	WBGene00021355	1.756159	4.617278	0.00258	0.065661
	WBGene00002889	2.120279	3.325552	0.002596	0.065661
	WBGene00011333	1.67633	5.535431	0.00261	0.065661
	WBGene00010475	-2.12886	4.96499	0.002615	0.065661
	WBGene00022497	-1.61998	11.94667	0.002701	0.067479
	WBGene00011639	1.520872	4.911743	0.002745	0.068249
	WBGene00001910	-2.89467	3.623369	0.002761	0.068302
	WBGene00019737	1.947082	3.188525	0.002813	0.069246
	WBGene00022736	-2.02149	2.797005	0.002833	0.069398
	WBGene00021979	-4.60712	3.914627	0.002889	0.070432
	WBGene00007053	2.166776	4.261663	0.002916	0.070725
	WBGene00020701	2.497141	2.847185	0.002934	0.070817
	WBGene00012462	-1.57568	5.689109	0.003066	0.073214
	WBGene00007886	-1.71238	4.126103	0.003074	0.073214
	WBGene00010306	2.00544	4.885833	0.003077	0.073214
	WBGene00018785	1.872857	2.928759	0.003116	0.073785
	WBGene00003077	-2.41418	3.25613	0.003234	0.076222
	WBGene00002271	-1.9166	11.28006	0.003263	0.076248
	WBGene00011248	1.634774	3.156345	0.003265	0.076248
	WBGene00000983	-2.36901	4.908285	0.003302	0.076414
	WBGene00016101	-6.51064	1.659778	0.003303	0.076414
	WBGene00017801	3.017258	1.916023	0.003353	0.077194
	WBGene00016339	3.368361	1.828307	0.003368	0.077194
	WBGene00004319	2.604071	3.09613	0.00342	0.078045
	WBGene00020673	3.20505	1.652794	0.003496	0.079398
	WBGene00017275	1.830189	3.983549	0.003575	0.080782
	WBGene00020259	1.777244	3.027714	0.003589	0.080782
	WBGene00021316	1.709852	4.726406	0.00362	0.081108
	WBGene00003392	1.854136	3.622998	0.00372	0.082987
	WBGene00006710	3.604323	1.360586	0.003768	0.083175
	WBGene00000097	-2.40492	6.11314	0.003784	0.083175
	WBGene00007703	1.751353	4.234232	0.00379	0.083175

MS1548 vs MS1809	WBGene00015547	3.31668	1.870907	0.003806	0.083175
	WBGene00000407	-1.55161	5.587532	0.003817	0.083175
	WBGene00044696	-1.46308	10.81894	0.003835	0.083175
	WBGene00013378	-1.94838	6.275668	0.003857	0.083175
	WBGene00000140	1.650383	5.78731	0.003862	0.083175
	WBGene00019215	-2.86689	2.786416	0.003989	0.085559
	WBGene00021491	-3.7829	4.118392	0.004046	0.085661
	WBGene00012835	2.022033	3.092488	0.004056	0.085661
	WBGene00007938	4.44466	1.226002	0.00406	0.085661
	WBGene00006431	5.585891	1.477465	0.00407	0.085661
	WBGene00015105	2.040609	3.291549	0.004084	0.085661
	WBGene00008034	-1.84864	5.133488	0.004097	0.085661
	WBGene00000163	1.661842	5.290843	0.004132	0.085686
	WBGene00010141	-1.9363	3.112051	0.004156	0.085686
	WBGene00020781	1.833197	5.539193	0.004177	0.085686
	WBGene00012089	2.699412	3.22634	0.004186	0.085686
	WBGene00021652	2.244791	3.813557	0.004191	0.085686
	WBGene00013267	2.395962	2.952777	0.004202	0.085686
	WBGene00011481	1.820817	3.612491	0.004218	0.085686
	WBGene00012811	4.053275	1.095495	0.004263	0.085953
	WBGene00020799	-1.78904	3.639577	0.004273	0.085953
	WBGene00016735	1.843866	3.862655	0.004283	0.085953
	WBGene00017466	5.708963	1.337013	0.004304	0.086035
	WBGene00011867	1.755347	4.189709	0.00434	0.086405
	WBGene00004754	1.254382	5.949293	0.004387	0.086991
	WBGene00019593	2.268415	2.834551	0.004448	0.087849
	WBGene00016117	2.733016	2.860472	0.004521	0.088946
	WBGene00016868	1.591283	4.196305	0.004577	0.089687
	WBGene00012748	-2.02665	3.65003	0.004623	0.090238
	WBGene00022335	4.139183	1.034831	0.004729	0.091421
	WBGene00004313	2.054481	4.088574	0.004737	0.091421
	WBGene00014140	-1.6231	6.664619	0.004739	0.091421
	WBGene00006050	-1.8228	5.555923	0.004757	0.091421
	WBGene00000172	1.687564	4.781946	0.004788	0.091666
	WBGene00018532	-1.64251	5.921511	0.004839	0.092292
	WBGene00019738	-1.58418	6.02954	0.004961	0.09426
	WBGene00015802	-1.52558	6.780664	0.005051	0.095607
	WBGene00016435	-1.26997	9.836614	0.005088	0.095775
	WBGene00022106	3.838237	1.575522	0.005098	0.095775

MS1548 vs MS1809	WBGene00017131	-1.82693	5.229899	0.00512	0.095822
	WBGene00021088	2.175632	3.445564	0.00518	0.095897
	WBGene00007097	-2.82744	3.323042	0.005183	0.095897
	WBGene00019118	1.962679	3.015351	0.005185	0.095897
	WBGene00011975	1.912503	2.972157	0.005218	0.095897
	WBGene00000366	1.866158	4.562906	0.00522	0.095897
	WBGene00012343	1.440485	4.97579	0.005308	0.09715
	WBGene00017317	2.523194	3.238503	0.005346	0.097488
	WBGene00008118	-5.45961	1.344806	0.005385	0.097852
	WBGene00018681	2.69536	1.940893	0.005422	0.098167
	WBGene00004806	1.625858	3.577582	0.005482	0.098294
	WBGene00016194	1.590343	4.84618	0.005486	0.098294
	WBGene00014226	-1.71934	5.35158	0.005496	0.098294
	WBGene00014000	-1.41267	6.870091	0.005508	0.098294
MS1810 vs MS404	ID	logFC	logCPM	PValue	FDR
	WBGene00019017	-10.461	10.71855	1.84E-38	9.22E-35
	WBGene00001581	7.56402	7.869656	1.23E-23	3.08E-20
	WBGene00003473	6.214737	3.916828	1.08E-08	1.79E-05
	WBGene00003099	8.720538	3.928255	4.48E-06	0.005593
	WBGene00017498	3.844177	6.602312	7.24E-06	0.007237
	WBGene00011979	4.550262	7.269957	9.43E-06	0.007849
	WBGene00012631	-2.89636	5.168807	1.34E-05	0.008835
	WBGene00008621	3.072237	3.930216	1.41E-05	0.008835
	WBGene00020606	-2.64376	8.126632	3.12E-05	0.017331
	WBGene00018393	-2.58358	9.095176	7.02E-05	0.035058
	WBGene00022645	-3.16895	9.346673	0.000212	0.096394
MS1810 vs MS1548	ID	logFC	logCPM	PValue	FDR
	WBGene00015344	-5.53259	3.809509	4.21E-09	2.1E-05
	WBGene00001725	7.328743	4.170203	4.77E-08	0.000119
	WBGene00003473	4.71086	3.916828	6.42E-06	0.010686
	WBGene00003099	8.049844	3.928255	2.36E-05	0.02942
MS1810 vs MS1809	ID	logFC	logCPM	PValue	FDR
	WBGene00011190	-4.51881	6.389841	5.92E-07	0.003347
	WBGene00021492	-7.39765	3.271298	2.56E-06	0.007241
	WBGene00009824	-2.8577	6.353315	7.09E-06	0.013353
	WBGene00017498	-5.63246	6.604285	1.23E-05	0.014105

MS1810 vs MS1809	WBGene00006418	-7.50678	4.042875	1.25E-05	0.014105
	WBGene00012089	5.759211	3.223983	3.07E-05	0.028949
	WBGene00019589	3.645094	4.540289	3.66E-05	0.029538
	WBGene00003099	7.895427	3.926871	4.77E-05	0.030371
	WBGene00007365	-2.37134	9.050661	4.83E-05	0.030371
	WBGene00008386	3.2977	3.24814	6.69E-05	0.03785
	WBGene00019593	4.512606	2.833813	8.08E-05	0.041544
	WBGene00021613	3.641736	3.032602	0.000112	0.052845
	WBGene00008599	7.386326	1.854507	0.000152	0.062047
	WBGene00019592	3.197721	3.812857	0.000154	0.062047
	WBGene00003571	3.647113	2.937751	0.000188	0.070728
MS1810 vs N2	ID	logFC	logCPM	PValue	FDR
	WBGene00016785	-7.11857	4.99414	1.25E-06	0.00627
	WBGene00022645	-3.28712	9.346673	6.39E-06	0.01597
	WBGene00017498	3.536313	6.602312	1.66E-05	0.027666
	WBGene00012757	-5.4266	6.678981	4.52E-05	0.047793
	WBGene00007717	-4.10764	3.376464	4.78E-05	0.047793
	WBGene00021602	-3.48845	4.013477	6.43E-05	0.053529
	WBGene00219422	-6.92283	1.913816	0.000102	0.072752
	WBGene00009257	-3.69772	6.749706	0.00012	0.074814
N2 vs MS1548	ID	logFC	logCPM	PValue	FDR
	WBGene00001725	8.261125	4.168956	3.15E-12	1.78E-08
	WBGene00011046	5.570394	5.363345	1.58E-11	4.48E-08
	WBGene00016785	8.798177	4.993369	1.75E-09	3.29E-06
	WBGene00017498	-4.24434	6.604285	2.9E-09	4.1E-06
	WBGene00015344	-5.09688	3.809031	2.15E-08	2.44E-05
	WBGene00004410	2.296868	10.71127	3.28E-08	2.71E-05
	WBGene00003473	4.497071	3.917687	3.35E-08	2.71E-05
	WBGene00004430	2.211365	13.19265	9.5E-08	6.72E-05
	WBGene00007097	4.95893	3.323113	3.21E-07	0.000202
	WBGene00003091	2.800651	11.8789	1.41E-06	0.000735
	WBGene00019525	4.541262	2.832968	1.56E-06	0.000735
	WBGene00003689	-2.39984	5.431452	1.56E-06	0.000735
	WBGene00001394	2.15142	10.16008	2.89E-06	0.001257
	WBGene00004278	1.989693	5.179548	3.83E-06	0.001464
	WBGene00003402	2.009565	4.768788	4.12E-06	0.001464
	WBGene00020285	2.217292	5.713758	4.37E-06	0.001464

N2 vs MS1548	WBGene00022644	2.50603	7.416869	4.61E-06	0.001464
	WBGene00044073	2.850076	3.630781	4.66E-06	0.001464
	WBGene00011801	-3.2242	5.988947	5.46E-06	0.001624
	WBGene00007440	6.95756	6.162316	8.14E-06	0.0023
	WBGene00001777	2.958971	3.62447	9.95E-06	0.002679
	WBGene00009499	2.324473	4.622915	1.21E-05	0.003122
	WBGene00014871	3.175833	2.966975	1.39E-05	0.003414
	WBGene00008296	3.274831	5.099252	1.77E-05	0.004163
	WBGene00045457	5.648061	4.671067	1.92E-05	0.004335
	WBGene00010745	2.791515	5.781232	2.02E-05	0.004386
	WBGene00018823	2.456388	7.962195	2.45E-05	0.005023
	WBGene00018724	2.474214	6.038452	2.49E-05	0.005023
	WBGene00016219	2.04897	6.860573	2.67E-05	0.005214
	WBGene00017493	6.203502	1.358167	2.78E-05	0.005249
	WBGene00012919	2.019599	4.494992	3.02E-05	0.005502
	WBGene00000783	2.613206	5.327956	3.57E-05	0.006307
	WBGene00012194	-1.92398	4.047324	4.65E-05	0.007964
	WBGene00011938	-1.87192	5.15043	5.39E-05	0.008716
	WBGene00009126	-2.09018	5.636232	0.000054	0.008716
	WBGene00009824	-1.8971	6.353315	6.38E-05	0.010015
	WBGene00015507	-2.70611	2.892906	6.68E-05	0.010206
	WBGene00009397	2.739811	4.745644	7.58E-05	0.011109
	WBGene00019580	2.916854	3.590508	7.66E-05	0.011109
	WBGene00013836	1.818326	4.104057	8.64E-05	0.012207
	WBGene00045416	4.057032	4.517487	8.97E-05	0.012324
	WBGene00018200	2.088206	4.983689	9.15E-05	0.012324
	WBGene00003163	-2.74934	4.183438	0.000103	0.01358
	WBGene00000245	-1.8718	4.223568	0.000109	0.013709
	WBGene00018153	1.961793	5.326716	0.000109	0.013709
	WBGene00020363	3.784823	1.753903	0.000112	0.01378
	WBGene00010124	2.687853	7.79098	0.000121	0.014251
	WBGene00009257	3.545393	6.747552	0.000121	0.014251
	WBGene00044696	1.677623	10.818	0.000129	0.014656
	WBGene00044492	2.739667	4.323067	0.000131	0.014656
	WBGene00015933	2.2086	6.290135	0.000132	0.014656
	WBGene00012583	3.009305	4.335899	0.000147	0.015934
	WBGene00007717	3.63122	3.375163	0.00015	0.015993
	WBGene00022645	2.557022	9.342585	0.000157	0.016487
	WBGene00022077	-2.44338	2.798781	0.000165	0.016935

N2 vs MS1548	WBGene00021337	3.501138	3.200473	0.00017	0.0172
	WBGene00010440	1.664473	9.618008	0.00019	0.018845
	WBGene00004993	1.630322	8.430533	0.000205	0.019981
	WBGene00020579	2.604923	3.997071	0.000232	0.022135
	WBGene00010133	-2.65609	2.720052	0.000235	0.022135
	WBGene00013464	1.809664	4.010877	0.00024	0.022277
	WBGene00010480	-3.07748	2.743447	0.00025	0.022766
	WBGene00006627	7.314697	1.965295	0.000254	0.022766
	WBGene00012885	1.824729	8.373993	0.000268	0.023708
	WBGene00010470	2.593259	5.849232	0.000288	0.02436
	WBGene00044644	3.430576	6.217343	0.000296	0.02436
	WBGene00012757	4.545969	6.674106	0.0003	0.02436
	WBGene00006920	-1.94056	5.261224	0.0003	0.02436
	WBGene00011474	6.470963	3.096075	0.000301	0.02436
	WBGene00008105	3.437903	1.96149	0.000304	0.02436
	WBGene00013923	3.075023	3.333533	0.000306	0.02436
	WBGene00017594	2.688621	3.350577	0.000311	0.02444
	WBGene00009012	-1.58361	5.895743	0.000317	0.024553
	WBGene00010924	-1.89213	3.813657	0.000329	0.02514
	WBGene00001993	-1.52467	6.072924	0.000337	0.025257
	WBGene00003099	3.402486	3.926871	0.000339	0.025257
	WBGene00003831	-1.46581	7.225304	0.000365	0.026796
	WBGene00001770	1.676095	7.509588	0.000384	0.027796
	WBGene00008565	-1.54242	5.366824	0.000388	0.027796
	WBGene00020185	-1.53261	5.714845	0.000398	0.028123
	WBGene00000522	1.883962	6.808601	0.000414	0.028903
	WBGene00019738	1.686265	6.027585	0.000465	0.032029
	WBGene00009254	-2.20172	3.22338	0.00047	0.032029
	WBGene00012469	-5.53998	1.605064	0.000498	0.033249
	WBGene00012622	3.233326	1.853444	0.0005	0.033249
	WBGene00006466	2.113283	3.872012	0.000509	0.033432
	WBGene00012959	5.145645	1.783239	0.000521	0.033874
	WBGene00219316	2.034251	5.201851	0.000605	0.038874
	WBGene00012348	-1.95529	5.310518	0.00063	0.040018
	WBGene00000190	4.637863	3.69691	0.000653	0.04076
	WBGene00017431	-1.89581	7.262332	0.00066	0.04076
	WBGene00021339	3.617883	3.177207	0.000663	0.04076
	WBGene00007811	-1.42625	6.677307	0.000701	0.042353
	WBGene00011190	2.185195	6.389841	0.00071	0.042353

N2 vs MS1548	WBGene00022329	1.821272	4.608745	0.000712	0.042353
	WBGene00019902	-2.14608	2.983252	0.000721	0.04244
	WBGene00018755	1.55267	6.792627	0.000778	0.045341
	WBGene00019218	4.227055	1.339906	0.000811	0.046804
	WBGene00016788	2.834207	4.924809	0.000892	0.050927
	WBGene00007201	1.73549	3.260003	0.000916	0.050927
	WBGene00023423	6.13192	1.886692	0.000916	0.050927
	WBGene00021602	2.774792	4.011597	0.000919	0.050927
	WBGene00044646	7.847815	3.2464	0.000935	0.051342
	WBGene00009614	6.171968	1.436308	0.000957	0.051899
	WBGene00022736	1.84288	2.796402	0.00097	0.051899
	WBGene00006431	-5.94485	1.47559	0.000976	0.051899
	WBGene00001523	-1.82214	4.035138	0.000982	0.051899
	WBGene00016756	-1.51845	4.996568	0.001	0.052341
	WBGene00017864	-2.01895	3.407061	0.001013	0.052541
	WBGene00010904	2.39183	7.224769	0.001072	0.055075
	WBGene00008915	-1.69085	3.518482	0.001203	0.061277
	WBGene00007455	1.436704	7.396531	0.001238	0.061883
	WBGene00021965	2.283001	4.281451	0.00124	0.061883
	WBGene00017634	6.246306	1.551077	0.001254	0.061883
	WBGene00019967	2.669512	3.132138	0.001259	0.061883
	WBGene00012089	3.993595	3.223983	0.00132	0.063707
	WBGene00000207	-1.28173	8.275167	0.001327	0.063707
	WBGene00009142	-2.30431	4.627902	0.00133	0.063707
	WBGene00044745	5.936228	1.26442	0.001383	0.065688
	WBGene00009125	4.165964	1.723155	0.001512	0.071248
	WBGene00011894	1.814959	4.789455	0.001527	0.071342
	WBGene00001794	-1.54002	6.862294	0.001545	0.071531
	WBGene00008034	1.713444	5.133418	0.001556	0.071531
	WBGene00008492	2.333033	4.691282	0.001583	0.071683
	WBGene00015186	-1.45701	6.7272	0.001585	0.071683
	WBGene00006608	-1.47946	5.405147	0.001598	0.071723
	WBGene00015759	2.630469	9.232832	0.001616	0.0718
	WBGene00020030	-1.79233	5.599624	0.001625	0.0718
	WBGene00016004	1.498325	5.629683	0.001718	0.075295
	WBGene00007932	1.814667	4.715698	0.001828	0.079505
	WBGene00012983	-1.41895	6.649577	0.001846	0.079656
	WBGene00014206	1.679753	4.232098	0.00189	0.080964
	WBGene00021427	-1.2598	9.869335	0.001936	0.081216

N2 vs MS1548	WBGene00194651	5.708893	1.340334	0.001939	0.081216
	WBGene00013602	3.778596	1.92936	0.001949	0.081216
	WBGene00011263	3.762506	2.989198	0.001958	0.081216
	WBGene00006565	-1.82972	3.586387	0.001978	0.081216
	WBGene00219274	2.874882	1.984211	0.001982	0.081216
	WBGene00044737	1.60829	7.775503	0.002014	0.081583
	WBGene00000975	1.760995	3.796359	0.002021	0.081583
	WBGene00001007	1.297799	5.856888	0.002061	0.081583
	WBGene00020511	-1.51062	3.99024	0.002071	0.081583
	WBGene00011676	-2.25389	3.85865	0.002086	0.081583
	WBGene00219422	5.73476	1.913517	0.002087	0.081583
	WBGene00006615	1.725524	3.666867	0.002093	0.081583
	WBGene00012399	5.498816	1.150639	0.002107	0.081583
	WBGene00009927	1.610745	4.879768	0.002196	0.084449
	WBGene00012910	2.33089	3.857962	0.002242	0.085656
	WBGene00006533	1.367077	7.897223	0.002257	0.085656
	WBGene00007833	2.192505	5.67507	0.002286	0.086185
	WBGene00016425	2.25767	4.775486	0.002334	0.087392
	WBGene00018729	2.384628	3.2414	0.002386	0.088737
	WBGene00009943	1.298894	6.30218	0.002455	0.090739
	WBGene00013037	1.286414	5.962164	0.002481	0.091088
	WBGene00020911	-2.04031	2.884736	0.002512	0.091616
	WBGene00008720	-1.93561	3.388482	0.002544	0.09221
	WBGene00018317	-2.33949	2.719227	0.00257	0.09254
	WBGene00009666	-3.13146	1.603736	0.002769	0.098228
	WBGene00044648	-1.32615	5.755858	0.002778	0.098228
	WBGene00007070	-1.67536	3.23381	0.00278	0.098228
N2 vs MS1809	ID	logFC	logCPM	PValue	FDR
	WBGene00017498	-9.22516	6.602312	3.53E-14	1.77E-10
	WBGene00009824	-4.07422	6.350844	1.54E-10	3.86E-07
	WBGene00006562	-6.42941	3.657355	5.2E-09	8.65E-06
	WBGene00010124	4.367638	7.793522	1.62E-08	2.02E-05
	WBGene00012089	6.71355	3.22634	3.9E-08	3.89E-05
	WBGene00006418	-8.59573	4.041526	5.65E-08	4.71E-05
	WBGene00017238	-3.54408	6.320762	7.77E-08	5.54E-05
	WBGene00021492	-7.84231	3.271604	9.1E-08	5.68E-05
	WBGene00020909	-3.57661	6.379305	3.34E-07	0.000185
	WBGene00012194	-2.96741	4.046088	6.94E-07	0.000347

N2 vs MS1809	WBGene00013650	-2.6838	6.36957	1.19E-06	0.00054
	WBGene00018200	2.684085	4.986562	2.18E-06	0.000908
	WBGene00011844	5.683895	1.795498	3.03E-06	0.001164
	WBGene00013340	3.255386	4.234819	3.31E-06	0.001181
	WBGene00007365	-2.59649	9.050518	4.16E-06	0.001385
	WBGene00015186	-2.53047	6.724491	6.89E-06	0.002151
	WBGene00021351	3.527183	3.134964	9.72E-06	0.002856
	WBGene00010425	2.19175	7.175216	1.21E-05	0.003349
	WBGene00001333	5.054331	1.614813	1.46E-05	0.00366
	WBGene00016037	5.845033	1.794157	1.46E-05	0.00366
	WBGene00020099	3.692104	1.90613	1.55E-05	0.003689
	WBGene00009521	3.422779	3.910922	1.86E-05	0.004168
	WBGene00008368	4.264305	1.807697	1.92E-05	0.004168
	WBGene00012757	5.696228	6.678981	2.13E-05	0.004435
	WBGene00012664	-2.14418	7.762934	2.33E-05	0.004666
	WBGene00019654	2.282516	5.885119	2.62E-05	0.005035
	WBGene00022114	2.28789	5.316332	2.79E-05	0.005161
	WBGene00007352	2.134306	4.915021	2.91E-05	0.005193
	WBGene00020579	3.104285	3.998763	3.11E-05	0.005363
	WBGene00194674	-7.39596	3.09734	3.35E-05	0.005587
	WBGene00000156	7.403212	1.611921	3.6E-05	0.005803
	WBGene00003392	2.737112	3.622998	4.07E-05	0.006348
	WBGene00009920	2.40294	5.948716	4.21E-05	0.006348
	WBGene00009057	4.596813	1.758069	4.32E-05	0.006348
	WBGene00004106	4.997889	1.611258	4.47E-05	0.00638
	WBGene00013737	-2.60379	4.813109	5.12E-05	0.007061
	WBGene00001544	5.540533	1.76	5.23E-05	0.007061
	WBGene00017861	2.0741	5.17908	6.09E-05	0.008008
	WBGene00008565	-2.10858	5.365757	6.64E-05	0.008505
	WBGene00022597	-3.23746	4.574926	7.39E-05	0.009229
	WBGene00013576	-2.63941	3.846609	7.75E-05	0.009445
	WBGene00019017	-1.85101	10.71855	8.49E-05	0.010103
	WBGene00044644	3.876441	6.220302	0.000108	0.012316
	WBGene00022645	2.772274	9.346673	0.00011	0.012316
	WBGene00007963	-2.6101	5.266101	0.000111	0.012316
	WBGene00000971	-2.49668	5.000579	0.000114	0.01238
	WBGene00010942	-2.07406	6.930079	0.000119	0.01238
	WBGene00018990	3.789805	1.843096	0.000119	0.01238
	WBGene00022644	2.209762	7.419962	0.000123	0.012513

N2 vs MS1809	WBGene00022130	-2.48035	3.953787	0.000125	0.012513
	WBGene00000659	6.563519	1.063205	0.000133	0.012703
	WBGene00000190	5.360325	3.699989	0.000135	0.012703
	WBGene00016875	-2.91004	4.393396	0.000135	0.012703
	WBGene00004410	1.667308	10.7139	0.000143	0.013047
	WBGene00006936	1.940457	3.600898	0.000144	0.013047
	WBGene00006533	1.825862	7.899793	0.000159	0.014232
	WBGene00000245	-2.17619	4.221429	0.000173	0.014959
	WBGene00008599	4.182068	1.856353	0.000174	0.014959
	WBGene00002041	2.540437	4.669716	0.000179	0.015082
	WBGene00008646	2.49886	3.423169	0.000181	0.015082
	WBGene00001226	2.441877	6.181435	0.000185	0.015174
	WBGene00004167	2.044811	4.870553	0.00019	0.015183
	WBGene00004003	2.430666	3.552034	0.000191	0.015183
	WBGene00002783	3.457315	3.547186	0.000214	0.01671
	WBGene00020056	7.390421	1.525449	0.000225	0.017313
	WBGene00015802	-2.02753	6.780664	0.000242	0.018108
	WBGene00022200	2.095864	4.239144	0.000243	0.018108
	WBGene00007703	2.264942	4.234232	0.000262	0.01927
	WBGene00045237	1.920575	6.616189	0.000266	0.019286
	WBGene00019318	6.522925	1.182347	0.000278	0.019593
	WBGene00002260	-2.88025	3.106558	0.000279	0.019593
	WBGene00010990	1.857204	4.216073	0.000282	0.019593
	WBGene00007073	3.885454	1.865318	0.000292	0.019883
	WBGene00004053	2.273565	3.775697	0.000297	0.019883
	WBGene00021640	3.965591	1.589734	0.000306	0.019883
	WBGene00004442	1.592661	11.09945	0.000308	0.019883
	WBGene00017855	2.542092	4.235345	0.000309	0.019883
	WBGene00021613	2.602303	3.034635	0.00031	0.019883
	WBGene00011308	1.960062	3.7253	0.000315	0.019951
	WBGene00020735	2.75709	3.110339	0.00032	0.019997
	WBGene00008444	2.35396	5.566001	0.000326	0.020084
	WBGene00015116	-2.1546	5.806641	0.000356	0.021644
	WBGene00044648	-1.87592	5.754166	0.000365	0.021644
	WBGene00004128	1.805501	3.988055	0.000368	0.021644
	WBGene00044063	-2.20343	4.422792	0.000368	0.021644
	WBGene00007258	3.140249	1.91079	0.00038	0.022086
	WBGene00020259	2.206221	3.027714	0.000388	0.022261
	WBGene00019660	3.279955	1.891407	0.000394	0.022379

N2 vs MS1809	WBGene00008068	-2.49842	5.754734	0.000408	0.022779
	WBGene00019525	3.634133	2.833148	0.00041	0.022779
	WBGene00018030	4.137002	1.326008	0.000424	0.023062
	WBGene00004130	4.919764	1.412755	0.000428	0.023062
	WBGene00002980	-1.9082	7.303187	0.000429	0.023062
	WBGene00009973	-1.82853	8.582508	0.000445	0.023653
	WBGene00022042	1.602062	9.782248	0.000455	0.023939
	WBGene00013901	3.017256	1.885901	0.000469	0.024403
	WBGene00007811	-1.72163	6.676018	0.000479	0.024609
	WBGene00021857	1.701941	6.466102	0.000483	0.024609
	WBGene00219410	6.921296	1.633666	0.000498	0.025134
	WBGene00006918	-1.89679	7.636419	0.000519	0.025945
	WBGene00019589	2.442469	4.541311	0.000526	0.025987
	WBGene00012583	2.935265	4.338443	0.000531	0.025987
	WBGene00021088	2.692779	3.445564	0.000542	0.025987
	WBGene00021491	-4.62489	4.118392	0.000545	0.025987
	WBGene00018794	3.783749	1.692641	0.000546	0.025987
	WBGene00000931	1.963207	3.195543	0.000568	0.02677
	WBGene00016406	-2.39785	3.656666	0.000585	0.027318
	WBGene00014215	1.859129	3.887799	0.000591	0.027346
	WBGene00022104	4.287105	1.479873	0.000625	0.02867
	WBGene00016073	2.047264	3.100324	0.000632	0.028726
	WBGene00021346	1.609295	5.082891	0.000647	0.028958
	WBGene00004447	1.485613	11.65257	0.000655	0.028958
	WBGene00004411	2.540133	3.521849	0.000655	0.028958
	WBGene00018015	-2.46078	3.68855	0.000663	0.029055
	WBGene00022531	4.454615	1.878314	0.000675	0.029302
	WBGene00019375	6.221396	1.121901	0.000682	0.029302
	WBGene00003738	2.145545	5.512652	0.000691	0.029302
	WBGene00023423	6.376885	1.887269	0.000692	0.029302
	WBGene00013680	2.556951	3.293176	0.000708	0.029405
	WBGene00018709	5.40662	1.6356	0.000712	0.029405
	WBGene00015759	2.99108	9.23603	0.000713	0.029405
	WBGene00000407	-1.82635	5.587532	0.000722	0.029405
	WBGene00015001	-2.34854	3.74259	0.000727	0.029405
	WBGene00007878	4.240183	1.407908	0.00073	0.029405
	WBGene00018823	2.061265	7.965541	0.000744	0.029738
	WBGene00001480	-1.80081	6.152715	0.000752	0.029835
	WBGene00018703	2.248841	4.367757	0.000759	0.029877

N2 vs MS1809	WBGene00002994	-1.82341	6.13373	0.000765	0.029882
	WBGene00003581	-1.74081	5.670586	0.000791	0.030622
	WBGene00003571	2.499954	2.939341	0.000825	0.031419
	WBGene00009628	1.868498	4.392299	0.000827	0.031419
	WBGene00022103	2.241956	3.370552	0.00083	0.031419
	WBGene00005394	7.015248	1.671163	0.00084	0.031554
	WBGene00019665	5.500845	1.632134	0.00086	0.032
	WBGene00006615	2.007703	3.669018	0.000865	0.032
	WBGene00006964	2.515105	3.681462	0.000885	0.032422
	WBGene00007122	-1.74087	5.80051	0.000894	0.032422
	WBGene00009952	-1.71615	8.29389	0.000898	0.032422
	WBGene00011248	1.813182	3.156345	0.000904	0.032422
	WBGene00019737	2.147975	3.188525	0.000908	0.032422
	WBGene00194742	-1.9388	5.408498	0.000923	0.032535
	WBGene00044073	2.242857	3.632234	0.000925	0.032535
	WBGene00012149	-1.8006	6.411616	0.000933	0.032588
	WBGene00018221	4.289164	1.734669	0.000965	0.033354
	WBGene00020185	-1.68163	5.714072	0.000968	0.033354
	WBGene00016913	3.791523	1.443505	0.001	0.033603
	WBGene00006626	2.0742	6.004925	0.001002	0.033603
	WBGene00020721	-2.03571	7.669917	0.001002	0.033603
	WBGene00018152	3.815174	1.965067	0.001002	0.033603
	WBGene00001352	2.452214	3.691516	0.001022	0.033974
	WBGene00003474	-1.47896	9.805017	0.001027	0.033974
	WBGene00014000	-1.66947	6.870091	0.001083	0.035618
	WBGene00008386	2.07261	3.249822	0.001102	0.036007
	WBGene00010279	1.840192	4.861259	0.001141	0.037038
	WBGene00007130	-1.78942	5.921371	0.001173	0.037709
	WBGene00019656	-1.66227	4.812944	0.001177	0.037709
	WBGene00012529	1.677205	4.481621	0.001207	0.038146
	WBGene00003099	3.235896	3.928255	0.001211	0.038146
	WBGene00021652	2.553468	3.813557	0.001214	0.038146
	WBGene00195093	1.899486	4.295119	0.001231	0.038295
	WBGene00007435	2.441155	1.937094	0.001236	0.038295
	WBGene00016391	1.76304	3.594228	0.001241	0.038295
	WBGene00015783	1.74692	3.802032	0.001259	0.038581
	WBGene00001250	1.921424	4.85874	0.001298	0.039549
	WBGene00000597	-4.89618	2.779606	0.001323	0.040071
	WBGene00009051	-1.60437	7.234073	0.001352	0.040357

N2 vs MS1809	WBGene00006752	3.969737	1.755435	0.001354	0.040357
	WBGene00020516	2.77504	3.152078	0.001357	0.040357
	WBGene00009932	4.481572	1.116178	0.001388	0.041026
	WBGene00005077	-5.85999	1.582422	0.001445	0.042395
	WBGene00020591	2.251344	2.966959	0.001452	0.042395
	WBGene00011046	2.605061	5.363409	0.001468	0.042395
	WBGene00000986	-3.01233	3.009349	0.001468	0.042395
	WBGene00010152	6.593094	1.172201	0.001524	0.043314
	WBGene00003507	-2.41342	2.720786	0.001526	0.043314
	WBGene00006780	1.765929	3.819125	0.001526	0.043314
	WBGene00022190	-2.59115	2.962341	0.001534	0.043314
	WBGene00011333	1.761602	5.535431	0.001551	0.043524
	WBGene00011156	1.465018	6.132409	0.001577	0.043524
	WBGene00011532	3.952422	1.580082	0.001578	0.043524
	WBGene00006364	3.308592	1.559891	0.00158	0.043524
	WBGene00013485	5.505079	1.062367	0.00159	0.043524
	WBGene00003084	6.000265	1.197256	0.001594	0.043524
	WBGene00011205	-2.03437	5.052869	0.001619	0.043962
	WBGene00019890	-2.11132	6.870173	0.001661	0.044853
	WBGene00021602	2.763644	4.013477	0.001671	0.044885
	WBGene00008547	4.947985	1.6409	0.001759	0.046992
	WBGene00009126	-1.87714	5.633591	0.001774	0.047161
	WBGene00018424	-2.36442	4.451599	0.001785	0.047181
	WBGene00016735	2.038331	3.862655	0.001834	0.048107
	WBGene00017066	4.437522	1.420132	0.001839	0.048107
	WBGene00007180	-2.1453	6.985219	0.00185	0.048147
	WBGene00000788	-1.33936	11.64701	0.001927	0.049708
	WBGene00010002	3.159203	1.403625	0.00193	0.049708
	WBGene00019295	-2.17544	3.913363	0.00195	0.049983
	WBGene00015743	2.24926	2.814244	0.001978	0.050425
	WBGene00010924	-1.94375	3.812276	0.001991	0.050492
	WBGene00012289	3.473339	1.222546	0.002013	0.050804
	WBGene00012983	-1.63876	6.647691	0.002032	0.051036
	WBGene00003551	4.253702	1.215479	0.002053	0.051293
	WBGene00022378	1.712109	3.090308	0.002116	0.052604
	WBGene00000097	-2.56302	6.11314	0.002128	0.052635
	WBGene00011984	-1.8919	4.271667	0.002167	0.053331
	WBGene00004266	-3.75661	1.625673	0.002187	0.053372
	WBGene00020911	-2.45653	2.883468	0.002197	0.053372

N2 vs MS1809	WBGene00022247	6.045097	1.073401	0.0022	0.053372
	WBGene00006382	2.129867	3.021647	0.002233	0.053894
	WBGene00001794	-1.74446	6.86026	0.002252	0.053976
	WBGene00008274	2.035194	3.654532	0.002258	0.053976
	WBGene00001182	4.163024	1.143092	0.002305	0.054856
	WBGene00007886	-1.73063	4.126103	0.00236	0.055811
	WBGene00013033	1.96705	3.738952	0.002368	0.055811
	WBGene00001817	3.90937	1.515062	0.002399	0.056079
	WBGene00016506	-2.40581	3.416553	0.002405	0.056079
	WBGene00018036	-1.36681	8.153702	0.002413	0.056079
	WBGene00017131	-2.00442	5.229899	0.002429	0.056183
	WBGene00000984	-1.67737	4.991492	0.002449	0.056303
	WBGene00194925	-2.77114	4.852987	0.002456	0.056303
	WBGene00045457	3.856919	4.671927	0.002536	0.057711
	WBGene00018681	2.841476	1.940893	0.00255	0.057711
	WBGene00004424	1.237625	11.27141	0.002552	0.057711
	WBGene00008985	3.27707	1.525196	0.002579	0.058048
	WBGene00021839	-1.64603	7.287999	0.002634	0.058799
	WBGene00010904	2.310085	7.227389	0.002645	0.058799
	WBGene00017245	2.134527	4.092588	0.002648	0.058799
	WBGene00012532	-2.016	5.24086	0.002738	0.060548
	WBGene00008336	-2.22417	4.651972	0.002756	0.060671
	WBGene00001684	-1.39218	7.049091	0.002806	0.061272
	WBGene00013730	1.536954	4.180491	0.002813	0.061272
	WBGene00004430	1.292176	13.19478	0.00282	0.061272
	WBGene00020386	-5.6918	1.635321	0.002838	0.061402
	WBGene00011166	3.226704	2.876796	0.002867	0.061753
	WBGene00017864	-2.21129	3.40667	0.002914	0.062335
	WBGene00014018	2.945377	2.848652	0.002919	0.062335
	WBGene00012727	2.945194	1.623868	0.002954	0.062616
	WBGene00009397	2.173523	4.748076	0.002957	0.062616
	WBGene00004478	1.378251	12.17353	0.003055	0.064418
	WBGene00009830	-2.0384	6.293415	0.003079	0.064639
	WBGene00022335	4.474909	1.034831	0.003094	0.064694
	WBGene00014205	2.682397	1.846291	0.003162	0.065833
	WBGene00000207	-1.36775	8.274046	0.003211	0.066586
	WBGene00007362	-1.56294	6.964895	0.003239	0.066878
	WBGene00010071	6.213359	1.060105	0.003282	0.067497
	WBGene00017905	-1.41922	8.230575	0.003367	0.068957

N2 vs MS1809	WBGene00000647	3.49217	1.261743	0.003396	0.069144
	WBGene00012094	-1.60683	5.142972	0.003404	0.069144
	WBGene00007030	1.870029	4.336489	0.00352	0.070932
	WBGene00018609	3.040909	1.836662	0.003533	0.070932
	WBGene00009137	2.375188	1.832123	0.003537	0.070932
	WBGene00018532	-1.69631	5.921511	0.003558	0.070932
	WBGene00021363	4.506366	1.354608	0.003563	0.070932
	WBGene00012929	2.145913	2.851502	0.003635	0.072076
	WBGene00016002	1.768781	4.29914	0.003672	0.072516
	WBGene00006725	1.269101	11.30747	0.003701	0.07277
	WBGene00007877	1.788867	3.948739	0.003714	0.07277
	WBGene00006606	1.984732	2.943056	0.003789	0.073871
	WBGene00018153	1.573341	5.328982	0.003799	0.073871
	WBGene00015268	4.650227	1.708203	0.003865	0.074847
	WBGene00019405	5.955002	1.012147	0.00389	0.074847
	WBGene00008492	2.249609	4.692605	0.003894	0.074847
	WBGene00008953	-2.01495	3.227725	0.003922	0.074971
	WBGene00004408	1.169047	11.90643	0.003931	0.074971
	WBGene00004419	1.340326	9.943567	0.003968	0.075278
	WBGene00022380	-1.41823	5.343316	0.003984	0.075278
	WBGene00015236	3.59407	1.772469	0.003997	0.075278
	WBGene00015545	3.297824	1.749377	0.004007	0.075278
	WBGene00011670	-1.60839	4.978944	0.004028	0.075388
	WBGene00044045	-1.7716	5.845126	0.004073	0.075942
	WBGene00010583	-3.10336	1.832901	0.004105	0.07623
	WBGene00015094	-5.60431	1.341489	0.004119	0.07623
	WBGene00001581	-1.86331	7.869656	0.004163	0.076585
	WBGene00004476	1.284382	13.1788	0.004169	0.076585
	WBGene00010475	-2.02137	4.96499	0.004279	0.078273
	WBGene00016343	2.945063	3.26514	0.004301	0.078273
	WBGene00019208	-2.17058	4.993372	0.004308	0.078273
	WBGene00044646	6.87	3.244799	0.004323	0.078273
	WBGene00016195	-1.96205	4.079689	0.004341	0.07831
	WBGene00011430	-1.95921	3.034235	0.004421	0.079463
	WBGene00009346	1.561015	4.415682	0.004485	0.080334
	WBGene00012896	1.798211	3.137694	0.004542	0.081056
	WBGene00006516	-1.8859	3.725219	0.004639	0.0825
	WBGene00006496	3.802376	1.122083	0.004708	0.083423
	WBGene00016639	6.254326	1.306253	0.004834	0.085363

N2 vs MS1809	WBGene00008205	-1.41026	9.222674	0.004862	0.08554
	WBGene00020507	-2.08481	3.535876	0.004903	0.085763
	WBGene00019287	4.465289	1.06889	0.004911	0.085763
	WBGene00012412	-2.51049	3.764316	0.004926	0.085763
	WBGene00012722	-1.42659	5.46133	0.005074	0.088044
	WBGene00003638	2.554884	1.867183	0.005139	0.088643
	WBGene00022181	-1.88591	4.11845	0.005144	0.088643
	WBGene00006820	2.738401	1.599113	0.005189	0.089108
	WBGene00013594	-1.34985	7.427735	0.005208	0.089118
	WBGene00000181	-1.4456	7.97239	0.005238	0.08933
	WBGene00017431	-1.81191	7.259351	0.005293	0.089967
	WBGene00044202	6.377416	1.008869	0.005397	0.091427
	WBGene00007720	-1.61667	4.595423	0.005484	0.091564
	WBGene00000968	-2.63135	2.83881	0.00549	0.091564
	WBGene00017310	1.466865	4.465561	0.005494	0.091564
	WBGene00016167	-3.6459	1.649776	0.005504	0.091564
	WBGene00003163	-2.21738	4.181543	0.005512	0.091564
	WBGene00009065	2.861574	1.484296	0.005515	0.091564
	WBGene00022856	-1.86424	7.746657	0.005555	0.091915
	WBGene00016524	-5.53689	1.55823	0.005679	0.093589
	WBGene00017594	2.14548	3.350907	0.005694	0.093589
	WBGene00001393	-1.74961	9.210412	0.005801	0.095049
	WBGene00008709	3.405303	1.421068	0.005823	0.095096
	WBGene00001404	-1.37536	6.330786	0.005854	0.09528
	WBGene00001403	1.881984	5.133447	0.005988	0.097037
	WBGene00000051	3.490908	1.703676	0.006	0.097037
	WBGene00003695	-2.04583	4.621373	0.006021	0.097054
	WBGene00006603	-1.45006	6.453654	0.006059	0.097349
	WBGene00004989	-1.6214	8.686785	0.006108	0.097823
	WBGene00010833	-1.86372	3.6389	0.006132	0.097889
	WBGene00017138	-1.48198	4.66986	0.006166	0.098124
	WBGene00000831	-1.44769	7.823921	0.006226	0.098763
	WBGene00006504	3.077819	1.675877	0.00625	0.098835
	WBGene00012718	2.958853	1.55102	0.00628	0.098989
	WBGene00008191	4.681181	1.169439	0.006306	0.099097
	WBGene00013686	1.812727	3.931953	0.006338	0.099287
	WBGene00012866	-1.87304	4.540106	0.006398	0.099548
	WBGene00012097	1.459201	4.333587	0.006412	0.099548
	WBGene00018643	-3.69761	3.130308	0.006415	0.099548

	WBGene00001155	-1.43291	11.57571	0.006471	0.099634
	WBGene00010266	-1.26093	9.351781	0.006479	0.099634
	WBGene00000480	1.814914	3.853769	0.00648	0.099634

Appendix B. Differential gene expression analysis using DEseq2

Differential gene expression analysis between N2 and MS404 using DEseq2 on cDNA libraries generated in Chapter 4. Differential expression genes with a P adj. value of less than .05 are reported below.

Gene	WormBase ID	N2 1	N2 2	N2 3	MS404 1	MS404 2	MS404 3	logFC	padj
F57F4.4	WBGene00019017	4695	9864	8158	4	0	0	-9.964	5.11E-110
gfi-1	WBGene00001581	18	46	37	2293	2186	3966	5.9446	3.80E-61
T24B8.5	WBGene00011979	50	72	38	1424	736	933	4.0695	2.56E-20
F33H12.7	WBGene00045457	4	1	1	100	57	91	4.0659	2.94E-12
mtl-1	WBGene00003473	0	3	2	142	82	41	4.1932	1.93E-11
metr-1	WBGene00010988	856	736	722	194	183	172	-1.945	9.52E-08
nhr-284	WBGene00020606	462	450	404	74	130	60	-2.211	9.52E-08
C53A3.2	WBGene00016892	192	162	112	28	30	18	-2.359	5.54E-07
C32H11.4	WBGene00007867	15	44	26	186	114	282	2.4639	3.67E-06
haf-6	WBGene00001816	134	47	108	9	10	12	-2.635	5.07E-06
B0024.4	WBGene00007097	1	0	3	12	29	149	3.3931	5.91E-06
C49G7.7	WBGene00016785	0	0	0	24	18	36	3.4114	9.48E-06
ckb-2	WBGene00000512	396	146	779	12	64	64	-2.679	1.82E-05
gpdh-1	WBGene00009824	336	402	395	87	148	97	-1.684	5.42E-05
hpo-6	WBGene00021518	1	2	1	42	16	30	3.0416	5.46E-05
lys-2	WBGene00003091	1562	3196	1968	14032	4511	6958	2.0872	0.000432
swt-6	WBGene00011190	50	62	49	380	207	122	2.0208	0.000611
C50F7.5	WBGene00016845	208	78	40	22	14	6	-2.346	0.000796
asp-12	WBGene00017678	43	36	38	308	150	86	2.0201	0.001264
tth-1	WBGene00006649	10	10	8	126	15	66	2.4056	0.001382
gst-20	WBGene00001768	12	10	8	50	88	35	2.0875	0.001825
Y46G5A.20	WBGene00012910	16	6	8	83	28	86	2.2016	0.001839
F01D5.1	WBGene00008492	8	13	16	172	46	32	2.2781	0.002431
aldo-1	WBGene00011474	0	0	1	24	20	4	2.689	0.004246
Y57G7A.1	WBGene00021965	14	8	13	140	37	30	2.1277	0.007746
clec-47	WBGene00011668	192	68	120	48	9	6	-2.049	0.008145
F49F1.7	WBGene00018647	6	8	3	59	24	30	2.1286	0.008145
C29F3.7	WBGene00007807	76	88	54	258	138	253	1.4841	0.012615

F13D12.3	WBGene00008739	12	3	4	140	8	26	2.3084	0.012615
clec-4	WBGene00012583	7	6	6	84	26	16	2.1127	0.012615
valv-1	WBGene00015216	4	4	2	40	8	36	2.2453	0.012615
F44E7.2	WBGene00018424	98	134	59	26	30	5	-1.883	0.012615
clec-43	WBGene00019917	5	1	4	10	31	48	2.2202	0.014514
Y54G2A.49	WBGene00044492	22	8	7	58	66	47	1.818	0.015296
spp-8	WBGene00004993	456	354	226	806	966	754	1.334	0.01615
dod-19	WBGene00022644	122	120	42	486	294	186	1.6734	0.01615
T19D12.4	WBGene00020579	11	4	4	20	49	46	1.9978	0.017524
msra-1	WBGene00018393	562	516	744	258	202	102	-1.441	0.017728
C14C6.2	WBGene00015756	82	56	10	178	165	252	1.7558	0.017819
F49F1.1	WBGene00018643	3	16	8	97	34	24	2.0105	0.018087
M02H5.8	WBGene00019744	135	78	90	490	173	272	1.5614	0.02
clc-1	WBGene00000522	114	85	82	354	200	198	1.4063	0.020054
oac-58	WBGene00019580	4	3	10	49	18	40	1.9789	0.021108
C34H4.2	WBGene00016425	16	9	24	142	34	48	1.8743	0.021592
C17H12.8	WBGene00015933	64	63	36	381	108	106	1.7412	0.022115
dhs-20	WBGene00000983	26	44	28	265	86	53	1.8036	0.023122
clec-66	WBGene00009397	16	10	8	134	28	20	1.9546	0.031957
K11H3.5	WBGene00010782	0	5	0	20	28	12	2.1834	0.036333
T24C4.4	WBGene00020760	30	12	8	39	85	82	1.6866	0.037963
ceeh-1	WBGene00019329	20	18	20	105	52	43	1.5743	0.040498
cdr-2	WBGene00008296	24	14	14	85	46	52	1.5522	0.042186
F49E12.9	WBGene00009902	62	74	110	438	131	192	1.548	0.043969