# Endogenous Control of Stochastic Gene Expression in the Development of *Caenorhabditis elegans*

by

Ni Ji

B.A., Biology and Physics
Berea College (2007)



Submitted to the Department of Brain and Cognitive Sciences
In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013



Author .....
Department of Brain and Cognitive Sciences
March 11, 2013

Certified by .....
Alexander van Oudenaarden, Ph.D.
Professor of Biology and Professor of Physics
Thesis Supervisor

Accepted by.....
Matthew A. Wilson, Ph.D.
Sherman Fairchild Professor of Neuroscience
Director of Graduate Education for Brain and Cognitive Sciences

# Endogenous Control of Stochastic Gene Expression in the Development of *Caenorhabditis elegans*

by

Ni Ji

Submitted to the Department of Brain and Cognitive Sciences in March 11, 2013 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Neuroscience

ABSTRACT

Studies in the past decade have established gene expression as an inherently variable process. Accompanying this exciting finding is a fundamental question: how do physiological events, such as cell fate specification, proceed so robustly in the face of gene expression variability? In this thesis, I took a fresh attack at this question by examining the control of variability in the context of the stereotyped development of the nematode *C. elegans*.

Specifically, I focused on the regulation of a Hox gene by the Wnt signaling pathway in a single *C. elegans* neuroblast. Analogous to vertebrate neural crest cells, Hox gene expression determines the migratory direction and the subsequent fate choices of cells that descend from the original neuroblast. Intrigued by the earlier observation that perturbation to Wnt signaling disrupts the wild-type stereotypy in migratory decision, I speculated that variable gene expression may underlie the partial penetrance in the mutants and subsequently questioned what mechanism safeguards against variability in the wild type.

Combining single-cell transcript counting with genetic manipulation, I quantified the variability in Hox gene expression in the Q neuroblasts in both the wild type and a series of Wnt signaling mutants. Interestingly, I observed increased expression variability in a number of mutants and an overall complex relationship between expression variability and mean expression

level. Distinct features in the gene expression profile embarked me on a search for network interactions, leading to the discovery of multiple novel feedback loops within the Wnt pathway. Applying computational network inference, I revealed a network of interlocking positive and negative feedback loops, which I subsequently show to have a topological advantage in dampening stochastic noise in gene expression.

Thesis Supervisor: Alexander van Oudenaarden

Title: Professor of Physics and Biology

# ACKNOWLEDGEMENT

During my six years of study at MIT, I have received a wonderful network of support from many people. First and foremost, I am deeply grateful to my thesis advisor, Professor Alexander van Oudenaarden for his support and guidance in the past four years. I came to the lab with little exposure to systems biology. Four years later, I have gained not only quantitative skills but also ways to think about biological questions from a quantitative, systems perspective. My intellectual and personal growth would not have been possible without Alexander's guiding hand.

I must thank Professor Rik Korswagen at the Hubrecht Institute (Utrecht, The Netherlands) for his instrumental role in my thesis work. My thesis project builds critically upon our close collaboration over the past four years. I could not count how many emails we have exchanged and how many Skype discussions we have held. Even though I am not his PhD advisee, I feel like I have received as much support and guidance from him as any of his own students would have. Looking back, I feel extremely fortunate to have worked in a team with Rik as well as two of his very talented students, Teije Middelkoop and Remco Mentink. Our collaboration has taught me the great value of teamwork and open discussion—a lesson I will continue to benefit from in my future scientific career.

I wish to thank Professor Bob Horvitz as another instrumental figure in my PhD training. When I started my thesis project, work in *C. elegans* had just begun in our lab. I was eager to build a solid understanding of *C. elegans* biology, and what could be better than being right downstairs from one of the founders of *C. elegans* research! Bob was extremely generous to join my thesis committee and provide me with valuable advice at all stages of my project. For past years, Bob has also kindly accepted me as a regulator participant at his weekly group meetings.

Interacting with and learning from members of the Horvitz lab provided me with invaluable inspirations for my own research. The experience also helped me build a deep appreciation for *C. elegans* biology.

I also owe the completion of my PhD thesis to the other two members of my thesis committee, Professor Elly Nedivi and Professor Jean-Jacques Slotine. They have always been open, critical and meanwhile encouraging on my scientific progress. Elly has always encouraged me to think deeply about the biological meaning of my experimental results, and Jean-Jacques has continuously motivated me to strengthen my analytical skills and to broaden my repertoire of computational tools. Their close and dedicated mentorship has accelerated my growth as a researcher.

Many colleagues and friends have played indispensable roles in my graduate study, among them are members of the van Oudenaarden lab (2009-2013), members of the Lois lab (2008-2009), members of the Tonegawa lab (2007-2008 and summer of 2006), members of the Fee lab (2008-2009). There are many more friends to thank, at MIT, in Boston, from Berea College, as well as from back home in Nanjing, China. Finally, I would like to dedicate my thesis to my dearest parents and grandparents, as well as my beloved boyfriend. Thank you for being there for me.

# CONTENTS

# TABLE OF FIGURES

# TABLE OF TABLES

# Chapter 1

# INTRODUCTION

Even in the age of Shakespeare, the question of why individuals of the same species differ from one another has been put forward (Shakespeare, *The Tempest 4:1*). The idea of nature versus nurture has been around for perhaps just as long. With the advent of modern biology, our eyes are opened to the precise molecular substrates on which nature and nurture act on. As scientists attempt to explain and predict biological variations with greater precision, they have also uncovered along the way that there is yet a third factor to account for—that is probability, or more precisely, stochastic variation in biochemical reactions.

Over half a century ago, Novick and Weiner first reported all-or-none induction of the enzyme β-galactosidase among clonal *Escherichia coliform* cells cultured in the same environment (Novick and Weiner, 1957). In this seminal study, they noted that the cell-to-cell heterogeneity was unlikely to result from inhomogeneous inducer concentration—because the bacteria culture is continuously well-stirred. With great foresight, they speculated that some key event in the β-galactosidase synthesis pathway may occur at random upon induction, leading to full enzyme synthesis in some cells and none in other cells. Further support for the idea that cellular events can be random or stochastic came some twenty years later, when Spudich and Koshland (1976) discovered nongenetic variability in the chemotactic behavior of bacteria. After ruling out the influence of cell cycle and random environmental fluctuations, the authors postulated that the observed stochasticity might have stemmed from reactions involving low numbers of molecules. Statistically, this makes sense, as the random birth, death, and union of a small number of molecules would indeed produce large fluctuations in the downstream processes

they control. Indeed, the authors noted that the mRNA products of several chemotaxis pathway genes are present at merely tens of copies per cell. Years later, computational modeling of the chemotaxis pathway brought forth theoretical support for the original hypothesis of Spudich and Koshland (Levin et al., 1998). The next step was to obtain evidence experimentally.

At the time of Spudich and Koshland, it was technically challenging to assess the abundance of molecules in single cells. Over the past decade, technical advance have allowed direct measurement and visualization of macromolecules, such as protein and mRNA, at the single cell level (Newman et al., 2006; Yu et al., 2006; Boeger et al., 2008). A series of pioneering work have not only confirmed that gene expression is inherently stochastic, but moreover revealed cellular mechanisms that influence the degree of stochasticity. The work in this thesis builds upon these invaluable findings, and attempts to extend our understanding of stochastic gene expression to the context of animal development.

## 1.1 Origins and Consequences of Stochastic Noise in Gene Expression

*Intrinsic Origins*

To understand the origin of stochastic fluctuations in the expression of a single gene, Elowitz and colleagues (2002) proposed to classify the source of stochasticity into two categories, namely intrinsic noise and extrinsic noise. Intrinsic noise refers to stochastic fluctuations originating from the transcription and translation of the gene itself. Microscopically, the biochemical reactions leading up to transcription and translation frequently involve a small number of molecules (e.g. the number of DNA molecules is typically 2 for a eukaryotic gene). Thus, uncertainty in the thermodynamic motion of individual molecules in turn leads to uncertainty in their collective outcome, i.e. the birth and death of mRNA and protein. Theoretical

models building on these ideas have successfully captured many aspects of the observed non-genetic variability in gene expression. They have further inspired the discovery of molecular mechanisms that control the magnitude of intrinsic noise in gene expression.

Among the mechanisms discovered to date, chromatin state figures prominently in the regulation of intrinsic noise. Specific nucleotide sequences at the promoter region have been shown to encode the affinity for DNA-binding proteins such as those of the pre-initiation complex, thereby influencing the frequency and duration of transcription initiation. Nucleosomes, consisting of segments of DNA wound around the histone proteins, restrict the access of DNA-binding protein to their cognate sequence (Weinberger et al., 2005; Blake et al., 2006; Boeger et al., 2008; Sanchez et al., 2011). The occupancy of nucleosomes at the promoter region thereby sets the threshold of transcription activation. As nucleosomes are removed and reformed at a dynamic equilibrium, stochasticity is introduced to the transcriptional machinery. On a similar notion, chromatin remodeling, marked by the enzymatic addition or removal of chemical groups, is stochastic in nature and also adds noise to the transcriptional output. In addition to regulation at the chromatin level, transcriptional pausing of polymerase (Lagha et al., 2012) as well as regulation of translation rates and mRNA and protein degradation rates (Chalancon et al., 2012) have also been shown to influence the level of intrinsic noise in gene expression (Figure 1-1).

*Extrinsic Origins*

While intrinsic noise stems from the transcription and translation of a gene itself, extrinsic noise arises from the intra- or inter- cellular environment (Figure 1-1). Global fluctuation in cellular environment may lead to small but significant differences in the temperature, osmolality, or nutrient concentration experienced by individual cells, thereby resulting in difference in gene expression. The environment within the cell is also hardly constant. Proteins that make up the

transcription and translation machinery (e.g. ribosomes) are themselves subject to stochastic noise in gene expression. Fluctuations in their abundance would globally affect gene expression levels across the genome. Furthermore, as genes in a genome interact extensively in a densely connected network, fluctuations in a few regulatory genes can propagate through the network causing widespread changes in gene expression (Pedraza and Van Oudenaarden, 2005). Interestingly, emerging evidence indicates that the structure of a network may critically influence its ability to dampen, amplify or transmit noise (Hasty et al., 2000; Austin et al., 2006; Eldar and Elowitz, 2010). As discussed in detail in the section 1.2, understanding how various network architectures influence noise propagation may prompt us to understand how stochasticity in gene expression is tolerated, controlled, or exploited in nature.

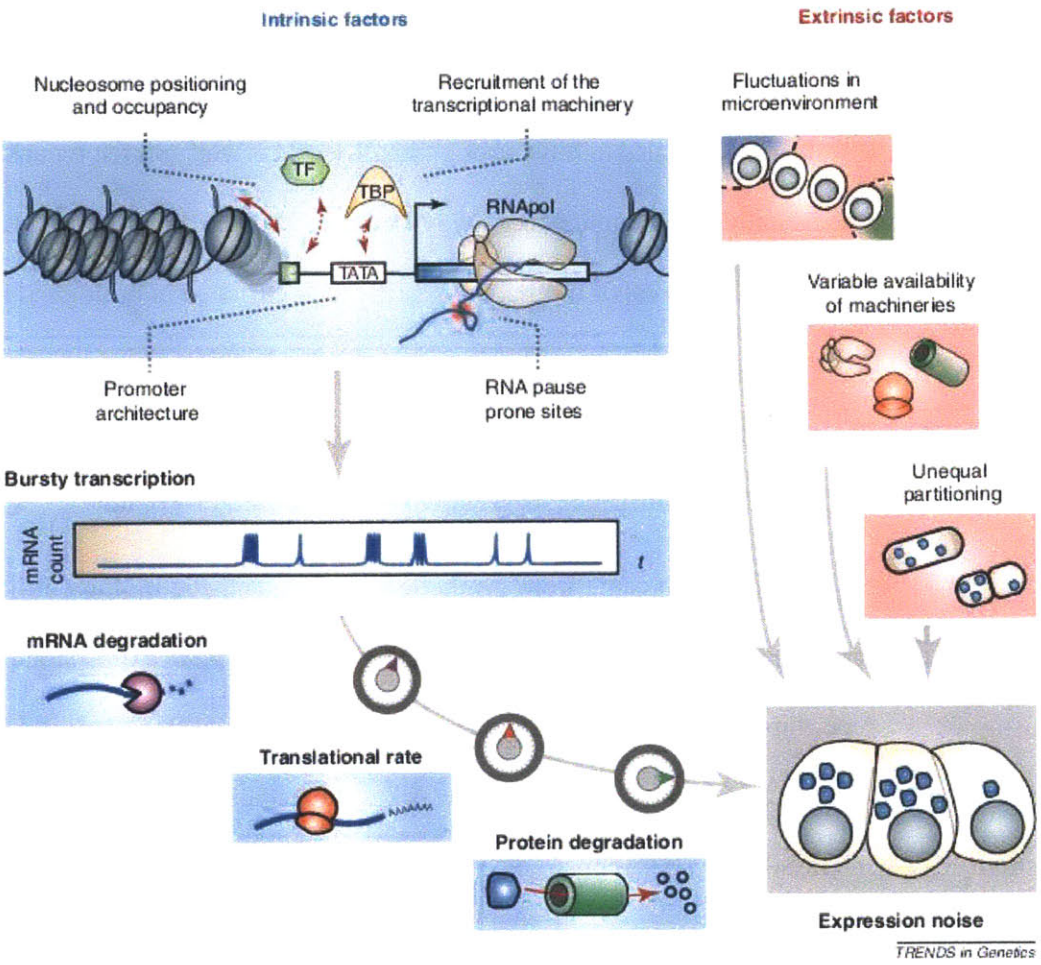



**Figure 1-1.** Intrinsic and extrinsic sources contribute to stochasticity in gene expression. Intrinsic sources of variability (left) and extrinsic fluctuations (right) together influence the level of noise in gene expression. Image courtesy of Chalancon et al., 2012.

## 1.2 Interplay between Network Architecture and Gene Expression Noise

As genes do not function in isolation but rather constitute parts of a densely interconnected network, the inherent noise in the expression of individual genes could in principle propagate through the network causing global effects on the physiological state of the organism. In parallel, studies have revealed that biological networks have distinct features from other types of

networks (e.g. social network, electrical circuit) (Barabási and Oltvai, 2004). Namely, biological networks tend to be characterized with a few highly connected nodes and many sparsely connected nodes ("scale-free"); it also exhibit strong "small-world effect", which means any two nodes in the network are only a few connections away from one another (Shen-Orr et al., 2002). Moreover, certain network structures, or motifs, are significantly enriched in biological networks (Milo et al., 2002). These structural features draw attention to the functional properties of biological networks. In the study of gene expression noise, researchers have begun to ask: do the structural properties of a network influence its sensitivity to noise? And if so, how?

The first answer to this question came from synthetic biology. Synthetic biology utilizes known open reading frames (ORFs) and regulatory sequences to build gene expression systems. Choosing from a repertoire of ORFs and regulatory sequences, one could flexibly build gene regulatory networks with a wide array of topologies. A key advantage of synthetic biology is that small scale gene networks can be constructed and manipulated in relative isolation from the rest of the genome, thus minimizing complicating factors from genes not directly studied. One of the pioneering study using synthetic gene circuit to understand expression noise was performed by Becskei and Serrano (2000). By building a synthetic construct where the protein product of a gene represses its own transcription, the authors asked how negative feedback affects noise in gene expression (Figure 1-2). Comparing to control circuit without feedback, the circuit with negative feedback exhibited less cell-to-cell variability in gene expression levels. This study demonstrated the effectiveness of synthetic biology approaches and fueled wide-spread interest in using synthetic circuits to understand biological networks.

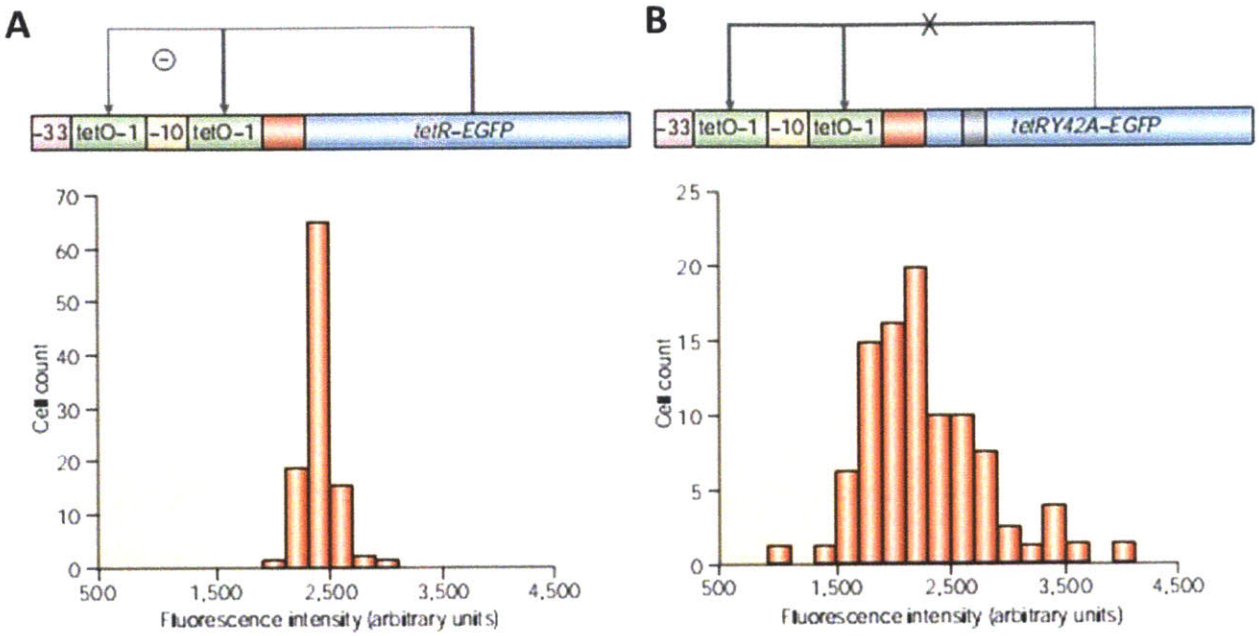



**Figure 1-2.** The use of synthetic gene circuit to probe the role of negative feedback on noise in gene expression. A) An autoregulated negative-feedback network in which a tetR-EGFP fusion protein binds to tetO operator sites in the promoter that drives its production (top panel). The levels of GFP expression in single cells show a narrow distribution of fluorescence intensity (bottom panel). B) An unregulated network constructed by mutating the DNA-binding domain of the tetR protein (top panel). Cells expressing GFP from this construct exhibits a wider distribution of GFP fluorescent intensity. Image courtesy of Hasty et al., 2001.

Following up on their initial study, Becskei and colleagues (2001) further explored the effect of positive autoregulation on stochastic gene expression. Consistent with their theoretical prediction, they found that positive feedback could convert an otherwise unimodal distribution of gene expression to a bimodal one. In other words, positive feedback could allow two distinct states of gene expression, one corresponding to high transcriptional activity and the other low levels of transcription, to coexist in the same isogenic cell population. Other researchers extended the initial studies on positive and negative feedback to provide more mechanistic understanding. For example, Austin and colleagues discovered a mapping between network

topology and the noise frequency. Importantly, they found that negative feedback shifts the noise spectra to higher frequencies, thus reducing the magnitude of noise at low frequencies. Since high frequency is often filtered out by the relative slow dynamics of transcription and translation, one would observe a net reduction in noise level as was the case in Becskei and Serrano (2000). Pedraza and van Oudenaarden (2005) elucidated how noise in gene expression propagates through a gene network. Using a combination of experimental and theoretical approaches, they ingeniously teased apart the relative contribution of intrinsic, pathway-specific, and global noise on the abundance of individual genes as well as the dynamic correlation between genes. In a complementary study, Hooshangi and colleagues (2005) also explored the propagation of noise through a cascade and found that increase in cascade length seemed to promote sharper transcriptional response to input signals.

Results from synthetic gene circuits have provided strong evidence that network topology can influence both the magnitude and the detailed distribution of gene expression noise. However, few studies exist to address whether network architecture influences gene expression noise endogenously. The recent study by (Cağatay et al., 2009) provides a positive answer to this question. The authors focused on the regulatory circuit regulating the differentiation of *Bacillus subtilis* into the competence state. By comparing the endogenous circuit to an engineered circuit with the same regulatory logic but different topological detail, the authors found that the two circuits differed in the distributions of competence durations. In this case, the endogenous circuit exhibited more stochastic variability than the engineered circuit, which appeared to endow the cell population with greater adaptability to a wide range of environment conditions. The finding of this study implicates that evolution may operate on network architecture to tune the amount of stochasticity in a genetic system. While this notion is likely to be true in eukaryotic multicellular

organisms as well, it awaits direct experimental evidence. This thesis work finds support for this notion in a developmental event of *C. elegans*. Future studies are likely to reveal more endogenous examples in which network architecture is adapted to amplify or minimize noise depending on the desired function.

## 1.3 Consequences and Control of Gene Expression Noise in Multicellular Organisms

In single-cell organisms, ample evidence has indicated the noise in gene expression could give rise to phenotypic diversity among isogenic individuals (Kaern et al., 2005; Mantzaris, 2007; Lidstrom and Konopka, 2010). Such phenotypic heterogeneity is sometimes viewed as beneficial, in that it serves as a bet-hedging strategy to ensure population survival in the face of fluctuating environments. Such reasoning may not generally apply to the control of gene expression in multicellular organisms. The complex physiology of multicellular organisms dictates that not all things can be left to chance. For example, the proper function of an organ depends on a balanced cooperation among multiple cell types; interruption of such balance often leads to disease (Komili and Silver, 2008). Furthermore, in both plants and animals, developmental events are often highly stereotyped and remarkably resistant to environmental fluctuations (Hornstein and Shomron, 2006; Geisen et al., 2008). Such phenotypic invariance suggests that noise in gene expression is either controlled or tolerated to render the physiological outcome deterministic.

As in single-cell organisms, gene expression noise is ubiquitously present in multicellular organisms and can induce profound functional consequences. In the development of the retina in *Drosophila*, stochastic expression of the gene *Spineless* gives rise to the mosaic distribution of two types of ommatidial cells across the retina (Figure 1-3, Wernet et al., 2006). The ratio of

between the two cell types, a parameter relevant to visual function, depends on the probability at which *Spineless* is turned on. In the *Caenorhabditis elegans* (*C. elegans*) embryo, cell-to-cell variability in the expression of *end-1*, the master regulator for the intestinal cell fate, is normally low but increase dramatically upon mutation of its upstream regulators. Interestingly, the highly variable *end-1* expression also appeared to correlate with the partially penetrant phenotype in the mutants. In addition to transcriptional networks, cellular response to external signals is also subject to stochastic noise (Raj et al., 2010). Upon application of an apoptosis-inducing agent, human tumor cells exhibit great variability in the timing and probability of apoptosis initiation (Spencer et al., 2009). This variability was attributed to the variable abundance of several key proteins that regulate apoptosis, implicating a role of stochastic gene expression. Of interest also is the finding that the expression of multiple house-keeping and cell-type specific genes in the murine heart and muscle appeared to become increasingly stochastic with aging. Together, these studies highlight the functional significance of gene expression noise in multicellular organisms and implicate the existence of active mechanisms to keep noise under control.

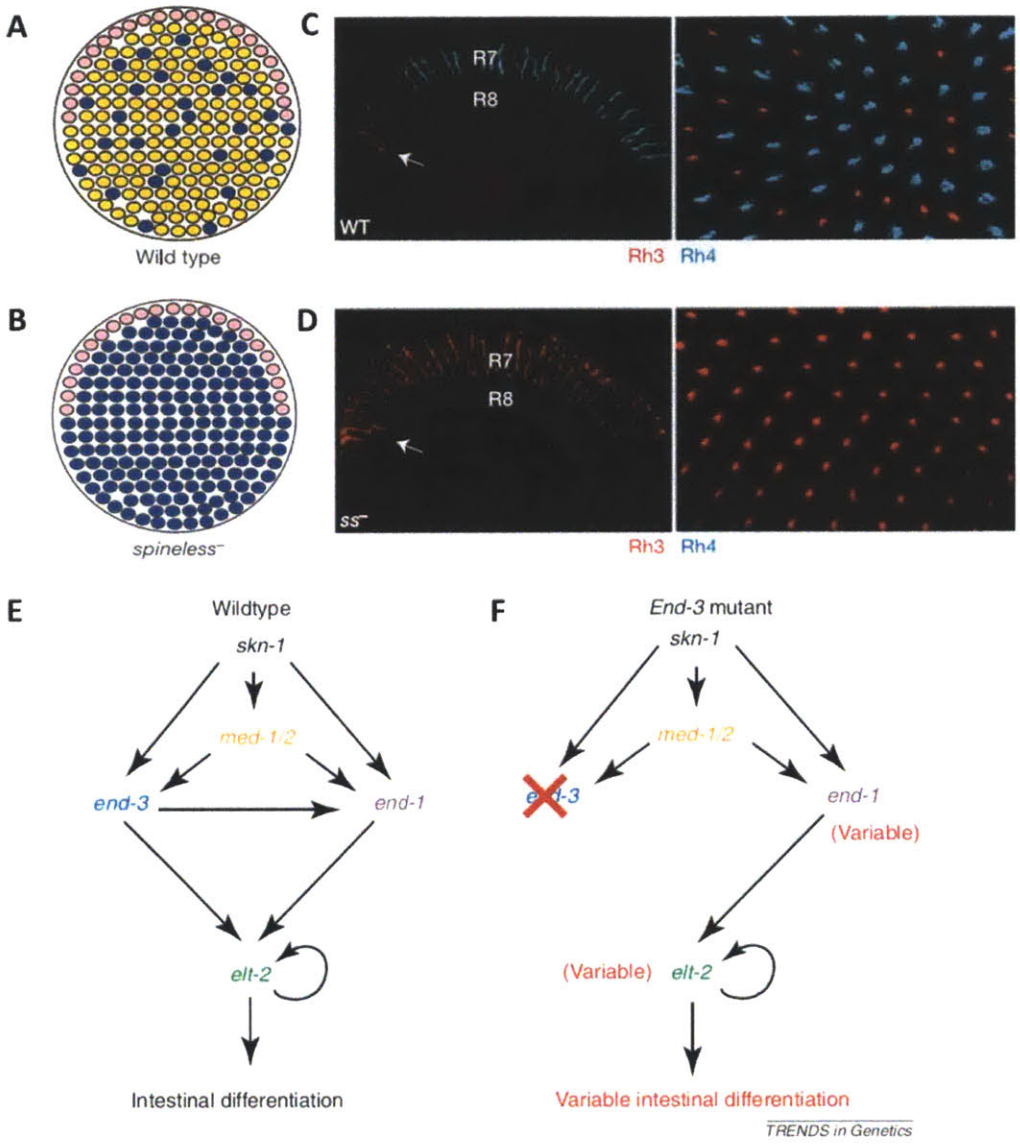



**Figure 1-3.** Examples of stochastic gene expression in endogenous systems. A)-D) Stochastic spineless expression in the *Drosophila* photoreceptor cells. A) Three subtypes of R7 ommatidia: "pale" (blue),"yellow" (yellow) and DRA (pink) together form the wild type retinal mosaic. Dorsal is up. B) Schematic representation of the spineless phenotype in R7 cells. C) Left: Transverse section through a wild type adult eye (dorsal is to the left). The arrow denotes the DRA. Right: Wild type whole-mount adult retina stained for Rh3 (red) and Rh4 (cyan) showing comparable proportions of opsins. D) Left: Transverse section through a *spineless* mutant adult eye. Rh3 (red) is expanded and Rh4 (cyan) is completely lost. Right: Opsin expression in the

mutant whole-mount adult retia. A)-D) Image courtesy of Wernet et al., 2006. E)-F) Stochastic gene expression underlies partial penetrance in *C. elegans* mutants. E) Summary of the genetic cascade governing intestinal cell specification in *C. elegans*. In the wild type, *skn-1* is maternally deposited and contributes to the activation of transcription factors *end-3* and *end-1*. *end-3* and *end-1* in turn activate elt-2, the key regulatory of intestine differentiation. F) In *end-3* mutants, *end-1* expression becomes significantly more variable, resulting in erratic expression of *elt-2* and abnormal intestine differentiation in a fraction of individuals. Image courtesy of Lagha et al., 2012.

## 1.4 Analytical Methods for Studying Noise in Gene Expression

Essential to both synthetic biology and theoretical approaches is the modeling analysis of gene expression noise. Mathematical modeling plays a crucial role in both the initial motivation as well as the final interpretation of the experiments. The mathematical tools used come from a variety of disciplines, including statistical physics, signal processing, control theory, and more (Friedman et al., 2006; Shahrezaei and Swain, 2008a). Among various approaches, the master equation provides the most accurate description of the microscopic molecular events underlying transcription and translation. From the master equation, one could in principle derive the exact distributions of mRNA and protein abundance at any point in time (as a function of the initial condition and parameter values). However, the master equation is difficult to solve analytically, especially when the network topology is complex (e.g. involving feedback) or the number of genes involved is large (Walczak et al., 2011). Such limitation has motivated the use of the Langevin and the Fokker-Planck equations, both are second-order approximations of the master equation (Walczak et al., 2011). The approximation builds on the assumption that the molecule of interest is relatively large numbers so that the system dynamics can be viewed as continuous (as opposed to discrete). These two approaches allow more complex network to be modeled

while allowing precise calculation of the gene expression distributions. The linear noise approximation (LNA) is another commonly used method (Elf and Ehrenberg, 2003). It further reduces the master equation by linearizing it around the steady state values of the system. Rather than the entire distribution, the LNA method provides estimates of the mean and variance values at steady state. The key advantage of the LNA method is that it remains analytically tractable for arbitrary network size and complexity.

## 1.5 General Overview of the Wnt Signaling Pathway

The Wnt signaling pathway is an evolutionarily conserved signal transduction cascade with wide-ranging roles in animal development and adult tissue maintenance. Disruption of normal Wnt signaling functions can lead to developmental anomalies as well as acquired diseases such as cancer, neurodegeneration, and diabetes (Katoh and Katoh, 2007; Liu et al., 2008; Schulte, 2010). The Wnt signaling pathway is typically activated through the binding of Wnt ligands to its receptors (e.g. the Frizzled, Ror, or Derailed families of receptors) and co-receptors (e.g. the Low-density lipoprotein receptor-related proteins (LRPs)) (Ling et al., 2009; van Amerongen and Nusse, 2009). The formation of the ligand-receptor complex modulates the interactions among an array of cytoplasmic signaling components, some directly affecting the cellular physiology by altering cytoskeletal dynamics or intracellular calcium flow, and others acting on transcription factors or co-factors to ultimately influence transcription. As the biochemical process downstream of ligand-receptor interaction is remarkably diverse, researchers have broadly grouped Wnt signaling into canonical and noncanonical pathways. The canonical Wnt pathway control transcription by regulating the subcellular localization of β-catenin (Korswagen, 2002; van Amerongen and Nusse, 2009, Figure 1-4). In the absence of Wnt ligands, cytoplasmic β-catenin interacts with member of the so-called "destruction complex", which include the APC

and Axin scaffold proteins and the kinases CKI and GSK3β. Interaction with the "destruction complex" leads to the phosphorylation of β-catenin and its subsequent ubiquitination and degradation by the proteosome. When Wnt ligands are bound to the Frizzled type receptors and coreceptors of the LRP-5/6/arrow family, the formation of the destruction complex is interrupted, which allows β-catenin to remain stabilized and translocate into the nucleus. Within the nucleus, the TCF/LEF transcription factors are bound at regulatory regions of Wnt-responsive genes. Without β-catenin, TCF/LEF factors interact with factors such as Groucho and histone deacetylase (HDAC) to repress transcription. Binding of TCF/LEF by β-catenin relieves the repression allowing transcription to proceed.

Compared to the canonical pathway, the noncanonical Wnt pathways utilize largely non-overlapping sets of intracellular components. At the ligand-receptor level, however, multiple Wnt ligands, Frizzled type receptors, and the disheveled family of cytoplasmic phosphoproteins have been shown to function in both canonical and noncanonical Wnt pathways (Cadigan et al., 1998; Rulifson et al., 2000; van Amerongen and Nusse, 2009). How individual pathways are activated to achieve signaling specificity is a question not yet fully understood.

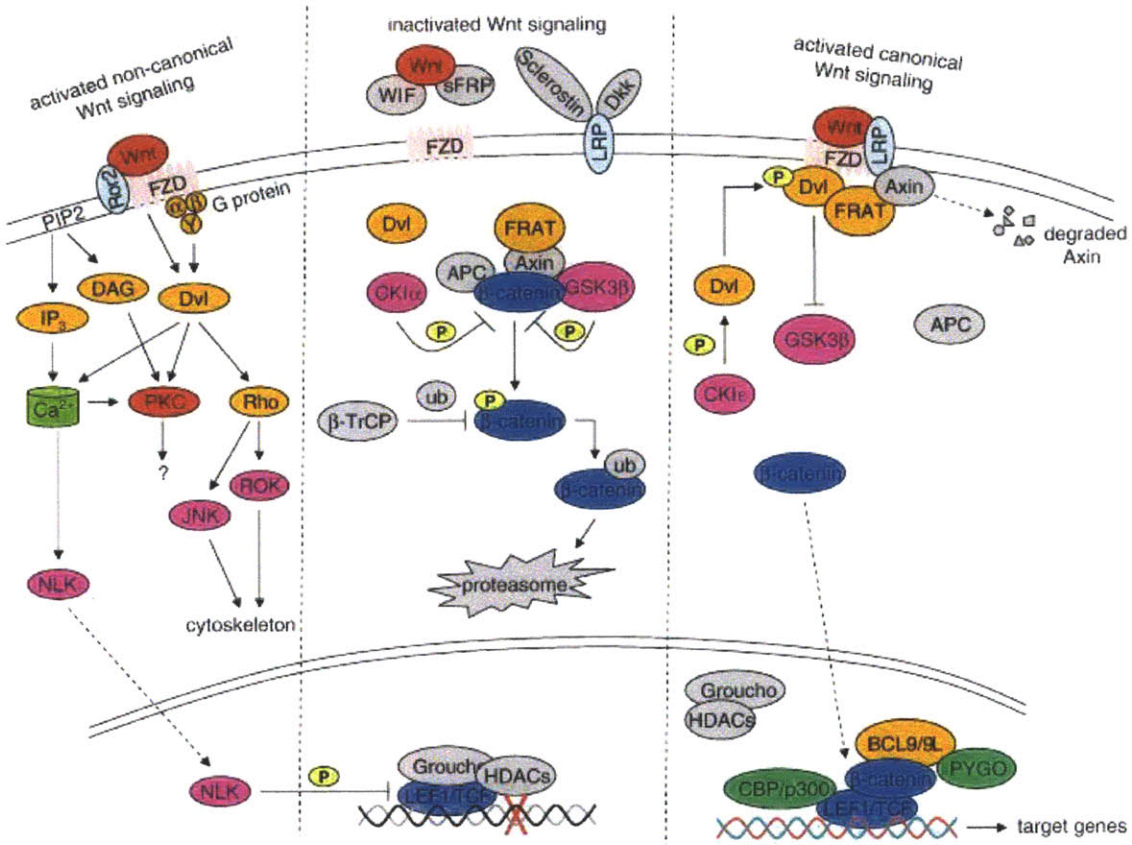



**Figure 1-4.** Schematic representation of the canonical and noncanonical Wnt signaling pathways. Shown are core components of the canonical (middle and right) and noncanonical (left) Wnt pathways. Note that noncanonical Wnt signaling encompasses other types of pathways (e.g. the Wnt asymmetry pathway in *C. elegans*) in addition to those outlined here. Image courtsey of Ling et al., 2009.

Compared to higher vertebrates, the *C. elegans* genome contains a small number of Wnt ligands and receptors. These include five Wnt ligands (*cwn-1, cwn-2, mom-2, egl-20*, and *lin-44*), four Frizzled type Wnt receptors (*mig-1, lin-17, mom-5*, and *cfz-2*), as well as the Ror/RTK type receptor *cam-1* and the Ryk/Derailed type receptor *lin-18*. As observed in other organisms, there seem to be no fixed one-to-one correspondence between Wnt ligands, receptors, and their downstream signaling cascades (Pan et al., 2006; Green et al., 2008; Zinovyeva et al., 2008). Rather, individual ligands and receptors often function in multiple Wnt pathways. Conversely,

multiple ligands or receptor often contribute redundantly to the same pathway (Pan et al., 2006; Silhankova and Korswagen, 2007; Green et al., 2008; Zinovyeva et al., 2008). In addition to the ligands and the receptors, the *C. elegans* genome also contains close homologs of many other Wnt signaling components found in higher organisms (Herman, 2003). The conserved features of the *C. elegans* Wnt pathway, together with the anatomical simplicity and genetic tractability of the organism, makes it a great model to study Wnt signaling.

## 1.6 Control of *C. elegans* Q Neuroblast Migration by Wnt Signaling

The QL and QR neuroblasts are two groups of lineally equivalent cells present at similar positions on the left and right side of the animal (Sulston and Horvitz, 1977). During the first stage of larval development, the Q mother cells, QL and QR, each generate a set of descendants that migrate in opposite directions: the QL descendants migrate a short distance towards the posterior, whereas the QR descendants migrate towards the anterior (Figure 1-5 A and B). The different migratory routes undertaken by the two Q lineages is specified by Wnt signaling (Whangbo and Kenyon, 1999). Specifically, migration towards the posterior is regulated by the Wnt ligand EGL-20, which activates a canonical Wnt pathway in QL to induce the expression of the homeobox gene *mab-5*, a key regulator of posterior migration (Salser and Kenyon, 1992). Two Frizzled receptors, MIG-1 and LIN-17, are required for *mab-5* expression to reach its wild type level (Whangbo and Kenyon, 1999, Figure 1-5 C). The regulatory mechanism underlying the anterior migration of the QR descendants is less well-characterized, but involves the joint action of multiple Wnt ligands (*egl-20, cwn-1* and *cwn-2*) and receptors (*mig-1, lin-17*, and *mom-5*) (Zinovyeva et al., 2008).

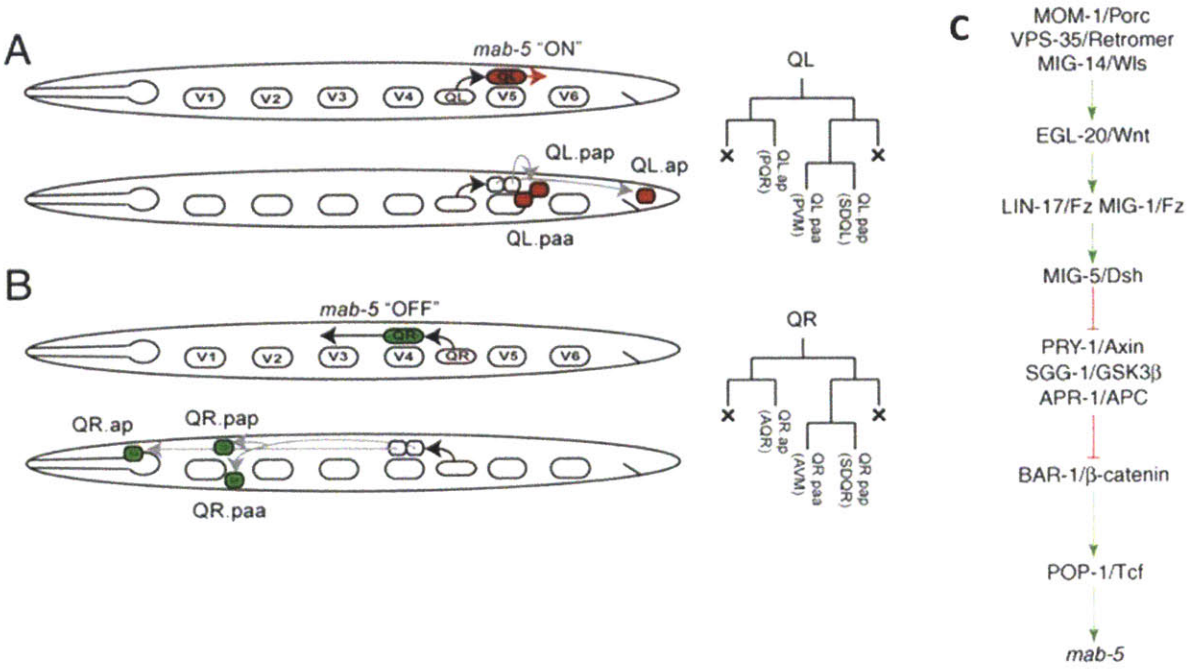



**Figure 1-5.** Control of Hox gene expression by the Wnt signaling pathway in the Q lineage of neuroblasts. A-B) Schematic overview of the Q neuroblast migration (top) and Q descendant migration (bottom). A) Left: Migratory trajectories of the left Q lineage. Right: Diagram of the left Q lineage. Each branch point corresponds to one cell division and each black cross indicates one apoptotic event. B) Left: Migratory trajectories of the right Q lineage. Right: Diagram of the right Q lineage, presented in the same format as in A. Image courtesy of Middelkoop et al., 2012. C) Components of the canonical Wnt pathway that activates *mab-5* expression in the wild type QL. Image courtesy Silhankova and Korswagen, 2007.

## 1.7 Thesis Overview

The thesis work presented here revolves around one central question: How do endogenous networks control noise in gene expression to ensure robust biological function? To answer this question, I chose to study the regulation of a Hox gene by the Wnt signaling pathway in the *C. elegans* Q neuroblasts. Analogous to vertebrate neural crest cells, the pattern of Hox gene

expression determines the migratory direction of the Q neuroblasts and the subsequent fate choices of their descendants. Intrigued by the earlier observation that perturbation to Wnt signaling disrupts the wild-type stereotypy in migratory decision, we speculated that variable gene expression may underlie the partial penetrance in the mutants and subsequently questioned what mechanism safeguards against variability in the wild type.

In Chapter 2, I focus on the control of Hox gene expression in the earliest cells of the Q lineage, QL and QR (for convenience, we refer to these two cells as the Q mother cells). Combining single-molecule fluorescent *in situ* hybridization (smFISH), high-throughput fluorescent imaging and mutant analysis, I quantified the degree of variability in the expression of a key Wnt target gene, *mab-5*, across the wild type and Wnt pathway mutants. Interestingly, Wnt pathway mutants exhibited increased cell-to-cell variability in gene expression, suggesting that a mechanism that normally controls gene expression variability was perturbed. To uncover such mechanism, I combined mutant analysis with computational network inference to uncover a novel set of interlocked positive and negative feedbacks in the Wnt pathway. I then constructed a network model to test the hypothesis that changes in network topology underlie the increased gene expression variability in the mutants. By systematic sampling the parameter space, I found that the interlocked feedback topology had a marked advantage in suppressing the stochastic variability in gene expression. I further validated this finding by analytically deriving the expected gene expression distribution under different mutant conditions and showing that the model prediction closely matched our experimental observation. Together, these findings revealed an endogenous network-based mechanism that dampens variability in gene expression. Experimental methods and materials, including a list of mutant alleles used in the study, are presented in the last section of the Chapter (section 2.4).

In Chapter 3, I extend the analysis of the Wnt pathway genes to the immediate descendants of QL and QR (from here on, we refer to these four cells as the Q daughter cells). Specifically, I ask three related questions: First, how does gene expression pattern change from the Q mother cells to the daughters? Second, whether and how does Wnt signaling and *mab-5* expression in the QL mother cell influence signaling and gene expression in the QL daughters? Third, do the genetic interactions found in the Q mother cells continue to exist in the daughter cells? The ultimate goal behind these questions is to understand the control of gene expression in the context of cell differentiation and cell cycle progression. The results reveal that lineage history and cellular context can act together to influence the transcriptional regulatory program of a given cell. Furthermore, since regulatory interactions affecting progenitor cells could exert long-lasting impact on the daughter cells, control of gene expression in the daughter cells may require the same regulatory program to switch on early in the mother cell. Experimental methods and materials, including a list of mutant alleles used in the study, are presented in the last section of the Chapter (section 3.7).

In the Appendix, I present a detailed review of the smFISH protocol optimized for fixed samples of *C. elegans* larvae. Since our approach is highly quantitative, it is crucial to ensure the sensitivity and specificity of our measuring technique. The protocol also serves as a useful resource that I hope will be of general interest to the *C. elegans* community.

## 1.8 Notes on the Genetic Nomenclature for *C. elegans*

*C. elegans* gene names typically consist of three italicized letters, a hyphen, and an Arabic number, e.g. *mab-5* or *lin-17*. The mRNA product of a gene is referred to by the gene name (e.g. *mab-5* transcript level), while the protein product must be capitalized and not italicized (e.g. MAB-5). For mutations, the name of the mutant allele usually consists of one or two letters

followed by several digits (e.g. *n671, e1787*); when used together with the name of the gene that is mutated, it is placed in a bracket after the gene name (e.g. *lin-17(n671)*). Transgenes can be described by listing the relevant components of the sequence (i.e. promoter, coding sequence, 3'UTR) interspaced by double colon, e.g. *Pegl-17::mig-1::unc-54 3'utr*. If the site of integration is known, the linkage group (I, II, III, IV, V and X corresponding to the six chromosomes) where the transgene is located can also be indicated along with the transgene. A strain carrying mutations and/or transgenes can be described to by the name of the mutant allele and/or the transgenes (i.e. *mig-1(e1787)* or *Pegl-17::gfp*). For strains carrying multiple mutations, the mutant alleles should be listed in the order of their chromosomal location, alleles residing on different chromosomes shall be separated by a semicolon (e.g. *lin-17(n671)* or *mig-1(e1787) lin-17(n671)*).

# Chapter 2

# FEEDBACK CONTROL OF GENE EXPRESSION VARIABILITY IN THE *CAENORHABDITIS ELEGANS* WNT PATHWAY

## 2.1 INTRODUCTION

Gene expression is inherently variable even among isogenic cells situated in identical environments (Raj and van Oudenaarden, 2008; Raj et al., 2008; Eldar and Elowitz, 2010; Balázsi et al., 2011; Li and Xie, 2011). On the one hand, variability in gene expression may confer beneficial phenotypic diversity. For example, it may serve as a "bet-hedging" strategy for isogenic microbial populations to ensure survival in fluctuating environments (Thattai and van Oudenaarden, 2004; Kussell and Leibler, 2005; Acar et al., 2008; Eldar et al., 2009), or as a "symmetry breaking" mechanism to induce multiple cell fates from a single progenitor cell type (Wernet et al., 2006; Chang et al., 2008; Kalmar et al., 2009). On the other hand, excessive variability in gene expression could disrupt normal development and tissue maintenance, leading to aberrant phenotypes (Aranda-Anzaldo and Dent, 2003; Chung and Levens, 2005; Henrichsen et al., 2009; Raj et al., 2010). The remarkable robustness of numerous physiological events implicates that endogenous mechanisms must exist to effectively control variability in gene expression (Nijhout, 2002; Félix and Wagner, 2008).

In a simple model of constitutive gene expression, the equilibrium level of mRNA transcripts is expected to follow a Poisson probability distribution. A distinct feature of the Poisson

distribution is that the ratio between the variance and the mean, termed the Fano factor, equals exactly one, regardless of the detailed parameters. For genes under transcriptional regulation, substantial deviations from the Poisson behavior have been theoretically proposed (Kepler and Elston, 2001; Friedman et al., 2006; Shahrezaei and Swain, 2008b) and experimentally observed in a series of seminal studies (see references below). Such deviation has often been attributed to transcriptional bursting, where the promoter transitions stochastically between its active and inactive states (Golding et al., 2005; Cai et al., 2006; Raj et al., 2006; Zenklusen et al., 2008). In addition, fluctuation in the abundance of the upstream regulators can also propagate to increase the variability of the target gene expression (Hooshangi et al., 2005; Pedraza and van Oudenaarden, 2005; Rosenfeld et al., 2005; Dunlop et al., 2008).

Pioneering theoretical and synthetic biology studies have highlighted the potential of regulatory networks in controlling gene expression variability. Negative feedback, a common mode of regulation, has been shown to suppress variability in synthetic gene expression systems (Becskei and Serrano, 2000; Austin et al., 2006). Positive feedback has been extensively studied for its ability to induce multimodal or "switch-like" behavior in both synthetic (Becskei and Serrano, 2000; To and Maheshri, 2010) and endogenous systems (Xiong and Ferrell Jr, 2003; Ozbudak et al., 2004; Acar et al., 2005; Weinberger et al., 2005; Cao et al., 2010). Whether positive feedback amplifies or dampens transcriptional variability, however, has not been established conclusively. In contrast to the simplicity of synthetic circuits, endogenous genes are embedded in densely connected networks with mixed feedback loops and multi-layered cascades (Milo et al., 2002; Davidson, 2010; Hirsch et al., 2010). Whether and how regulatory networks regulate gene expression variability endogenously remains to be explored.

*C. elegans* provides an excellent model for studying the endogenous control of gene expression variability. Its highly stereotyped development (Sulston and Horvitz, 1977) implicates underlying mechanisms that robustly control transcriptional variability. Here, we study specifically the stereotyped migratory decision of the *C. elegans* Q neuroblast. Two Q neuroblasts, QL and QR are born at bilaterally symmetrical positions in the *C. elegans* embryo, but migrate oppositely along the anterior-posterior axis upon hatching (Figure 1-5A and B and Figure 2-1A). In the left Q neuroblast (QL), expression of the Hox gene *mab-5/Antennapedia* is necessary and sufficient to ensure the posterior migration of the QL descendants ( Salser and Kenyon, 1992; Harris et al., 1996). In the wild type, *mab-5* expression in QL is dependent on the canonical Wnt signal transduced through the posteriorly produced Wnt ligand, EGL-20 (Figure 2-1A and 1B, Whangbo and Kenyon, 1999; Coudreuse et al., 2006). Two out of the four *C. elegans* Frizzled type Wnt receptors, MIG-1, and *LIN-17*, are required for *mab-5* expression in QL (Harris et al., 1996). The other Frizzled homologs, *mom-5* and *cfz-2*, have also been implicated in the regulation of QL descendant migration (Zinovyeva et al., 2008). Interestingly, Frizzled mutants exhibit varying degrees of partially penetrant migratory defects, where a fraction of QL descendants reverse to migrate anteriorly (Zinovyeva et al., 2008; Figure 2-1C and S1A). Whether this phenotypic heterogeneity originates at or downstream from *mab-5* expression is unclear.

By combining single-cell transcript counting with genetic manipulation, we identified a strong link between the variability in *mab-5* expression and the penetrance of the migratory phenotype. We observed a complex relationship between the variability and the mean levels of *mab-5* expression, implicating feedback regulation. A systematic search for regulatory interactions revealed a network of novel positive and negative feedback loops between the Frizzled receptors

and the Wnt signaling pathway. A minimal computational model could reproduce the changes in variability in the mutants and provide mechanistic insights on how the wild type network achieves robustness. Our results demonstrate, in a developmentally relevant context, the contribution of a regulatory network to controlling gene expression variability.

## 2.2 RESULTS

### Wnt signaling activates *mab-5* expression to a stable range in wild type QL

To probe the putative relation between *mab-5* expression and the phenotypic heterogeneity in the Wnt pathway mutants, it is necessary to quantitatively compare *mab-5* expression among wild type and mutants. We started by characterizing *mab-5* expression in the wild type QL neuroblasts (Figure 2-1D-E). Using single molecule fluorescent *in situ* hybridization (smFISH, Raj et al., 2008), we counted *mab-5* transcripts at various stages of QL migration (Figure 2-1D). We used the total distance migrated by QL and QR (MD), a quantity increasing monotonically with time (Figure 2-1E upper panel and Figure 2-S1B), as an indicator of migratory stage. We then combined data from many cells to obtain a population profile of *mab-5* expression dynamics (Figure 2-1E, lower panel).

Before the onset of migration, *mab-5* transcripts were present at low levels in QL (Figure 2-1D and E, MD=0-2). Thereafter, QL began to polarize and *mab-5* transcripts started to accumulate in the cytoplasm. Concurrently, nascent transcripts began to accumulate in the nucleus in the form of bright transcription sites (Figure 2-1D and S1C). The frequent appearance of paired transcription sites likely indicates heightened transcriptional activity on both alleles (Raj et al., 2006).

After a period of initial variability, *mab-5* expression converged to around 50-60 transcripts per cell towards the end of QL migration (MD≥8, Figure 2-1E). The variability in *mab-5* expression stabilized to a Fano factor value (variance/mean) of 2.4. This value is greater than the average measurement of 1.6 in *Escherichia coli* (Taniguchi et al., 2010), yet more than 10 fold lower than those reported for mammalian mRNAs (Raj et al., 2006).

While Wnt signaling has been suggested as the main activator of *mab-5* transcription (Korswagen, 2002), whether it acts directly within QL remains uncertain. We tested the cell-autonomous requirement of Wnt signaling by blocking EGL-20 mediated Wnt signaling either globally (via loss-of-function mutation of *egl-20*) or Q cell specifically (by expressing a dominant-negative form of POP-1/TCF (*DN-pop-1*) under the control of the Q cell specific promoter, *Pegl-17*) . With both manipulations, we observed more than 95% reduction in *mab-5* transcripts in QL (Figure 2-S1D). These findings confirm that Wnt signaling is required within QL to activate *mab-5* expression.

Based on the above finding, we speculated that *mab-5* expression may be used to assess Wnt pathway activity in QL. To explore this possibility, we compared the expression dynamics of *mab-5* to a transgenic reporter with seven POP-1 binding sites fused to a *pes-10* minimal promoter driving *mCherry* (POPTOP, POP-1 and TCF Optimal Promoter; Green et al., 2008). The dynamics of *mCherry* transcripts closely resembled those of *mab-5* (Figure 2-1E and 1F). This observation supports *mab-5* transcript level as an endogenous readout of Wnt signaling in QL.

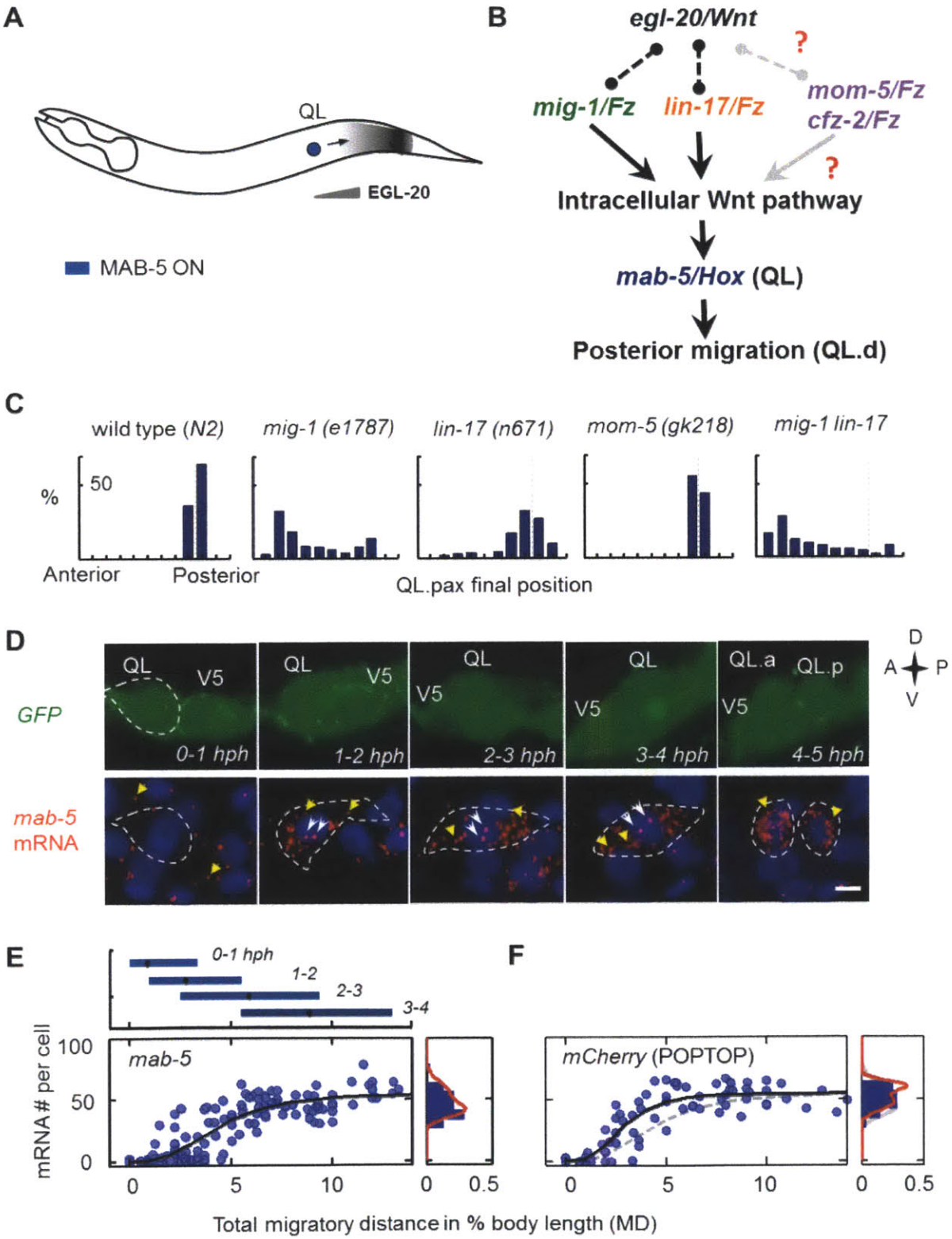

A
QL
EGL-20
MAB-5 ON
B
egl-20/Wnt
?
mig-1/Fz
lin-17/Fz
mom-5/Fz
cfz-2/Fz
?
Intracellular Wnt pathway
mab-5/Hox (QL)
Posterior migration (QL.d)
C
wild type (N2)
mig-1 (e1787)
lin-17 (n671)
mom-5 (gk218)
mig-1 lin-17
%
50
Anterior
Posterior
QL.pax final position
D
GFP
mab-5 mRNA
QL
V5
0-1 hph
1-2 hph
2-3 hph
3-4 hph
QL.a
QL.p
4-5 hph
D
A
P
V
E
0-1 hph
1-2
2-3
3-4
mab-5
mRNA # per cell
100
50
0
0
5
10
0
0.5
F
mCherry (POPTOP)
0
5
10
0
0.5
Total migratory distance in % body length (MD)

**Figure 2-1. Using single-cell transcript counting to study the control of *mab-5* expression.**

(A) Cartoon showing activation of MAB-5 expression in QL in response to the posterior-to-anterior gradient of EGL-20/Wnt.

(B) Model of Wnt signaling based on published studies. Question marks and grey edges indicate lack of definitive evidence.

(C) Final position of QL descendants in wild type and various Frizzled loss-of-function mutants. Unless otherwise noted, compound mutants carry the same alleles as single mutants.

(D) Detection of *mab-5* transcripts using smFISH over the course of QL migration. Upper row: The migrating QL neuroblast at different stages of its migration. A membrane-bound GFP marker demarcates the outline of QL and its non-migratory neighbor cell, V5. Lower row: smFISH staining of *mab-5* transcripts in the same cells as shown above. smFISH staining (red) is overlaid with DAPI staining (blue). Yellow arrowheads: single *mab-5* transcripts; white arrows: transcription centers in the nucleus of QL. While not easily discernible in a colored image, transcription centers often stain much brighter than individual transcripts in the cytoplasm. Scale bar represents 2.5 μm.

(E) *mab-5* mRNA levels in QL in wild type animals. Upper: Normalized total migratory distance (MD) for worms collected at different time points after hatching. Black dots mark the mean and blue bars span 2.5-97.5 percentiles. Lower: Number of mab-5 transcripts per cell plotted against MD. Each blue data point corresponds to the number of *mab-5* smFISH spots measured in a single QL cell from a single animal. The histogram to the right is generated using data points to the left with MD>8.

(F) *mCherry* mRNA levels in QL in animals carrying the Wnt signaling reporter transgene, POPTOP.

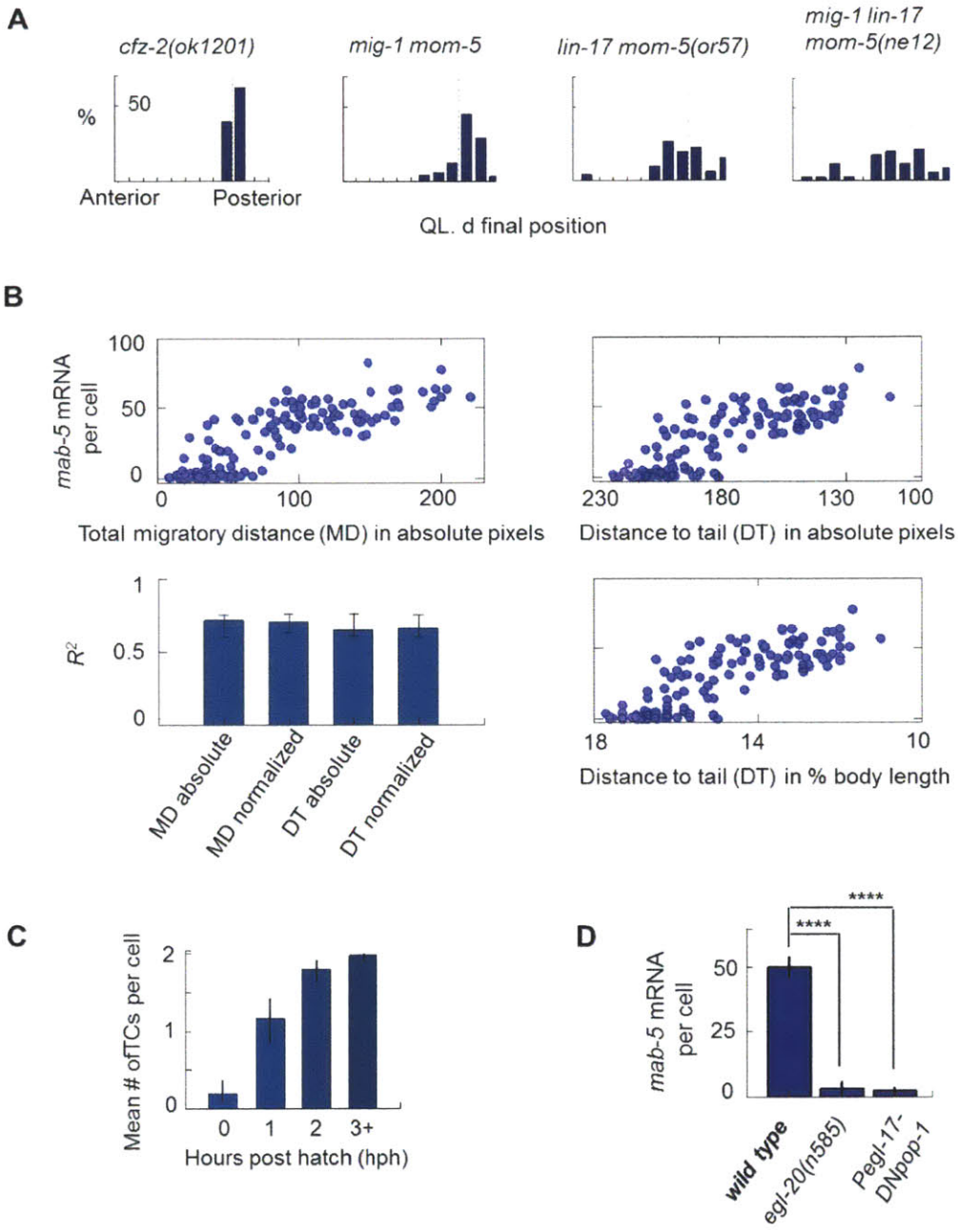



**Figure 2- S1.** A) Final position of QL descendants in additional Frizzled mutants. B) Upper left clockwise to lower right: mab-5 transcript counts plotted against other parameters associated with Q cell migration. Lower left: Squared correlation coefficients between mab-5 transcript count and various migration-related parameters. C) Mean number of transcription centers (TCs) per cell at different stages of migration. D) Final (MD>8) mab-5 transcript counts in wild type

compared to mutants with global or QL-specific blockade of EGL-20 dependent Wnt signaling. N>40 for each data set. All error bars in this figure denote 95% confidence intervals (95% CI).

**Three Frizzled receptors are expressed in QL and exhibit distinct expression dynamics**

Following our observation that different Frizzled mutants exhibit different penetrance in the migratory phenotype of the QL descendant (Figure 2-1C and S1A), we speculated that different Frizzled receptor may differ in their expression pattern. To test this hypothesis, we characterized the expression of the four *C. elegans* Frizzled paralogs. Using paralog-specific smFISH, we detected QL-specific expression of *mig-1/Fz*, *lin-17/Fz* and *mom-5/Fz*, but not *cfz-2/Fz* (Figure 2-2A and S2A). In addition to a difference in the average abundance, the Frizzled paralogs also exhibited diverse expression dynamics. *mig-1* transcripts were initially enriched (average 27 copies per cell) but decreased to less than 10 copies per cell over the course of migration. *lin-17* exhibited the opposite dynamics, rising from less than 10 copies per cell to an average of 34. *mom-5* in contrast was expressed at low levels (on average 4 copies per cell) throughout QL migration (Figure 2-S2A). Outside QL, the four Frizzleds also exhibited distinct global expression patterns (Figure 2-S2B).

The dynamics of *mig-1* and *lin-17* transcripts suggest they may exhibit single-cell level correlation with *mab-5*. While a positive correlation may be intuitively expected between receptor and target gene abundance, *mig-1* instead exhibited a significant negative correlation with *mab-5* (Figure 2-2B, Pearson's $R=-0.67$, $p<0.001$). *lin-17*, while causing a weaker migration defect when mutated (Figure 2-1C), exhibited a strong positive correlation with *mab-5* (Pearson's $R = 0.91$, $p<0.001$). No significant correlation was observed between *mom-5* and *mab-5* (Pearson's $R= 0.02$, $p>0.5$). Together, the distinct transcriptional and correlation profiles

suggest that divergent transcriptional regulatory programs exist upstream of the Frizzled receptors.

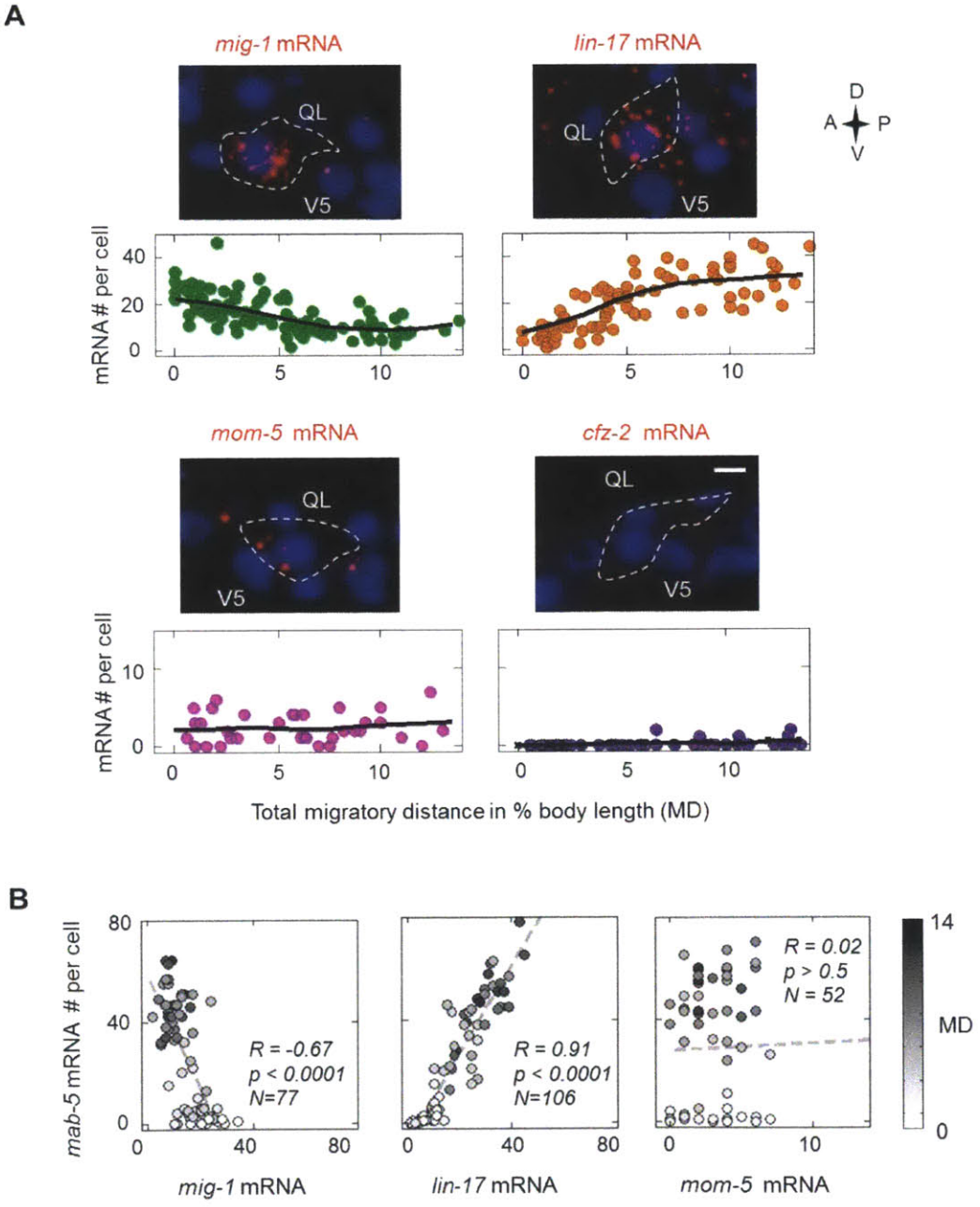



**Figure 2-2.** Three Frizzled paralogs are dynamically transcribed in QL. A) smFISH staining and single-cell transcript counts for the four *C. elegans* Frizzled paralogs over the course of QL

migration. B) Single-cell correlation between Frizzled and *mab-5* transcript counts. Shades of dots indicate corresponding MD value.

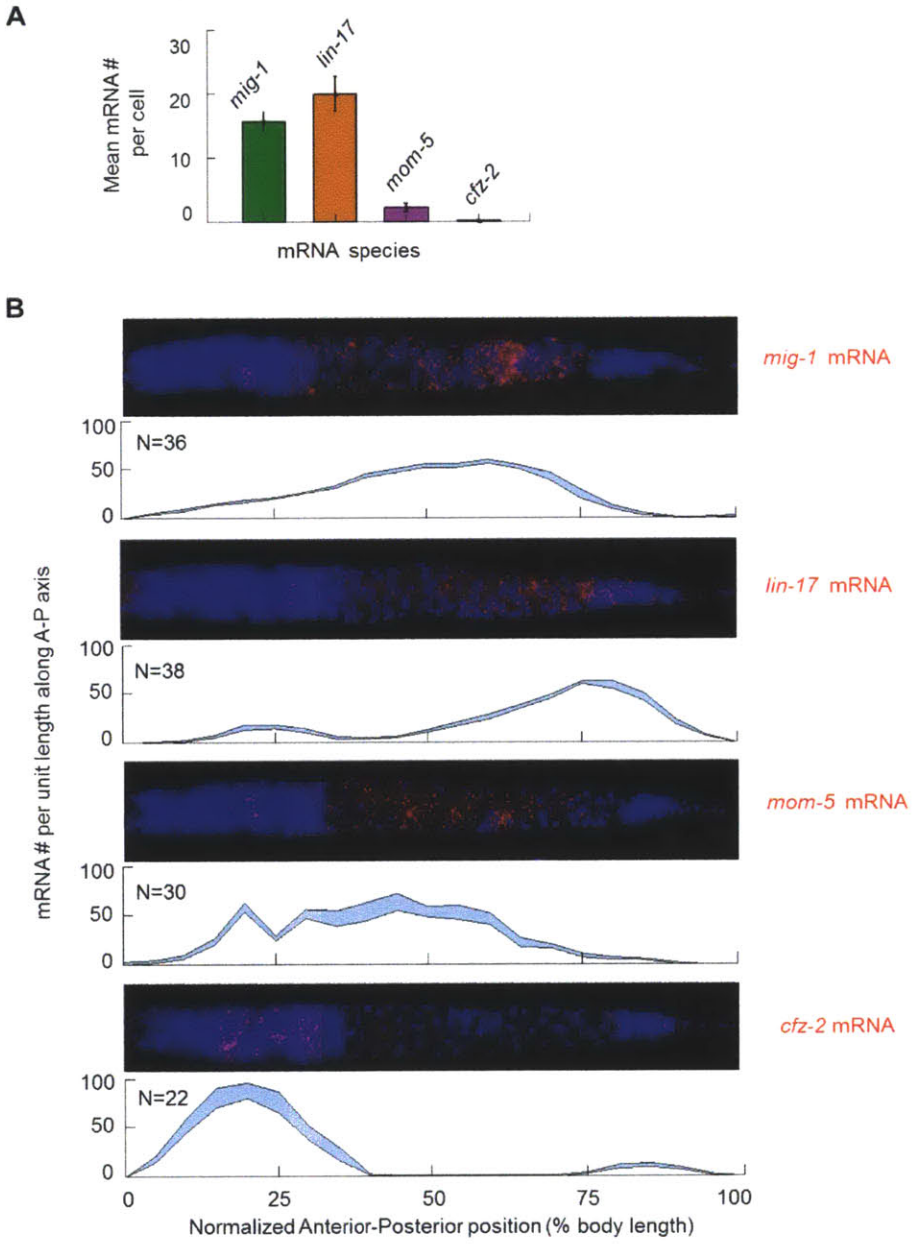



**Figure 2-S2.** A) Average transcript abundance over the course of migration for the four Frizzled paralogs. Error bars are 95% confidence intervals. N>40 for each data set. B) smFISH staining

and quantification of transcript abundance along the anterior-posterior axes for the four Frizzled genes. Error bars (light blue patches) are 95% CI.

**Frizzled mutants exhibit different levels of variability in *mab-5* expression**

Having assessed *mab-5* and Frizzled gene expression in the wild type, we next asked whether and how *mab-5* expression is altered in a series of mutants carrying single or compound Frizzled loss-of-function mutations.

MAB-5 antibody staining was previously shown to be reduced in *mig-1* and *lin-17* single mutants (Harris et al., 1996). In agreement, we observed a strong reduction in *mab-5* transcripts (less than 10 transcripts per cell) in the majority of *mig-1(e1787)* single mutant (Figure 2-3A). A small fraction of QLs, however, retained significant levels of *mab-5* expression (20 transcripts per cell or higher). Cell-to-cell heterogeneity was also evident in the *lin-17(n671)* single mutant. Individual QLs exhibited between very low to near wild type amount of *mab-5* transcripts. The *mom-5(gk812)* mutant, unlike the wild type, exhibited high variability in *mab-5* expression beyond the initial phase of QL migration (MD>5, note the existence of cells with < 25 copies of *mab-5*). In comparison, *mab-5* levels in the *cfz-2(ok1201)* mutant were indistinguishable from the wild type.

While mutations in more than one of the three Frizzleds (*mig-1, lin-17* or *mom-5*) nearly completely abolished *mab-5* expression in QL, heterozygotes of these mutants exhibited near wild type levels of *mab-5*. Interestingly, heterozygotes of the Frizzled triple mutant (triple het) exhibited increased cell-to-cell variability in *mab-5* expression. While the majority of QLs exhibited wild type levels of *mab-5*, a small fraction of the late stage QLs contained less than 20 *mab-5* transcripts (Figure 2-3A). Thus, heterozygous mutations in the three Frizzleds appeared to

have perturbed Wnt signaling to a level sufficient to reduce *mab-5* expression in some cells. Taken together, increased variability appeared to accompany the decrease in *mab-5* expression across the Frizzled mutants.

Motivated by the recent discovery that variability in gene expression underlies partial penetrance (Raj et al., 2010), we questioned whether variability in *mab-5* transcript abundance could predict the phenotypic penetrance of different mutants. We hypothesized that *mab-5* expression must exceed a certain threshold level; otherwise the QL descendants would migrate anteriorly (i.e. become defective in posterior migration). Under this hypothesis, we searched (Figure 2-S3C) and found threshold values of around 25 transcripts per cell (Figure 2-3C) to yield accurate predictions of phenotypic penetrance. *mom-5* single and compound mutants were not included in this analysis due to the *mab-5*-independent requirement of *mom-5* for anterior migration (Zinovyeva et al., 2008). Thus, upregulating *mab-5* expression above a certain threshold may be critically in driving a robust migratory decision of QL descendants.

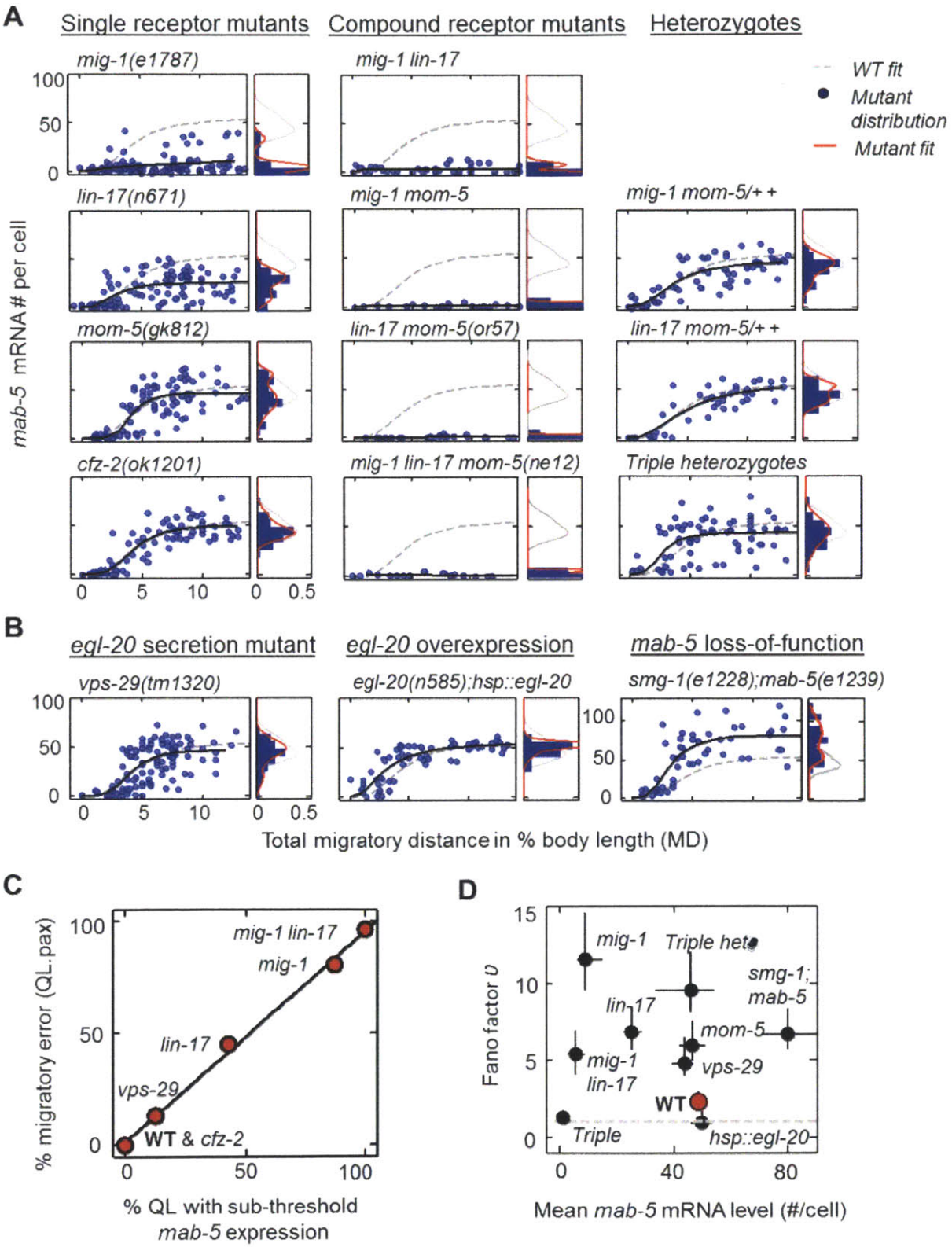

A
Single receptor mutants
Compound receptor mutants
Heterozygotes
mig-1(e1787)
mig-1 lin-17
lin-17(n671)
mig-1 mom-5
mig-1 mom-5/+ +
mom-5(gk812)
lin-17 mom-5(or57)
lin-17 mom-5/+ +
cfz-2(ok1201)
mig-1 lin-17 mom-5(ne12)
Triple heterozygotes
WT fit
Mutant distribution
Mutant fit
mab-5 mRNA # per cell
B
egl-20 secretion mutant
egl-20 overexpression
mab-5 loss-of-function
vps-29(tm1320)
egl-20(n585);hsp::egl-20
smg-1(e1228);mab-5(e1239)
Total migratory distance in % body length (MD)
C
% migratory error (QL.pax)
mig-1 lin-17
mig-1
lin-17
vps-29
WT & cfz-2
% QL with sub-threshold mab-5 expression
D
Fano factor υ
mig-1
Triple het
smg-1; mab-5
lin-17
mom-5
mig-1 lin-17
vps-29
WT
Triple
hsp::egl-20
Mean mab-5 mRNA level (#/cell)

**Figure 2-3.** Wnt signaling mutants exhibit different variability in *mab-5* expression. A) Dynamic and steady state *mab-5* expression in Frizzled single and compound mutants. Black lines are guide to the eye generated by fitting to a sigmoidal function. Red curves are fits using two Gaussian distributions. B) Dynamic and steady state *mab-5* expression in mutants with altered EGL-20/Wnt gradient or loss of MAB-5 function. C) Correlation between *mab-5* transcript levels and the migratory phenotype of the QL descendants in various Wnt pathway mutants. Same mutant alleles as listed in (A) and (B). D) Fano factor versus the steady state mean of *mab-5* in various Wnt pathway mutants. Red indicates wild type. Grey broken line indicates expected Fano factor value for Poisson distribution. Error bars are 95% confidence intervals (CI).

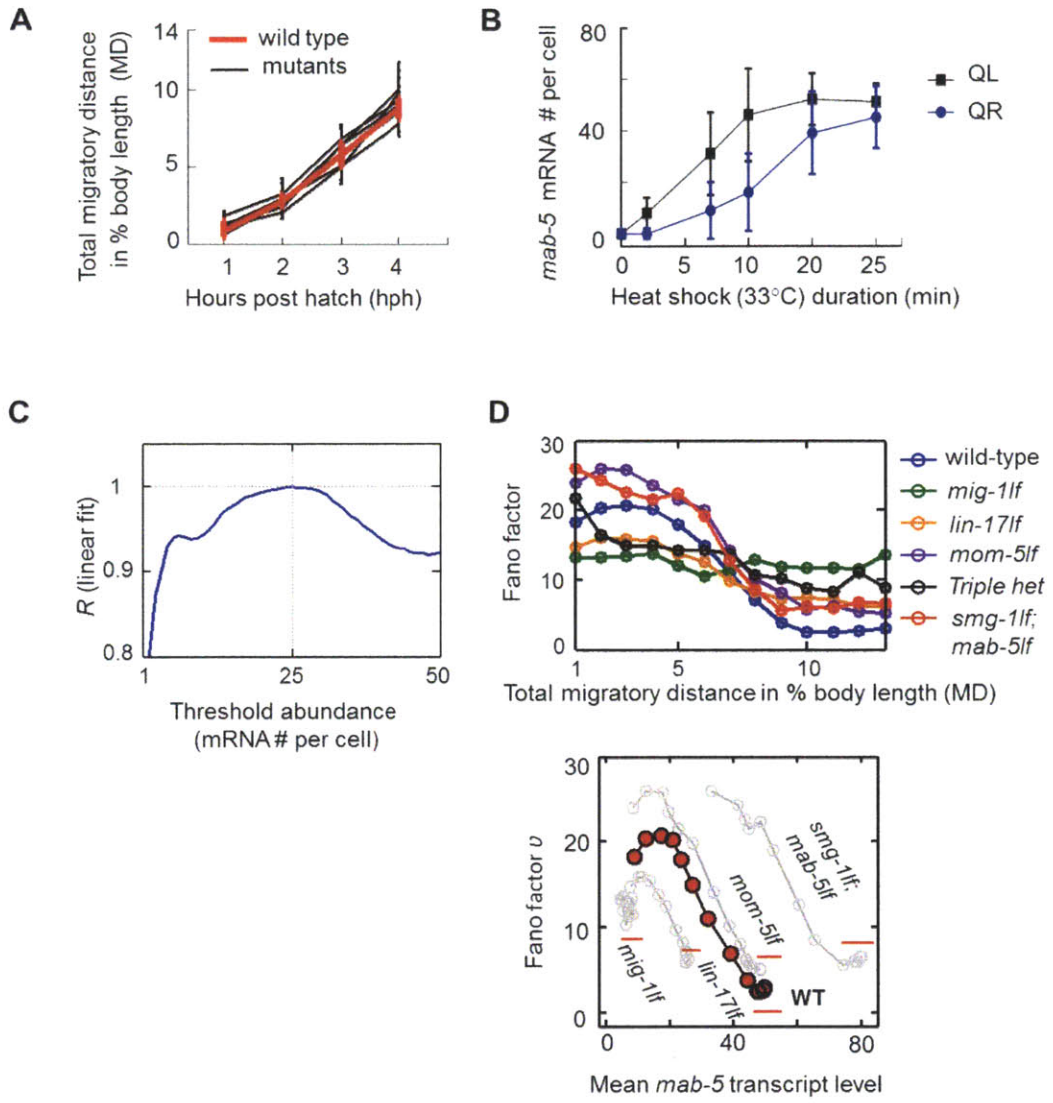



**Figure 2-S3.** A) Total migratory distance (MD) plotted against worm age (hours post hatching) in wild type (red) and all mutants (black) listed in Figure 3. Error bars mark 2.5-97.5 percentiles. B) *mab-5* transcript levels in the *egl-20 (n585); hs::egl-20* strain in response to various durations of 33°C heat shock. Note QR in the wild type does not express *mab-5*. Extended heat shock activated *mab-5* expression to wild-type levels in both QL and QR. Error bars are standard deviations of the mean. C) Correlation coefficients for the analysis in Figure 3C calculated for a full range of hypothetical thresholds in transcript abundance. D) Upper: Evolution of Fano factor values over the course of QL migration for wild type and various mutant strains. Lower:

Evolution of the relation between Fano factor and mean transcript abundance throughout QL migration. lf: loss of function.

**Perturbing EGL-20 and MAB-5 function increase variability in *mab-5* expression**

To test whether the increase in *mab-5* variability is unique to the Frizzled mutants, we next perturbed the input to the Wnt pathway, the EGL-20/Wnt gradient, using the *vps-29 (tm1320)* mutant. Loss of function of *vps-29* destabilizes the retromer complex, leading to a shortened and reduced EGL-20 gradient (Coudreuse et al., 2006). In these mutants, we found that *mab-5* expression was reduced to below 25 transcripts per cell in around 10% of QLs (Figure 2-3B). The cell-to-cell variability in *mab-5* expression again predicted the phenotypic penetrance; about 13% of the QL descendants were misplaced anteriorly (Figure 2-3C).

Conversely, we tested the effect of EGL-20 overexpression. Using an EGL-20 transgene under the control of a heat-shock promoter (Whangbo and Kenyon, 1999), we were able to drive EGL-20 concentration to a level higher than the wild type (see data in Figure 2-S3B). Increased EGL-20 concentration, however, did not significantly increase the average level of *mab-5* transcripts (Figure 2-3A, Mann-Whitney test, $p>0.1$). Rather, *mab-5* expression in late-stage QLs appeared less variable than the wild type, as indicated by a significant reduction in variance (F test $p<0.05$, Figure 2-3 B).

The transcriptional response to altered EGL-20 concentration may be explained by a "ceiling effect" whereby *mab-5* expression cannot exceed its wild type level. This possibility was, however, overturned by the observation of increased *mab-5* transcription in the *smg-1(e1228); mab-5(e1239)* double mutant (see Supplemental Information for the rationale of using the *smg-1* mutation background). While *mab-5* expression is essentially wild type in the *smg-1* single mutant (Figure 2-4B), significant increase in *mab-5* transcripts was observed in the *smg-1; mab-*

*5* double mutant (Mann-Whitney test, $p<0.001$), indicating potential feedback regulation of *mab-5* on its own transcription. This increase in average *mab-5* level was further accompanied by an increase in cell-cell variability (F test $p<0.001$). These observations suggest that low variability in gene expression cannot be simply attributed to saturation in *mab-5* transcription.

**A complex relationship exists between *mab-5* variability and average expression level**

To systematically compare the variability in *mab-5* expression across different genetic backgrounds, we calculated the Fano factors for both wild type and mutants over the course of QL migration.

In most strains, Fano factors were initially high and decreased over the course of migration to settle down at a stable value (Figure 2-S3D). Plotting the steady state Fano factors against the final transcript levels revealed several interesting features (Figure 2-3D). First, the Fano factors varied greatly across strains (range: 0.95-11.5). Thus a model of constitutive transcription with Poisson dynamics is insufficient to explain our observations. Alternatively, a model of bursty transcription would predict the Fano factor to increase (if burst size is modulated) or decrease (if burst frequency is modulated) monotonically with the mean (Raser and O'Shea, 2004). However, the observed relation between the Fano factor and the mean could not be summarized in a simple monotonic function. Mutants sharing similar average *mab-5* expression (e.g. the wild type and the triple heterozygotes) displayed up to 3 fold difference in Fano factor value. Additionally, the lowest Fano factor values were found in the wild type and the EGL-20 over-expression strains, which showed high but not the highest mean expression level. Taken together, the mode of transcription alone could not account for the observed variability in gene expression. Other

mechanisms, possibly upstream of *mab-5* transcription, may thus be at play to influence gene expression variability.

**All three Frizzleds are transcriptional targets of the Wnt pathway**

As *mab-5* expression consisted of distinct high and low subpopulations in a number of mutants (e.g. the *mig-1*single mutant and the triple heterozygotes), a feature attainable in systems with positive feedback (Becskei et al., 2001), we wondered whether feedback regulation exist within the Wnt pathway in QL. While Wnt signaling is conventionally viewed as a feedforward cascade, evidence from non-*C. elegans* species suggest that feedback regulation exists and may play a role in Wnt pathway regulation (Cadigan et al., 1998; Sato et al., 1999; Willert et al., 2002).

Should transcriptional feedback exist between the Frizzleds and the Wnt pathway, one would expect perturbations in Wnt signaling to induce a detectable change in the expression of the Frizzled paralogs. Indeed, loss of Wnt signaling either globally or Q-cell-specifically led to more than two fold difference in the expression of all three Frizzled genes (Figure 2-4A). *mig-1* and *mom-5* were upregulated, whereas *lin-17* was downregulated in both the *egl-20(n585)* and the *Pegl-17::DN-pop-1* strains. Interestingly, the difference in Frizzled expression between the wild type and mutant strains became progressively enlarged over the course of QL migration, consistent with the dynamics of Wnt signaling activation (Figure 2-1F). The observation that *mig-1* and *lin-17* levels stayed rather constant in the mutants indicates that feedback regulation likely underlies the dynamics of Frizzled expression in the wild type (Figure 2-4A and 2-2A).

To further understand how the Frizzleds respond transcriptionally to changes in Wnt signaling, we made use of all the aforementioned Wnt pathway mutants by ranking them according to their average *mab-5* expression levels. Thereby, we essentially obtained a mutant series in which Wnt

signaling activity was varied in a graded manner (Figure 2-4B, left panel). Notably, Frizzled transcript levels varied in a consistent manner across the strains. Strains with low *mab-5* expression carried low levels of *lin-17* and high levels of *mig-1* and *mom-5*, and the opposite were true for strains with high *mab-5* levels.

Intuitively, the increase in *mab-5* expression in the *smg-1; mab-5* mutants should predict a concomitant increase in *lin-17* and a decrease in *mom-5*. Our results however indicated otherwise (Figure 2-4B, right panel). This conflict is reconciled if we recognize *lin-17* and *mom-5* as downstream targets of *mab-5*. Loss of functional MAB-5, regardless of upstream Wnt signaling level, would then predict a decrease in *lin-17* and an increase in *mom-5*. The observation that *mig-1* levels were unaltered in the *mab-5* mutants suggests that the feedback to *mig-1* is likely *mab-5*-independent. These results suggest the existence of both *mab-5*-dependent and *mab-5*-independent feedback in the Wnt pathway.

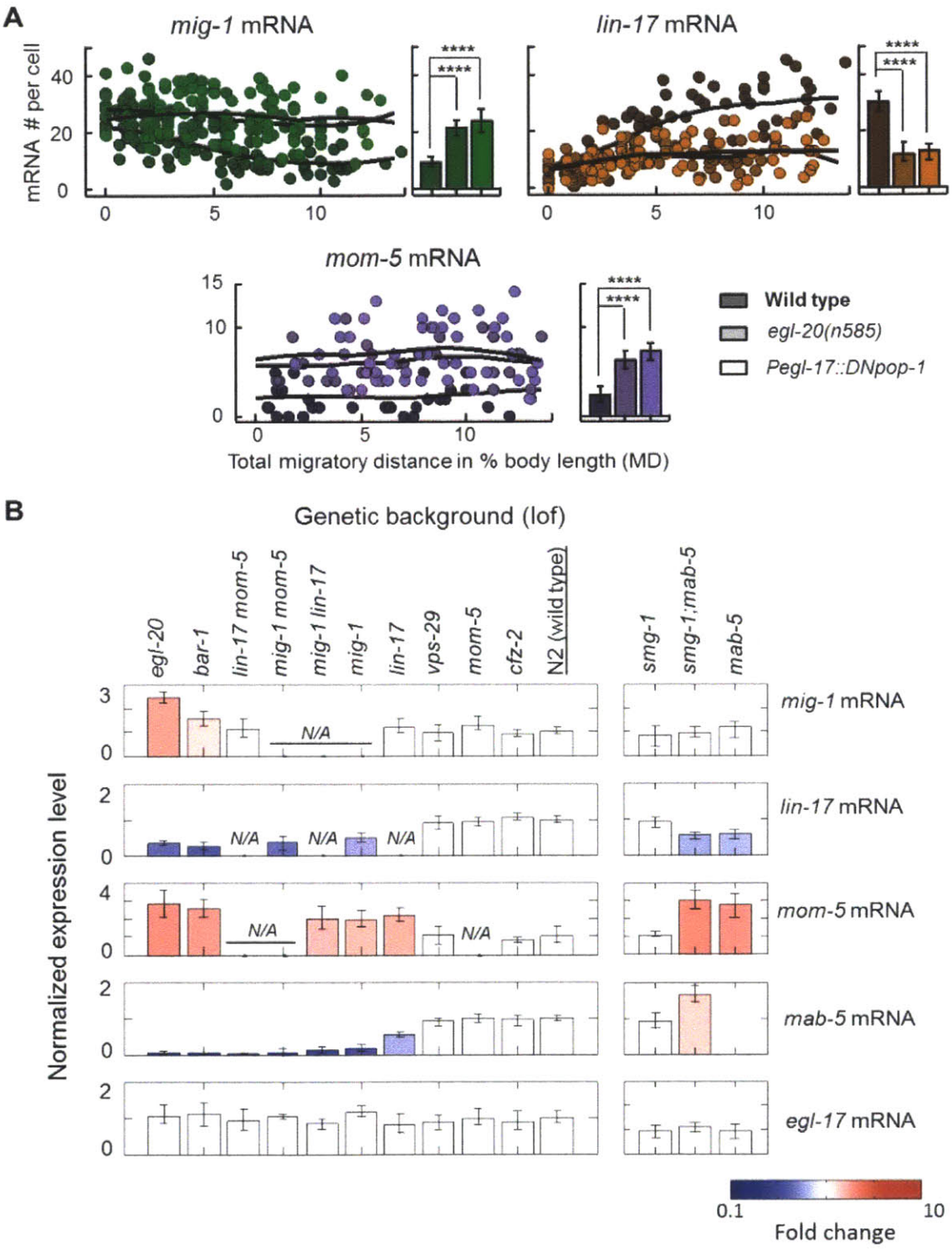

A
mig-1 mRNA
lin-17 mRNA
mom-5 mRNA
mRNA # per cell
Total migratory distance in % body length (MD)
Wild type
egl-20(n585)
Pegl-17::DNpop-1
****
B
Genetic background (lof)
egl-20
bar-1
lin-17 mom-5
mig-1 mom-5
mig-1 lin-17
mig-1
lin-17
vps-29
mom-5
cfz-2
N2 (wild type)
smg-1
smg-1;mab-5
mab-5
Normalized expression level
mig-1 mRNA
lin-17 mRNA
mom-5 mRNA
mab-5 mRNA
egl-17 mRNA
N/A
0.1
10
Fold change

**Figure 2-4.** Frizzled paralogs, *mig-1, lin-17* and *mom-5* are transcriptional targets of the Wnt pathway. A) Dynamic (left of each panel) and final average (right of each panel, corresponding to data on the left with MD>8) Frizzled transcript levels in mutants with global or QL-specific blockade of EGL-20 dependent Wnt signaling. Same wild type data as Figure 2-2A. ****$p<0.0001$. B) Normalized expression levels of Frizzleds and *mab-5* in various genetic backgrounds. Only values significantly different from the wild type (FDR corrected $p<0.05$) were colored. Genotypes of strains are indicated above the bar graph with same mutant alleles as indicated previously. Error bars are 95% CI of the mean.

**Interlocked positive and negative feedback loops exist within the Wnt pathway**

We next sought to incorporate the novel feedback interactions into a network model of the Wnt pathway. With feedback, perturbation to a single gene can propagate to affect many genes in the network, making it difficult to deduce the immediate targets of the perturbed gene. We tackle this general challenge in network inference by employing the Modular Response Analysis (MRA, Kholodenko et al., 2002). By iteratively perturbing every component of the network (a "component" could consists of one or multiple interacting genes) and measuring the response of the unperturbed components, MRA could uniquely infer the most probable network topology (Figure 2-5A). The inferred topology consists of only interactions between the "closest neighbors", thus avoiding redundant reference to the same network structure. Additionally, the inferred topology continues to hold true with the discovery of new network components (Kholodenko et al., 2002; Andrec et al., 2005).

The inferred topology consisted of a complex network of interlocked feedback loops (Figure 2-5C and 2-S4A-C). At the receptor level, the positive feedback targeting *lin-17* and the negative feedback targeting *mig-1* and *mom-5* intersect at the output end of the Wnt pathway. Such interlocked feedback topology may allow the Frizzled paralogs to cross-regulate each other's expression, as observed in the Frizzled mutants (Figure 2-4B). Downstream of the Wnt pathway, *mab-5* not only mediates feedback to the Frizzleds, but also negatively regulates its own transcription.

With multiple interlocked feedback loops, mutations in Wnt pathway components would not only perturb their protein function, but also alter the topology of the network. Supporting this is the preservation of strong single-cell correlation between *lin-17* and *mab-5* transcript levels in all strains with intact *LIN-17* and MAB-5 function (Figure 2-5D). In contrast, in the *smg-1; mab-5*

double mutant, where MAB-5 function was abolished and all *mab-5*-mediated feedback lost, the correlation between *lin-17* and *mab-5* was both altered and weakened. Thus, mutations in individual components of the Wnt pathway may also induce specific changes in the network topology.

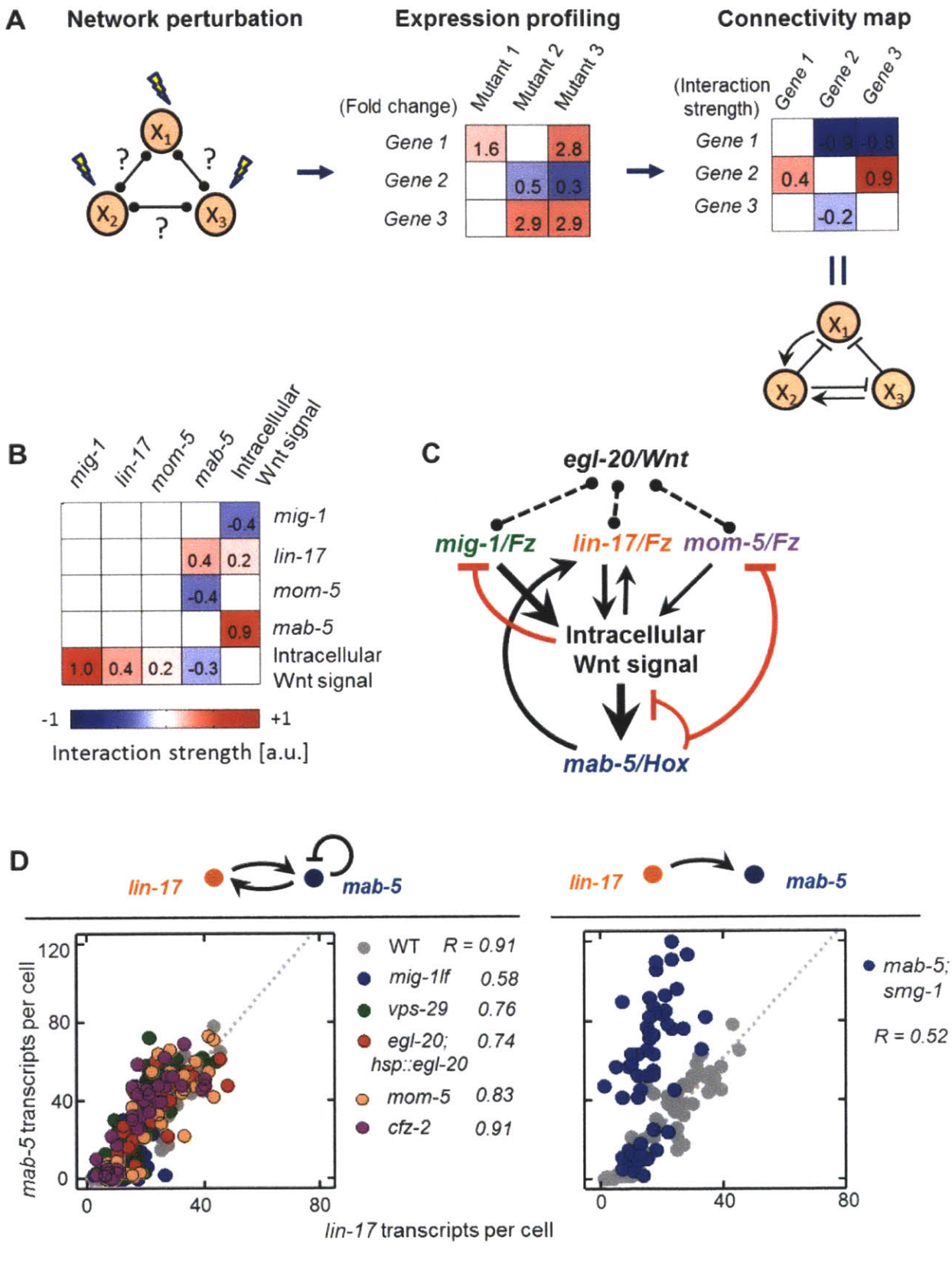

A
Network perturbation
Expression profiling
Connectivity map
X1
X2
X3
?
(Fold change)
Mutant 1
Mutant 2
Mutant 3
Gene 1
Gene 2
Gene 3
1.6
2.8
0.5
0.3
2.9
2.9
(Interaction strength)
Gene 1
Gene 2
Gene 3
-0.9
-0.8
0.4
0.9
-0.2
B
mig-1
lin-17
mom-5
mab-5
Intracellular Wnt signal
-0.4
0.4
0.2
-0.4
0.9
1.0
0.4
0.2
-0.3
-1
+1
Interaction strength [a.u.]
C
egl-20/Wnt
mig-1/Fz
lin-17/Fz
mom-5/Fz
Intracellular Wnt signal
mab-5/Hox
D
lin-17
mab-5
mab-5 transcripts per cell
lin-17 transcripts per cell
WT R = 0.91
mig-1lf 0.58
vps-29 0.76
egl-20; hsp::egl-20 0.74
mom-5 0.83
cfz-2 0.91
mab-5; smg-1
R = 0.52
0
40
80
120

**Figure 2-5.** Inferring the regulatory network within the Wnt pathway using the MRA algorithm. A) Schematic of the work flow for implementing the MRA algorithm. B) Inferred connectivity matrix. Only significant (p-value with Bonferroni correction <0.05) interactions are colored based on the inferred interaction strengths. C) Revised Wnt pathway model based on the inference results.

Strong correlation between *lin-17* and *mab-5* was preserved in strains with intact positive feedback from *mab-5* to *lin-17*, but was altered and reduced in the *smg-1; mab-5* mutant where the positive feedback was disrupted. All correlation coefficients (R) are with $p<0.001$.

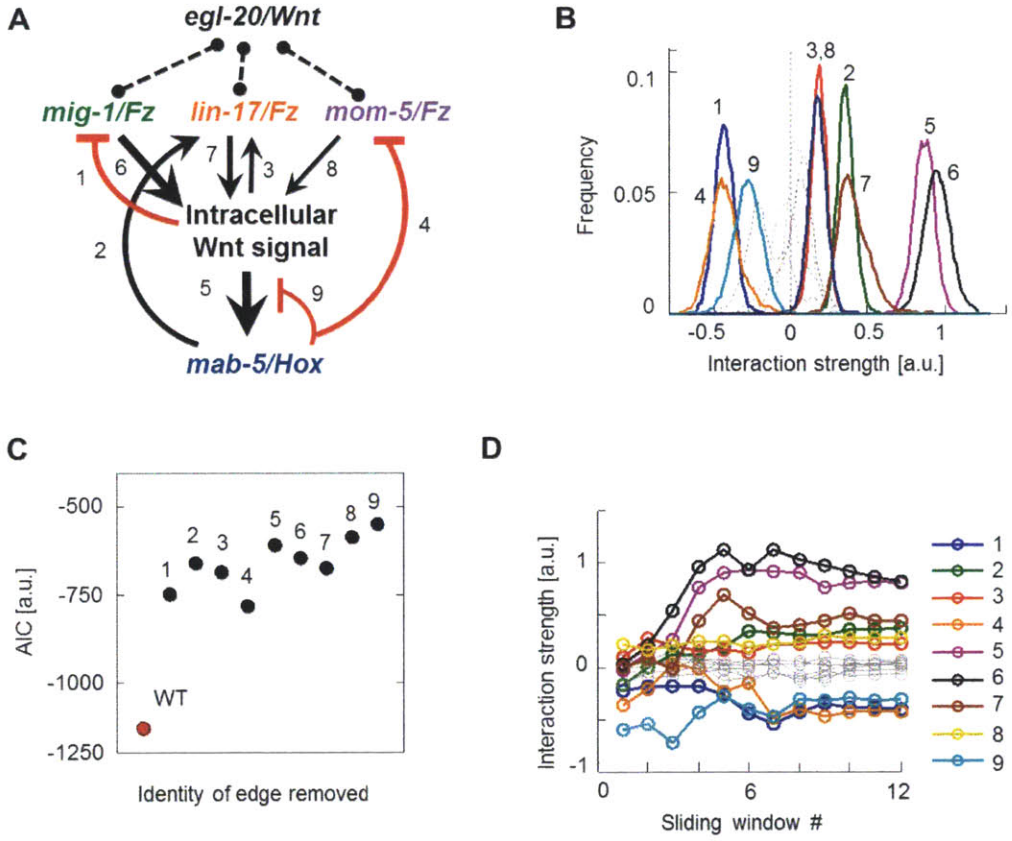



**Figure 2-S4.** A) The inferred network interactions are labeled 1-9 to correspond to the references in (B)-(D). B) Bootstrap distributions of inferred connectivity strengths. C) Akaike Information Criterion (AIC) values for the inferred network (red) and alternative topologies (black). Numbers on the data points indicate the network interactions that were forced to be zero. D) Network interaction strengths inferred using data from different ranges of MD values (sliding window # 1-

12, see Experimental Procedures). Colors correspond to individual network interactions as listed to the right.

**Positive and negative feedback cooperate to minimize variability**

To probe whether and how network topology influences variability in gene expression, we constructed an ordinary differential equation (ODE) model of the inferred network. Starting from the full 5 component network, we first obtained model parameters by fitting to both the steady state and the dynamic gene expression data (Figure 2-S5A and Supplemental Information). The full model was then reduced to a one-dimensional (1D) model by exploiting time scale differences between individual reactions (Figure 2-S5B and Supplemental Information).

The 1D model enabled us to directly calculate the Fano factor and mean output level for four classes of networks: those with no feedback, with negative feedback only, with positive feedback only, and with interlocked positive and negative feedback (Figure 2-6B). To explore general features of the four network classes, we varied the strengths of the feedback randomly between 0 to 10 times the wild type values while keeping other parameters fixed (Experimental Procedures).

As illustrated in Figure 2-6B, different classes of networks occupied distinct domains of the Fano factor -mean output space. Among the networks with no or a single type of feedback, there was a general trade-off between the Fano factor and the mean expression level. Specifically, negative feedback generally led to low output variability at the expense of the mean; positive feedback, while increasing the expression level, also made it more variable. Interestingly, this trade-off was alleviated in networks with interlocked positive and negative feedback. Many of the randomly sampled networks (blue circles) occupied the lower right quadrant (i.e. low variability and high mean), which was inaccessible by networks with no or a single type of feedback (see also Figure 2-S5C). Thus, with the same set of components, a network with

interlocked positive and negative feedback may more readily achieve low variability without compromising the mean expression level.

Since positive and negative feedback by themselves exhibited opposite effects on mean and variability, it is intriguing how the two types of feedback when combined can promote both high mean and low variability. We thus systematically examined the dependence of the mean and variability on feedback strength. As shown in Figure 2-6C, the mean expression level consistently decreased with increasing negative feedback strength, and increased with increasing positive feedback strength. The effect of negative feedback was essentially compensated by the positive feedback, resulting in intermediate mean values when both feedback loops are strong (Figure 2-6C, lower panel).

As for the Fano factor, increasing negative feedback generally led to a decrease in the Fano factor (Figure 2-6D; see also Figure 2-S5E), consistent with observations from synthetic circuits (Becskei and Serrano, 2000; Austin et al., 2006). The extent to which the Fano factor value decreases, however, depended strongly on the strength of the positive feedback. While the Fano factor decreased to around 0.8 at low positive feedback strength, it rapidly dropped to less than 0.5 at high positive feedback strength (Figure 2-6D, lower panel). As a result, the lowest Fano factor values were found when both the positive and the negative feedback were strong. This observation is surprising given that networks with positive feedback only showed increased output variability (Figure 2-6B). An intuitive explanation was found by recognizing that the "dampening" effect of the negative feedback is stronger at high mean expression levels (Figure 2-S5D). Thus, positive feedback indirectly contributes to the dampening of variability by promoting high mean expression levels.

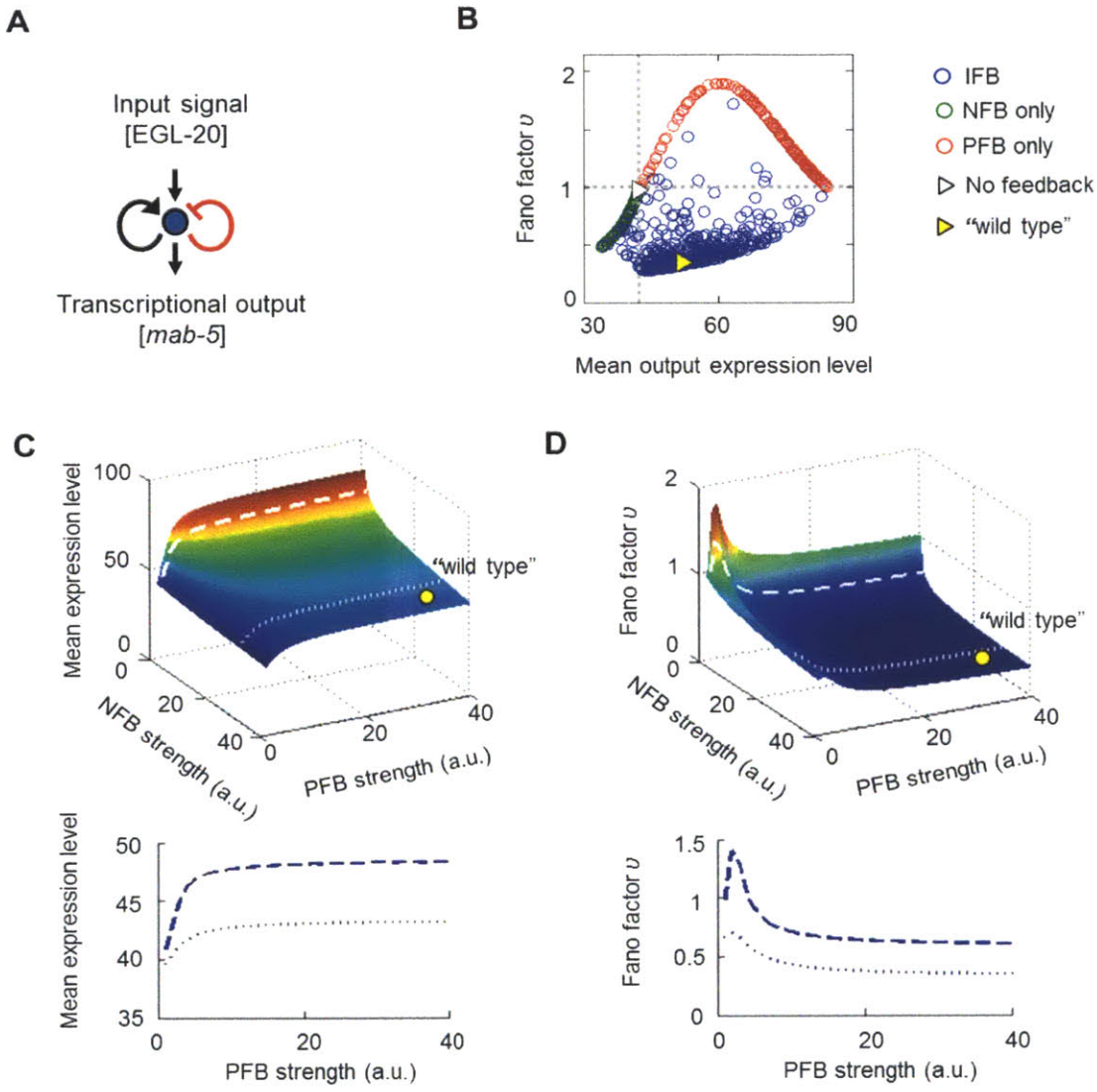



**Figure 2-6**. Modeling reveals synergistic contribution of positive and negative feedback in reducing variability. A) Schematic of the reduced 1D model. B) Analytically derived Fano factor versus mean output values for networks with interlocked feedback (IFB), negative feedback (NFB) only, positive feedback (PFB) only, and no feedback. "Wild type" indicates that the model parameters were obtained by fitting to the wild type data. C) Upper: Mean output value of the interlocked feedback network as a function of feedback strengths. Lower: Re-plotting of the broken and dotted lines in the upper panel. D) Upper: Fano factor value of the interlocked feedback network as a function of feedback strengths. Lower: Re-plotting of the broken and dotted lines in the upper panel. Note difference in Fano factor value at high PFB strength.

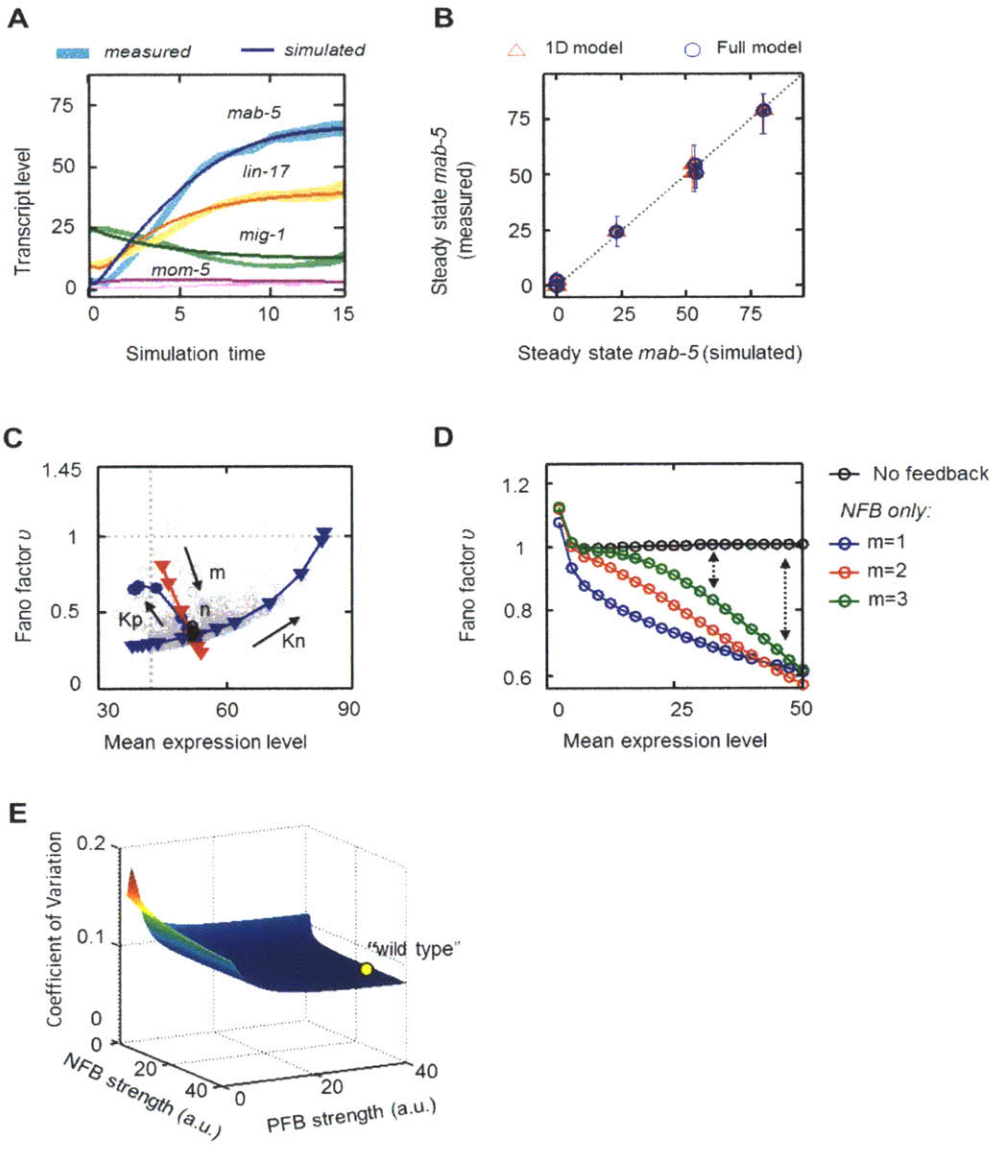



**Figure 2-S5.** A) Simulation of gene expression dynamics with the fitted 5-component model (full model). Colored patches represent experimentally measured gene expression levels whose boundaries indicate 95% CI of the mean. B) Simulation with the 1D model (red triangles) and the full model (blue circles) produced identical steady state *mab-5* expression values. Error bars are 95% CI of the mean. C) Dependence of the Fano factor value on the feedback activation thresholds and Hill coefficient values. Kp: activation threshold of the positive feedback (blue circles), Kn: activation threshold of the negative feedback (blue triangles), n: Hill coefficient of

the positive feedback (black circles, clustered around the intersection), m: Hill coefficient of the negative feedback (red triangles). D) Dependence of the Fano factor value on average gene expression level in a network with negative feedback. In comparison to a network with no feedback (black), negative feedback always reduces the Fano factor value, and does so more dramatically when the average gene expression level is high. E) Dependence of the coefficient of variation (standard deviation/mean) on the feedback strength. Same analysis as in Figure 2-6C&D.

**Model predicts *mab-5* variability in the mutants**

Since the network model was inferred and parameterized based on average transcript levels, we wondered whether it could predict the observed variability in various strains. In principle, variability in *mab-5* expression could result from multiple sources, including stochasticity in its own transcription as well as fluctuations internal and external to the Wnt pathway. We thus combined the effect of all unmodeled fluctuation into a single term (D) and determined the magnitude of this "extrinsic fluctuation" by fitting to the wild type *mab-5* distribution (see Supplemental Information). Remarkably, the revised model not only captured the distribution of *mab-5* levels in the wild type, but also predicted the direction of changes in *mab-5* variability in various Wnt pathway mutants (Figure 2-7 A, 2-7 B and 2-S6A)*. Thus, alterations in network topology likely underlie the changes in *mab-5* variability across the mutants. Conversely, the intact network in wild type may contribute strongly to the observed low variability in *mab-5* expression.

*Note that the goal of the model is to predict qualitative changes (i.e. increase or decrease) in the level of gene expression variability, rather than to produce perfect matching in the shape of the distribution. The shape of the distributions predicted by the model is influenced by the assumption on the distribution of the extrinsic noise (e.g. white noise, or Poisson-distributed noise), which can not be inferred from the current measurements.

Together, our results support a model in which variability in gene expression is controlled through an intricate network of feedback regulation between the Wnt pathway and its own components. Specifically, the signal amplifying effect of the positive feedback appears to be co-opted to ensure a strong negative feedback, one which is needed to effectively dampen fluctuations in gene expression (Figure 2-7C). Increasing evidence of feedback regulation challenges the conventional notion of signaling pathways as linear, unidirectional cascades. It is likely the rule rather than the exception that feedback regulation is widely exploited as a flexible mechanism to achieve phenotypic stereotypy when desired.

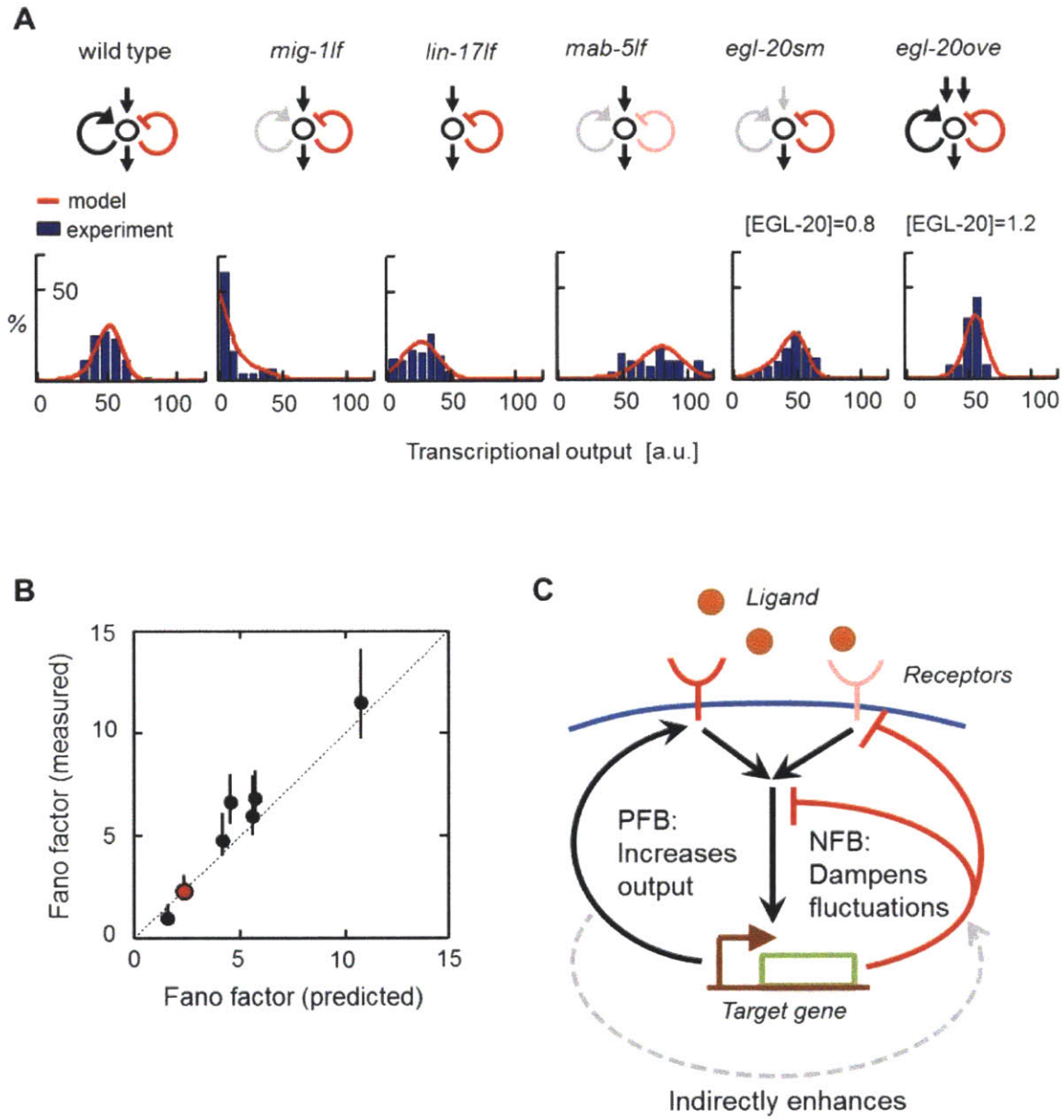



**Figure 2-7.** Model predicts observed gene expression variability across strains. A) Model prediction (red curves) of the distribution of *mab-5* transcript levels in wild type and various Wnt signaling mutants. Network diagrams indicate the speculated changes in network topology. Grey arrows indicate weakened interactions and double arrows symbolize an increase in EGL-20 concentration. B) Theoretically predicted versus the experimentally measured Fano factor values

for the strains shown in (A). Error bars are 95% CI of the mean. C) Conceptual explanation of the interplay between the positive and the negative feedback in reducing variability.

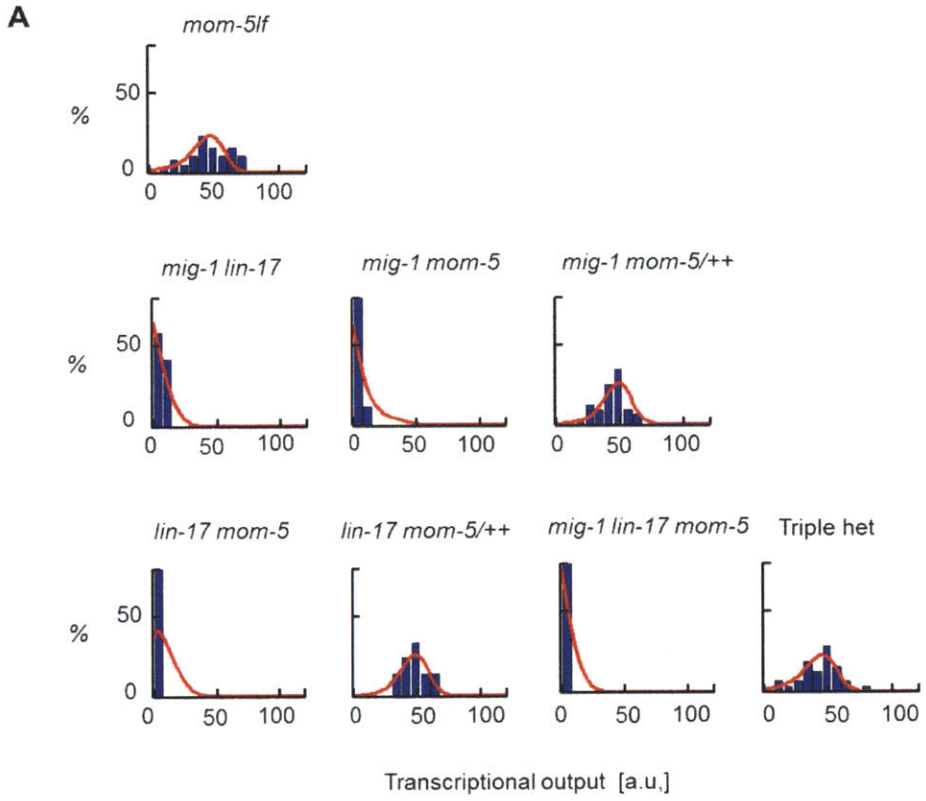



**Figure 2-S6.** A) Comparison of the theoretical (red curves) and measured (blue histograms) steady state mab-5 distribution in mutants not listed in Figure 7A. Note for the heterozygote mutants, functional mig-1 and mom-5 transcripts were assumed to be 70% of their wild-type levels, and functional *lin-17* transcript was modeled to be at 80% of its wild-type level. The exact amounts of these transcripts could not be measured due to the presence of residual transcripts from the mutant allele.

## 2.3. DISCUSSION

### Regulatory network as an endogenous mechanism to control variability

A major effort in biology has been to understand how biological events such as animal development proceed so robustly and how this robustness is disrupted in the face of disease. As gene expression critically drives the phenotypic outcome, understanding how endogenous gene expression is controlled to ensure robustness is of great interest. Our results underscore the importance of regulatory networks in controlling transcriptional variability in a multicellular organism. We show how two common regulatory modules, positive and negative feedback, can cooperate to yield tightly regulated gene expression in a single cell. While individual properties of each feedback motif have been examined in detail, the joint action of the two appears more complex (Acar et al., 2005; Brandman et al., 2005). In fact, the interlocked feedback motif was mainly explored for its role in generating oscillations (reviewed in Ferrell et al., 2011). Our findings suggest that the same motif can be adapted to ensure stable gene expression even at elevated levels. The versatility of the interlocked feedback motif exemplifies the rich potential of regulatory networks in implementing robust gene expression control.

**Sources of variability in *mab-5* expression**

While wild type *mab-5* expression is less variable compared to the Wnt signaling mutants examined in this study, its steady-state Fano factor is well above the theoretical expectation of 1. What are the sources of variability in addition to stochastic transcription? Studies in mammals and yeast suggest the infrequent bursts in transcript production (transcription bursting) as a dominant source of variability. Bursty transcription, however, does not seem to apply to *mab-5* expression in the wild type QL. This is because bright transcription sites (Figure 2-1D and S1C) and high *mab-5* expression were observed for all QLs at the end of migration.

Given the dependence of *mab-5* expression on Wnt signaling, it is plausible that fluctuations in the levels of the Wnt ligand or receptors led to the heightened variability in *mab-5*. For

example, the stochastic diffusion of a morphogen molecule may lead to fluctuations in its local concentration (Gregor et al., 2007). Variable expression of the receptors, which was observed for the three Frizzleds (Figure 2-2B), could also affect the transcriptional output of the Wnt pathway. Fluctuations in downstream components of the Wnt pathway could also propagate to affect the transcription of *mab-5*. As multi-layered cascades, signaling pathways are subject to many sources of fluctuations and thus in strong need to control variability in their output.

**Variability in gene expression carries signatures of the network topology**

An emerging view in the study of stochastic gene expression argues that variability, or noise, can inform about the underlying mechanism of regulation (Cağatay et al., 2009; Chalancon et al., 2012; Munsky et al., 2012). In this study, we used the average gene expression to infer network topology, and found a surprising link between network topology and the variability in gene expression. In retrospect, signatures of the inferred network topology may be readily present in the gene expression distributions. For example, low variability in the wild type and the inability to increase *mab-5* level by EGL-20/Wnt overexpression suggest the existence of a negative feedback loop (Figure 2-3B and D). A perturbation experiment that eliminates the putative negative feedback loop was thus carried out to test this possibility (Figure 2-3B). Similarly, the distinct subpopulations of *mab-5* ON and OFF cells in strains such as the *mig-1* single mutant implicate the existence of positive feedback. Further, the strong single-cell correlation between *lin-17* and *mab-5* indicates a strong positive coupling which could arise either through a common upstream activator or a strong positive feedback. Both mechanisms turned out to exist in the network (Figure 2-5C). Taken together, variability in gene expression may carry distinct signatures of the underlying network and can serve as a useful guide to network identification.

## 2.4 EXPERIMENTAL PROCEDURES

### *C. elegans* strains and culture

*C. elegans* strains were grown at 20°C using standard culture conditions and techniques. The Bristol N2 strain was used as wild type. Mutant alleles and transgenes used in the study are:

Linkage Group I (LGI*): mig-1(e1787), lin-17(n671), smg-1(e1228), mom-5(gk812), mom-5(or57), mom-5(ne12);*

LGIII: *mab-5(e1239), vps-29(tm1320);*

LGIV: *egl-20(n585);*

LGV: *heIs63[Pwrt-2::gfp-ph; Pwrt-2::h2b-gfp; Plin-48::tomato], muIs53[hs::egl-20; unc-22(dn)]* (Whangbo and Kenyon, 1999);

LG unknown: *syIs187[POPTOP]*;

Extra-chromosomal array: *huEx278[Pegl-17::DNpop-1;Pmyo-2::tomato].*

Note on the *smg-1(e1228); mab-5(e1239)* strain: The *smg-1 (e1228)* background was used to allow visualization of *mab-5* transcripts in the *mab-5(e1239)* mutants. Normally, transcripts from nonsense mutant alleles (as in the case of *mab-5(e1239)*) are subject to nonsense-mediated RNA decay (Pulak and Anderson, 1993), which abolishes the majority of the mutant transcripts. We circumvented this problem using the *smg-1(e1228)* mutation which compromises the nonsense mediated decay pathway (Denning et al., 2001; Grimson et al., 2004).

### Scoring QL descendent migration

The precise positions of the Q descendants QL.pap/QL.paa were scored by DIC microscopy in late L1 stage larvae as described (Coudreuse et al., 2006).

### Single molecule fluorescence in situ hybridization

SmFISH was performed as described (Raj et al. 2008). Manual segmentation of GFP-marked QL periphery was performed, followed by automated spot counting in MATLAB-based custom-written software. Total migratory distance (MD) was assayed by manually marking the nuclear position of QL and QR, tracing the A-P axis of the worm, and automatically computing the distance between QL and QR along the A-P axis.

**Heat shock activation of *hsp::egl-20***

Heat shock experiments were performed on animals carrying *muIs53 [hsp::egl-20 ; unc-22(dn)]* as described (Whangbo and Kenyon, 1999). Briefly, heat shock treatment was given to 0-0.5 hour synchronized L1 larvae in a total volume of 50 µl at 33°C for a desired length of time. Heat shock was terminated by chilling tubes on ice for 10 seconds and worms were then grown on fresh plates at 20°C for an additional 2-2.5 hours.

**Network inference**

Following procedures outlined in Kholodenko et al. 2002, the raw transcript count for a given gene in a given strain is transformed into the central fraction difference (CFD) value using the following formula:

$$\Delta ln\, x_j = 2 * \frac{\left(x_j{}^{(1)} - x_j{}^{(0)}\right)}{x_j{}^{(1)} + x_j{}^{(0)}}$$

where $x_j{}^{(0)}$ denotes the transcript count of gene j in the wild type, and $x_j{}^{(1)}$ the transcript count in the strain of interest (e.g. all wild type measurements would yield a CFD value of 0). Based on the transformed CFD value, a 5x5 matrix was assembled by bootstrap sampling for each gene in each mutant background. This "gene expression matrix" was then substituted into the following inference algorithm (Van Kampen, 1983):

$$\boldsymbol{r} = -[diag(\boldsymbol{R}^{-1})]^{-1} * \boldsymbol{R}^{-1}$$

Where **r** stands for the connectivity matrix, **R** represents the gene expression matrix, and "diag" denotes the matrix diagonal. Bootstrapping followed by the inference routine was repeated 10,000 times and the resulting connectivity matrices were pooled to obtain the probability distribution and FDR adjusted p-value for each connectivity strength.

**Computational validation of the inferred topology**

To control for network complexity, we compared the Akaike Information Criterion (AIC) value (Akaike, 1974) of the wild-type network to a series of 9 alternative networks using the following equation (Burnham and Anderson, 2002):

$$\mathrm{AIC} = n \ln(\mathrm{RSS}/n) + 2k + C$$

Where RSS is the residual sum of squares from model fitting, n denotes the size of the data set used for model fitting, k denotes the number of model parameters, and C is a constant independent of the model (which we set to zero for simplicity). Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. The alternative networks were constructed by eliminating one at a time of the inferred network interactions. Specifically, this was done by forcing the corresponding model parameter to be zero and proceed with least-squares parameter fitting. The AIC value allows for a fair comparison between models with different number of parameters (model complexity). As shown in Figure 2-S4C, the inferred model yielded a much lower AIC value than all the alternative models tested, which validates the inferred model as the most likely model given the gene expression data set.

We next tested the sensitivity of inferred topology to the definition window of steady state gene expression. Briefly, a sliding window that spans MD=X-1 to MD=X+3, where X varied

from 1 to 12, were used to selected data points for network inference. Following the same data transformation as outlined in Methods, the MRA algorithm was applied and the resulting average interaction strengths were reported in Figure 2-S4D. The inferred values varied greatly for X=1 to 6, but stabilized for X>6. The initial fluctuation in interaction strengths reflected the initial variation in transcript abundance, and the stabilization of inferred interaction strengths validate the use of MD>8 as a window for steady state gene expression.

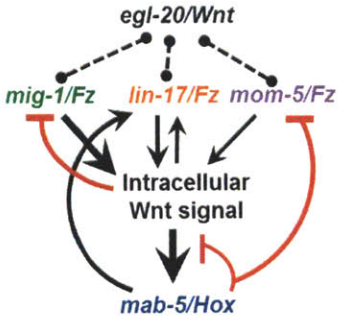
**Statistical analysis**

The Mann-Whitney test was used to compare mean expression levels and the F test was used to test equal variance between the wild type and mutants. Non-parametric bootstrap was used to derive confidence intervals on the mean transcript counts and the Fano factors values. To control for multiple comparisons, the Benjamini-Hochberg procedure was used to achieve a false discovery rate (FDR) of less than 0.04 for comparison of transcript abundance; the Bonferroni correction with n=20 was applied to the bootstrap p values of the inferred network interactions. Corrected p-value of less than 0.05 was considered significant.

**Modeling**

**1. Model construction**

Based on the inferred network topology (Figure 2-5C),




we constructed the following ordinary differential equation (ODE) model. As listed below, the variables R1-3, W and T denote the per-cell abundance of each of the five network components.

R1: *mig-1*
$$\dot{R1} = \gamma 1 * \left(V1 * \left(\frac{K1^{n1}}{K1^{n1}+W^{n1}}\right) + \beta 1 - R1\right) \quad (1)$$

R2: *lin-17*
$$\dot{R2} = \gamma 2 * \left(V2 * \left(\frac{c*T^{n2}+W^{n2}}{K2^{n2}+c*T^{n2}+W^{n2}}\right) - R2\right) \quad (2)$$

R3: *mom-5*
$$\dot{R3} = \gamma 3 * \left(V3 * \left(\frac{K3^{n3}}{K3^{n3}+T^{n3}}\right) - R3\right) \quad (3)$$

W: Intracellular Wnt signal
$$\dot{W} = \gamma w * \left(Vw * L * \left(\frac{R1 + s1 * R2 + s2 * R3}{+s3 * R1 * R2 + s4 * R1 * R3}\right) * \left(\frac{Kw^{n4}}{Kw^{n4}+T^{n4}}\right) - W\right) \quad (4)$$

T: *mab-5*
$$\dot{T} = \gamma t * (W - T) \quad (5)$$

**Note 1:** $L$ denotes the amount of EGL-20 ligand available. Since *egl-20* was expressed at similar levels across strains (data not shown), we assumed the local EGL-20 level to be unaltered in all strains except for mutants directly affecting the function or the secretion of EGL-20 (i.e. *egl-20 (n585)* and *vps-29 (tm1320)*). The value of $L$ denotes the local concentration of $L$ relative to the wild type. To minimize the number of free parameters, we assumed $L$ to stay constant in all genetic backgrounds and assigned $L$ with an arbitrary value of 1. Potential interactions between

the Frizzleds and the Wnt ligand, while not explored in the current study, can be incorporated into the model by modeling L as a function of the R1~3 (i.e. $L=L(Ri)$).

**Note 2:** To model the non-additive relation between Frizzled single mutants and compound mutants (e.g. the reduction in *mab-5* in the *lin-17 mom-5* double mutant is much greater than the summed loss in the two single mutants (Figure 2-3A)), it was necessary to include the interaction terms (i.e. $s * Ri * Rj$) in (4). This way the near complete loss of mab-5 expression in Frizzled compound mutants can be recapitulated by the model.

## 2. Parameter estimation

We next fitted the above model to the gene expression profile of wild-type and mutant strains. To constrain the number of parameters and facilitate fitting, we manually tested Hill coefficient values of 1, 3, 6 and 9. We only accepted higher Hill coefficients when the increase in Hill coefficient led to considerable decrease in the mean squared error (MSE) without a compromise in the $p$-values. We used similar criteria to determine whether to include the basal transcription rates ($\beta i$) in a given equation. In general, keeping the minimal number of parameters yielded close approximation of the experimental data. The following parameter estimates were obtained from the least squares fitting routines in MATLAB (MathWorks, Natick, MA):

| Parameter | Estimated value | Standard Error | p-Value | RMSE |
|---|---|---|---|---|
| *V1* | 11.1 | 0.9 | 6.40E-26 | 2.8 |
| *K1* | 0.2 | 0.2 | 0.043 | |
| *beta1* | 9.8 | 0.4 | 2.58E-57 | |
| *n1* | 1.0 | - | - | |
| *V2* | 34.0 | 1.4 | 1.30E-56 | 1.6 |
| *c* | 7.9 | 1.9 | 6.00E-05 | |
| *K2* | 58.4 | 12.0 | 2.50E-06 | |
| *n2* | 3.0 | - | - | |
| *V3* | 6.1 | 0.3 | 9.80E-53 | 1.1 |
| *K3* | 42.8 | 7.9 | 1.80E-07 | |
| *n3* | 1.0 | - | - | |
| *Vw* | 3.6 | 0.2 | 5.62E-21 | 17.7 |
| *s1* | 0.4 | 0.1 | 8.90E-09 | |
| *s2* | 0.27 | 0.1 | 0.015 | |
| *s3* | 0.0014 | 3E-4 | 2.6E-5 | |
| *s4* | 0.13 | 0.02 | 3.8E-5 | |
| *Kw* | 62.1 | 8.9 | 8.80E-17 | |
| *n4* | 6.0 | - | - | |
| *γ1* | 0.4 | - | - | 7.5 |
| *γ2* | 1.1 | - | - | |

| | | | |
|---|---|---|---|
| $\gamma 3$ | 10.0 | - | - |
| $\gamma w$ | 0.06 | - | - |
| $\gamma t$ | 5.7 | - | - |

Note the parameters V1 through n4 were fit using steady state gene expression data using the NonLinearModel.fit routine in MATLAB. The degradation rates were fit using time series gene expression data using the lsqnonlin function in MATLAB.

### 3. Model reduction

With the fit parameters, the model exhibits temporal dynamics that approximates the wild-type expression profile (Fig S5A). Next, by observing that $\gamma w$ is much smaller than the other time scales ($\gamma 1$-$3$, and $\gamma t$), we reduce the model down to a deterministically equivalent 1D model by setting equations (1)-(3) and (5) to zero and substituting the resulting equalities into equation (4):

$$\dot{W} = \gamma w * \left(Vw * L * Rtot * \left(\frac{Kw^{n4}}{Kw^{n4}+W^{n4}}\right) - W\right) \tag{6}$$

where $Rtot = R1 + s1 * R2 + s2 * R3 + s3 * R1 * R2 + s4 * R1 * R3$ with $R1$-$3$ defined as in (1)-(3).

### 4. Analytical calculation of the Fano factor

To obtain the theoretical Fano factor values, we first rewrite the 1D model in the following format:

$$\dot{x} = f(x) - g(x) \tag{7}$$

where $f(x) = \gamma w * Vw * L * Rtot * \left(\frac{Kw^{n4}}{Kw^{n4}+W^{n4}}\right)$, $g(x) = \gamma w * W$, and $Rtot$ defined the same as in (6).

To model the effect of stochastic fluctuations, we construct the following Langevin model:

$$\dot{x} = f(x) - g(x) + \sqrt{f + g + D} * \Gamma \tag{8}$$

where the magnitude of the intrinsic and extrinsic fluctuations are defined by the their autocorrelation functions:

Intrinsic fluctuation $\epsilon i$: $< \epsilon i(t)\epsilon i(t') >= 2 * (f + g) * \delta(t - t')$

Extrinsic fluctuation $\epsilon e$: $< \epsilon e(t)\epsilon e(t') >= 2 * D * \delta(t - t')$

Here we assume the intrinsic and the extrinsic fluctuations are uncorrelated. The stationary probability distribution of x can then be analytically calculated using the Fokker-Planck formalism. The drift ($A(x)$) and diffusion ( $B(x)$) terms in a corresponding Fokker-Planck model can thus be represented as (Van Kampen, 1983):

$$A(x) = f - g \text{ and } B(x) = f + g + D$$

The value of D was set to 0 for the computational analysis on the effect of network topology on output variability (Figure 2-6 and S5) and determined by fitting to the wild-type *mab-5* distribution to allow the prediction of *mab-5* variability in the Wnt signaling mutants (Figure 2-7 and S6).

It follows that the probability density of the stochastic model is:

$$p(x) = \frac{const}{B(y)} * \exp[2 * \int_0^y \frac{A(y')}{B(y')}]dy' \tag{9}$$

The mean and variance then follow from the probability density, with the ratio of the two leading to the expression of the Fano factor $v$:

$$v(x) = \frac{\int [x^2 * p(x)]\, dx - (\int [x * p(x)] dx\})^2}{\int [x * p(x)] dx\}}$$

**5. Predicting mab-5 variability in the mutants:**

Following the derivation of the probability density function, the distribution of *mab-5* expression can be directly calculated for different network topologies. As the exact level of the extrinsic fluctuations (*D*) is unknown, we first fit the wild-type *mab-5* distribution, D was varied incrementally from 0 up and a value of 450 was found to well approximate the wild-type distribution. The same value of D was then kept fixed while the topology of the model network was varied to mimic the altered network topology in the mutants. The probability density function was then derived and the Fano factor value calculated for each "mutant" network.

# Chapter 3

# TRANSCRIPTIONAL REGULATION OF WNT PATHWAY GENES IN THE Q DAUGHTER CELLS

## 3.1 Introduction

While past studies have convincingly established the requirement of wild type *mab-5* in the posterior migration of the QL descendants, an outstanding question remains. Namely, is *mab-5* expression in QL required for its continued expression in the QL daughter cells? Past studies have shown that forced activation of *mab-5* expression in the Q descendants (through laser activation of a transgene containing heat shock promoter driven *mab-5* cDNA specifically in the Q descendants) can induce posterior migration even in cells already migrating towards the anterior (Figure 3-1, Salser and Kenyon, 1992). This result demonstrates that *mab-5* expression in QL is not absolutely required for its continued expression in the QL descendants. However, it does not exclude the possibility that the expression of *mab-5* in QL contributes to the posterior migration of the QL descendants. The goal of this chapter is thus to characterize how Wnt signaling and the expression of *mab-5* in QL influences the gene expression dynamics of its daughters. Addressing this question would not only further our understanding of the genetic basis of Q neuroblast migration, but also provide general insights into how transcriptional regulation is coordinated with cell cycle and lineage progression.

To test the requirement of QL-specific *mab-5* expression in the posterior migration of the QL daughter cells, one would ideally like to block Wnt signaling or disrupt *mab-5* function in the

QL neuroblast but not its daughters. To proceed with such experimental manipulation, two conditions need to be satisfied. First, one needs to identify a promoter that is active selectively in the QL but not its descendants. Second, even with the use of a QL-specific promoter, one still needs to ensure that the effect of the experimental manipulation is confined within the lifetime of QL. This is because the expression of a transgene or an siRNA in QL could easily affect the state of the daughter cells via direct transfer of cytoplasmic and nuclear materials during cell division. As I show later in this chapter, while the first condition can be met, the second is more difficult to satisfy with existing technology. Taken into account this technical difficulty, I present here evidence from quantitative gene expression assay and genetic manipulation that the QL neuroblast may play a positive role in promoting the posterior migration of its daughters.

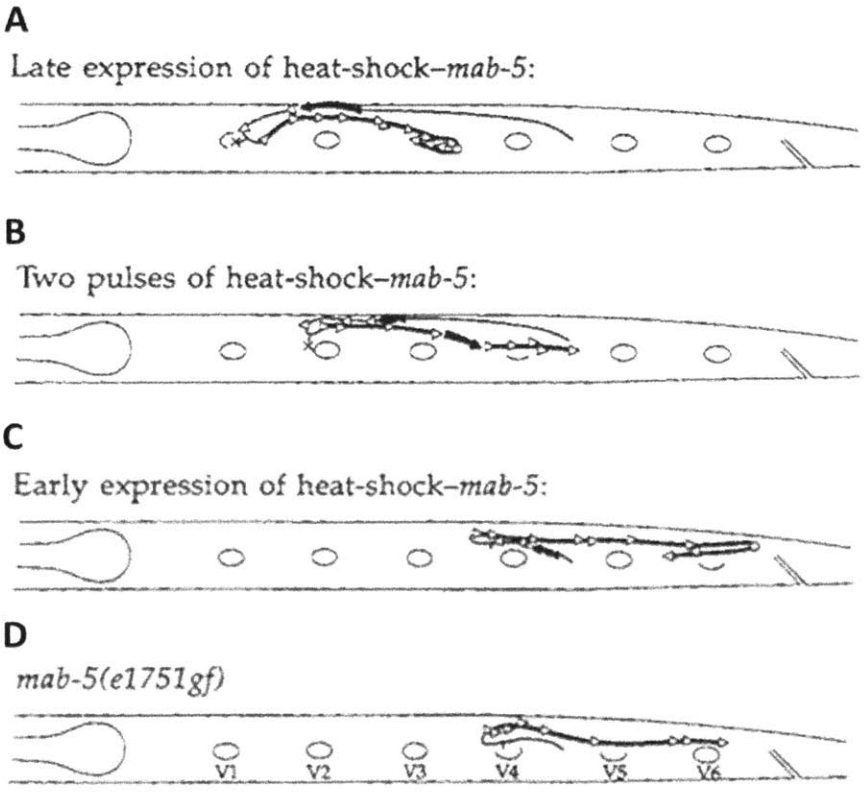


**Figure 3-1.** Laser activation of a heat shock promoter driven *mab-5* transgene in the Q daughter cells is sufficient to drive posterior migration. Shown here are combined results from experiments on both QL and QR descendants, all in a *mab-5(e2088lf)* background. Black traces mark the combined migratory trajectory of Q , Q.a and Q.ap cells, with the trajectory of Q.ap cells marked in bold. Open arrowheads indicate the positions and inferred migratory directions of Q descendants when individual data points are taken. Thick arrows indicate timing and duration at which heat shock is given. Image courtesy Salser and Kenyon, 1992.

## 3.2 Transcript Abundance in the Q Daughter Cells Is Influenced by Both Direct Inheritance through Cell Division and Cell Type Specific Transcriptional Regulation

Using smFISH and semi-automated image analysis as outlined in the previous Chapter, we quantified the transcript level of five Wnt receptor and two Hox genes in the Q mother cells and Q daughter cells in both the QL and, as a comparison, the QR lineages. In addition to the five genes analyzed for QL (*mig-1/Fz, lin-17/Fz, mom-5/Fz, cfz-2/Fz, mab-5/Hox*) in Chapter 2, I additionally included *cam-1*, a Ror receptor tyrosine kinase family member found to inhibit canonical Wnt signaling in the left Q lineage (Forrester et al., 2004, Figure 3-2), and *lin-39*, another *C. elegans* Hox homolog involved in regulating the long-distance anterior migration of the Q descendants (Maloof et al., 1999, Figure 3-3).

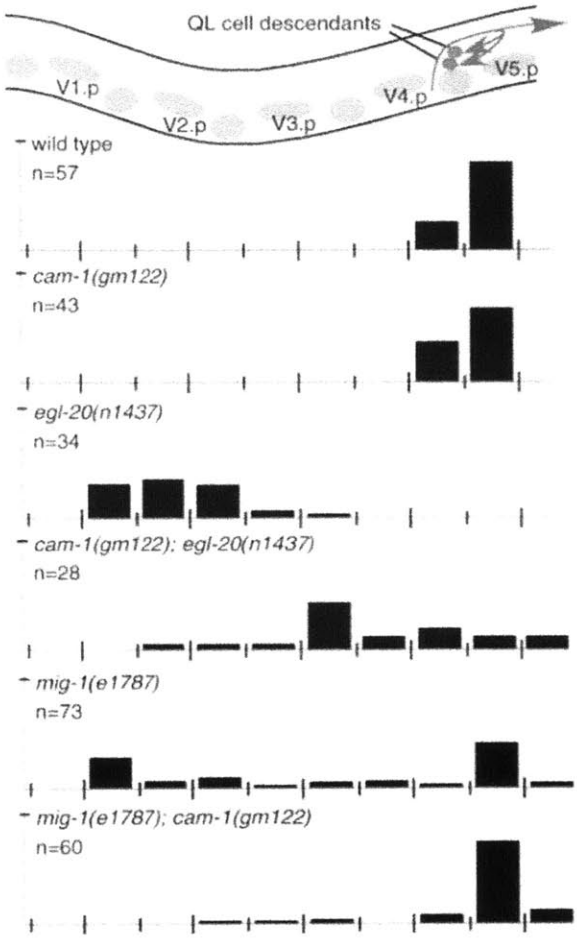



**Figure 3-2.** Q neuroblast migration in wild type, *cam-1(gm122)* and other Wnt signaling mutants. Top panel is a schematic of the lateral view of a late L1 animal. Anterior is to the left and dorsal is up. Histograms below the schematic indicate the final positions of the cell bodies of the Q descendants, QL. paa and QL.pap. The long tick marks on the x-axis indicate the location of Vn.p nuclei and the short tick marks indicate the location of Vn.a nulcei. The tick mark on the y-axis denotes 100%. Image courtesy Forrester et al., 2004.

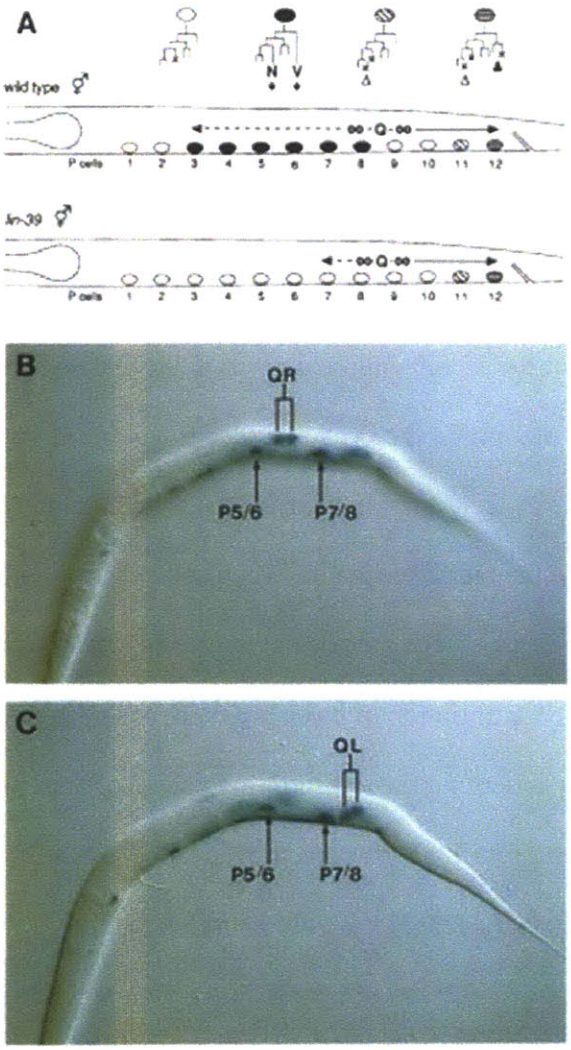



**Figure 3-3.** Function and expression pattern of *lin-39*. A) Summary of P cell lineages and Q cell migrations (Sulston and Horvitz, 1977) in wild type and *lin-39* mutants. Diamonds and arrowheads denote cell fate dependent on Hox gene function. Closed diamonds denote *lin-39*-dependent cell fates. Open arrowheads indicate cell deaths (x) that require *mab-5* activity (Kenyon, 1986). Q neuroblast migration on the left is indicated by a solid arrow, while the migration on the right is marked by a dashed arrow. B) and C) Expression of an integrated *lin-39::lacZ* fusion transgene in an L1 larva. The two panels show the right (B) and the left (C) sides of the animal. LacZ staining is present in the Q daughters of both the right and the left lineages, as well as in the P lineage of ectoblasts. Image courtesy of Wang et al., 1993.

Figure 3-4 summarizes the transcription dynamics of all seven genes in QL/R and their daughters. Immediately prior to QL/R division, QL expresses high levels (>20 transcripts per cell) of *lin-17* and *mab-5* and low levels (<10 transcripts per cell) of *mig-1*, *lin-39*, *mom-5* and *cam-1*; QR expresses intermediate levels (between 10-20 transcripts per cell) of *mig-1* and low levels of *lin-17*, *mom-5* and *cam-1*(Figure 3-4 A and B). Immediately after division, the transcript levels for each of the aforementioned genes in the Q daughter cells are about half of those in the mother cell. Later on, as the Q daughter cells continue to migrate, their gene expression patterns change as well. In particular, *mig-1*, *lin-17*, *mom-5* and *mab-5* levels rise over time in one or both of the QL descendants, whereas *lin-17*, *mom-5*, *cam-1* and *lin-39* increases in both QR descendants. The increase in transcript levels for both the receptor and the Hox genes indicates active transcription of these genes in the Q daughter cells. Thus, the Q daughter cells are not only inheriting the transcription profile of their mothers, but also selectively transcribing a subset of the Wnt receptor and Hox genes. It is also noteworthy that the transcriptional states of the daughter cells are not entirely identical to that of the mothers. For example, while *mig-1* level decreases over time in QL, it is upregulated in QL.p shortly post QL division (Figure 3-4 A). Furthermore, the transcriptional state also differ between pairs of sister cells. In fact, except for *cfz-2* which is absent in both Q lineages, the other four Wnt receptor genes (*mig-1*, *lin-17*, *mom-5*, *cam-1*) and both Hox genes (*mab-5* and *lin-39*) exhibit sister cell asymmetry in either the QL or QR lineage (Figure 3-4 and 3-5). Thus, two mechanisms together influence the transcription profiles of the Q daughter cells: the inheritance of transcripts from the mother cells and the activation of transcriptional regulatory programs specifically in the daughter cells.

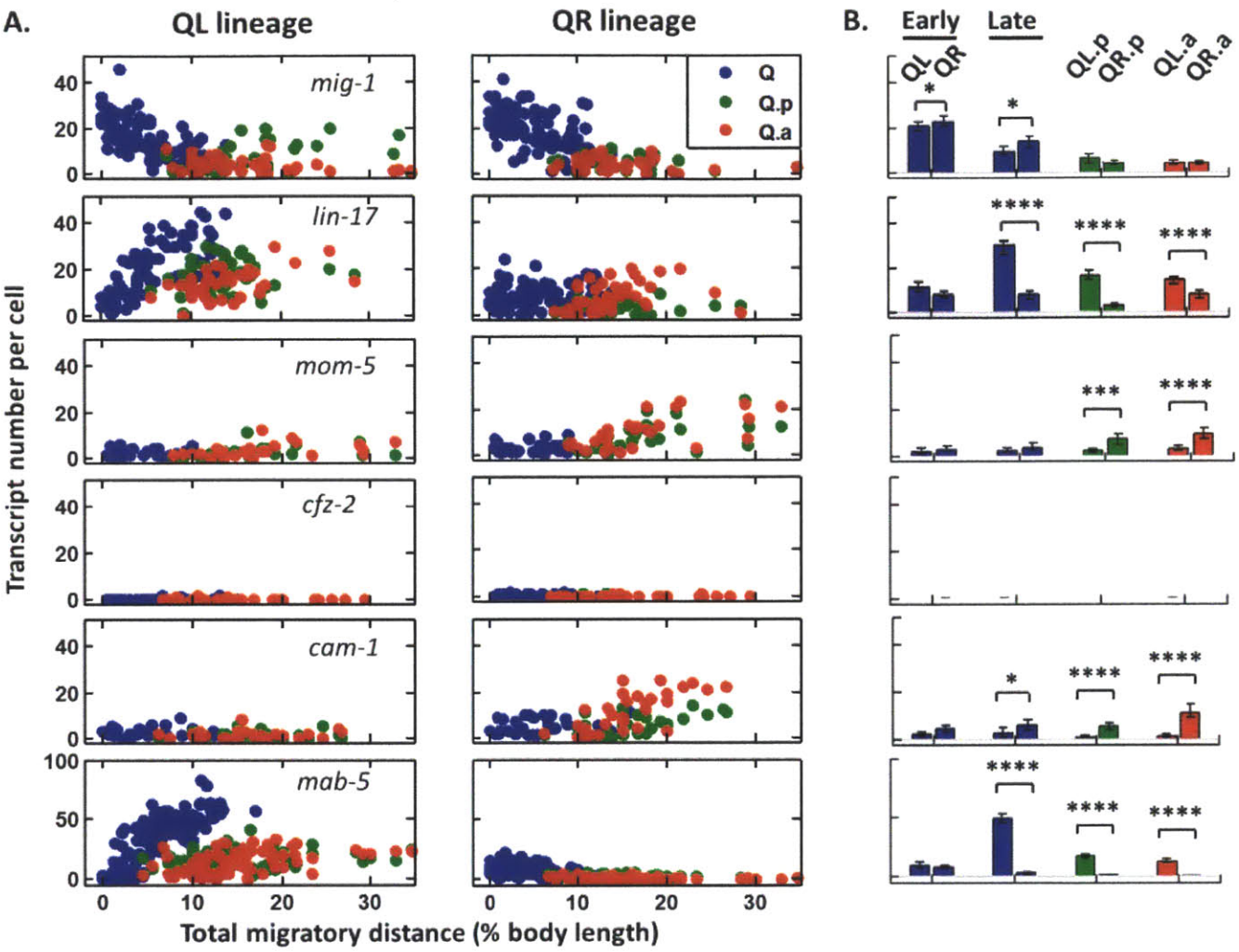



**Figure 3-4**: Transcription dynamics of the four Frizzled paralogs, *cam-1*/Ror and the Wnt target gene *mab-5*/Hox in the migrating Q neuroblasts. A) Single cell transcript counts acquired at different time points for QL (blue), QR (blue), and their daughter cells (green and red). As a proxy for time, we used the total distance migrated by cells of the QL and QR lineage as it is expected to increase monotonously with time. B) Comparison of transcript abundance between the QL and QR lineage, for the six genes mentioned above. Note distinct asymmetry in the transcript levels of *lin-17, mom-5, cam-1*, and *mab-5*. Errorbars are 95% confidence intervals of the mean. The asterisks denote *p*-values from the Mann-Whitney U test.

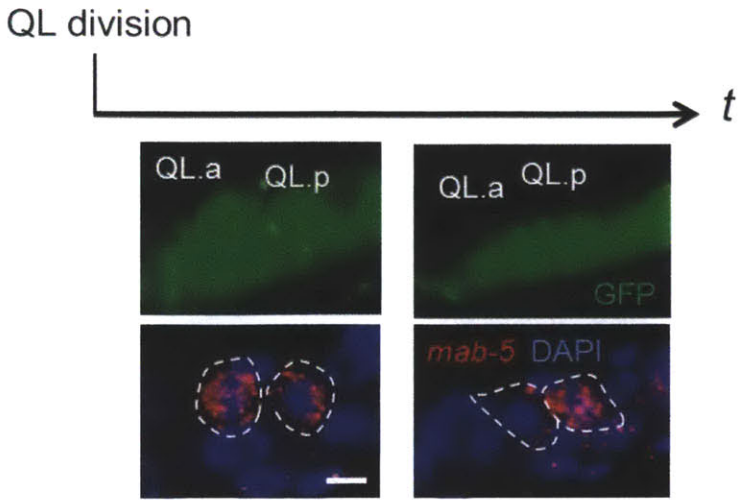



**Figure 3-5.** Downregulation of *mab-5* transcripts selectively in QL.a. The top arrow marks the passage of time (or increase in worm age) since QL division. (Middle) Membrane bound GFP (expressed from the transgene *Pegl-17::gfp*) demarcate the boundaries of the Q daughter cells. DAP. (Bottom) *mab-5* smFISH staining demonstrates the distribution of *mab-5* mRNA shortly after QL division (left) and the subsequent downregulation of *mab-5* mRNA in QL.a.

## 3.3 Loss or Restoration of Wnt signaling in QL Respectively Abolish or Rescue *mab-5* Expression and Posterior Migration in the QL Daughter Cells

To probe how Wnt signaling and *mab-5* expression in QL affects signaling and gene expression in the QL daughters (i.e. QL.a and QL.p), I first searched for known promoters that are active selectively in the Q mother cells but not their daughters. The *egl-17* promoter is known to be active specifically in the Q lineage and have been commonly used to drive transgene expression in the migrating Q neuroblasts (Blelloch et al., 1999; Ou and Vale, 2009; Middelkoop et al., 2012). In animals expressing GFP driven by the *egl-17* promoter (*egl-17p*), GFP fluorescence can be detected in both the Q mother cells and all their descendants (Ou and Vale,

2009). While this observation implys continuous activity of *egl-17p* throughout the Q lineage, it is likely confounded by the long life time of the GFP protein (Esposito et al., 2007). To directly assay the activity of the endogenous *egl-17* promoter, I used smFISH to stain the endogenous *egl-17* mRNA. Consistent with previous reports (Ou and Vale, 2009), *egl-17* mRNA is enriched in the Q lineages while present at very low levels elsewhere. However, unlike GFP protein expressed from an *egl-17* transcriptional reporter, endogenous *egl-17* mRNA is present at appreciable levels only in the Q mother cells (Figure 3-6 A). In both QL and QR, *egl-17* transcripts are present at around 20 copies/cell at the onset of Q cell migration, and drop to less than 5 copies/cell just prior to the 1$^{st}$ round of cell division. *egl-17* transcripts are further downregulated in the Q daughter cells to rapidly reach an average of less than 1 copy per cell (i.e. most observed Q daughter cells express no *egl-17* mRNA, Figure 3-6 A). The endogenous dynamics of *egl-17p* is closely recapitulated by a transgene driven by the same promoter. Using the same *egl-17* promoter (Ou and Vale, 2009) to drive the expression of *mig-1/Fz* cDNA in a *mig-1(n1787)* background, I observed the *mig-1* mRNA to be enriched mainly in the Q mother cells but rarely in the Q daughter cells (Figure 3-6 B). The lack of *egl-17p* (both endogenous and transgenic) activity in the Q descendants strongly indicates that *egl-17p* is a promoter active specifically in the Q mother cells.

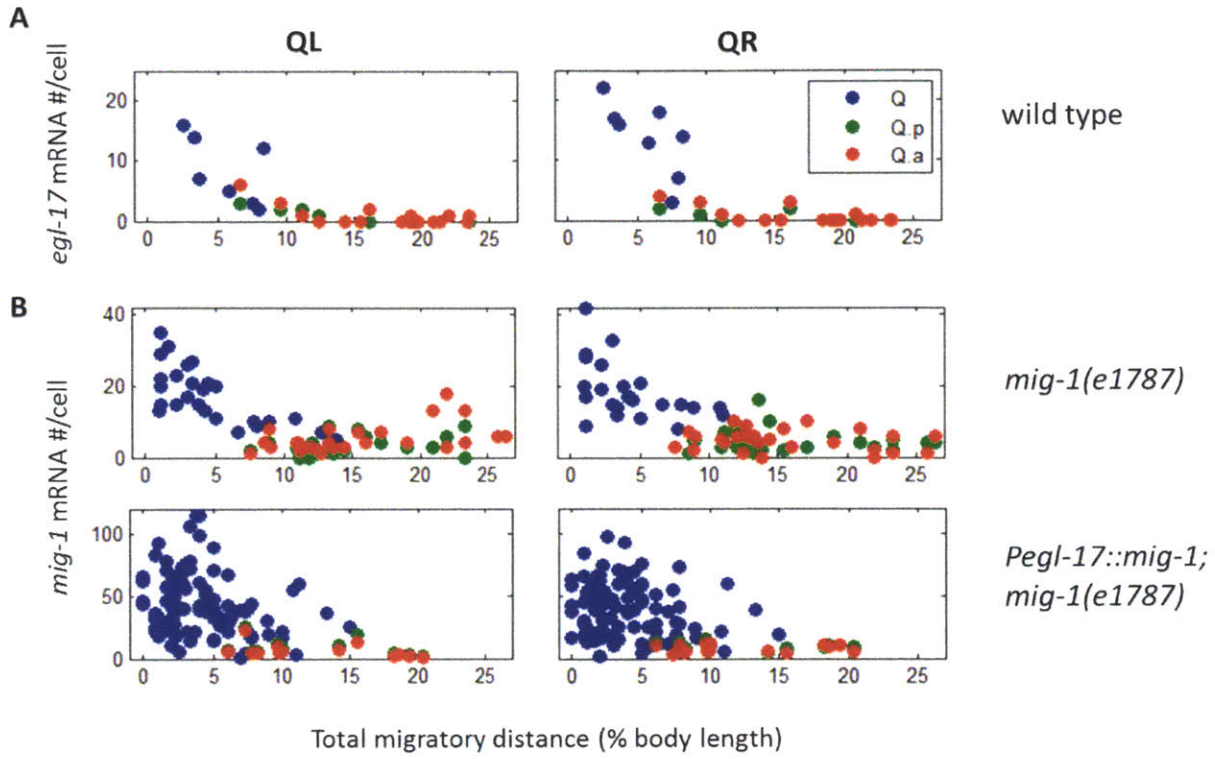



**Figure 3-6.** Endogenous and transgenic *egl-17* promoter activity in the left and right Q lineages. A) Endogenous *egl-17* transcript levels in the left and right Q lineages in wild type animals. B) Top panels: Endogenous *mig-1*transcript levels in the *mig-1(e1787)* mutants; bottom panels: Endogenous and transgenic *mig-1* transcript levels in *mig-1(e1787)* mutants carrying the *Pegl-17::mig-1cDNA* transgene The large variability in *mig-1* mRNA level is likely a result of the variable copy numbers of the extra-chromosomally located transgene.

Having established *egl-17p* as a promoter specifically active in the Q mother cells, I next utilized transgenes driven by this promoter to manipulate Wnt signaling specifically in the Q mother cells. First, I used a transgene expressing a dominant negative form of POP-1/TCF from the *egl-17* promoter (*Pegl-17:DNpop-1*, also used in Chapter 2). Expression of the dominant negative POP-1 disrupts Wnt signaling and leads to complete loss of *mab-5* transcripts in QL (Chapter 2, Figure 2-S1 D). Should the lifetime of the dominant negative POP-1 be relatively

short, Wnt signaling should ultimately resume in the QL daughter cells. However, low levels of *mab-5* transcripts (<10 copies per cell for both QL.p and QL.a, Figure 3-7) were consistently observed in both QL daughters. The inability of the QL daughters to resume *mab-5* transcription may be due to a combination of the following reasons: First, a substantial amount of dominant negative POP-1 protein was transferred from QL to its daughters, which continues to inhibit Wnt signaling in the QL daughters. Second, as described in Chapter 2 (Figure 2-4), loss of Wnt signaling and *mab-5* expression in QL would lead to a decrease in *lin-17/Fz* transcripts in QL. Since *lin-17* transcripts in QL is inherited by both QL daughters (Figure 3-4 A), the reduced *lin-17* level in the mother cell would lead to a decrease in *lin-17* level in the newly generated QL daughters (which was indeed observed, Figure 3-8). Reduced Frizzled receptor expression may thus prevent the QL daughters from effectively activating Wnt signaling and *mab-5* expression. Thirdly, without sufficient MAB-5 expression, the QL daughters would soon start migrating to the anterior. Migration away from the posteriorly localized source of the Wnt ligand, EGL-20, may also attenuate Wnt signaling in the QL daughters, further preventing *mab-5* from transcribing. While the first reason is attributed an experimental caveat, the other two point to a critical role of QL-specific *mab-5* expression in priming the daughter cells for robust Wnt signaling and posterior migration.

To further test whether Wnt signaling and *mab-5* expression in QL could promote the posterior migration of the QL daughters, I then attempted to restore Wnt signaling in QL in animals carrying loss-of-function mutation of the Frizzled receptor gene *mig-1*. *mab-5* expression is strongly reduced in both QL (Chapter 2, Figure 2-4) and the QL daughters (Figure 3-6) in the *mig-1(e1787)* mutants. Expression of the *mig-1* cDNA under the control of the *egl-17* promoter restores *mab-5* expression in both QL (data not shown) and the QL daughters (Figure

3-7). As shown in Figure 3-6 B, *mig-1* transcripts synthesized from the *Pegl-17::mig-1* transgene is specifically enriched in QL but not its daughters. The fact that transcription of *mig-1* only in QL is sufficient to rescue Wnt signaling and *mab-5* expression the QL daughters suggest that the activation of Wnt signaling and *mab-5* expression in QL can effectively promote Wnt signaling and *mab-5* expression in the QL daughters. Thus, while the posterior migration of QL does not depend on Wnt signaling or the expression of *mab-5* (Salser and Kenyon, 1992), activation of Wnt signaling and *mab-5* transcription early on in the mother cell may be a endogenous mechanism to ensure the robust posterior migration of the QL daughter cells.

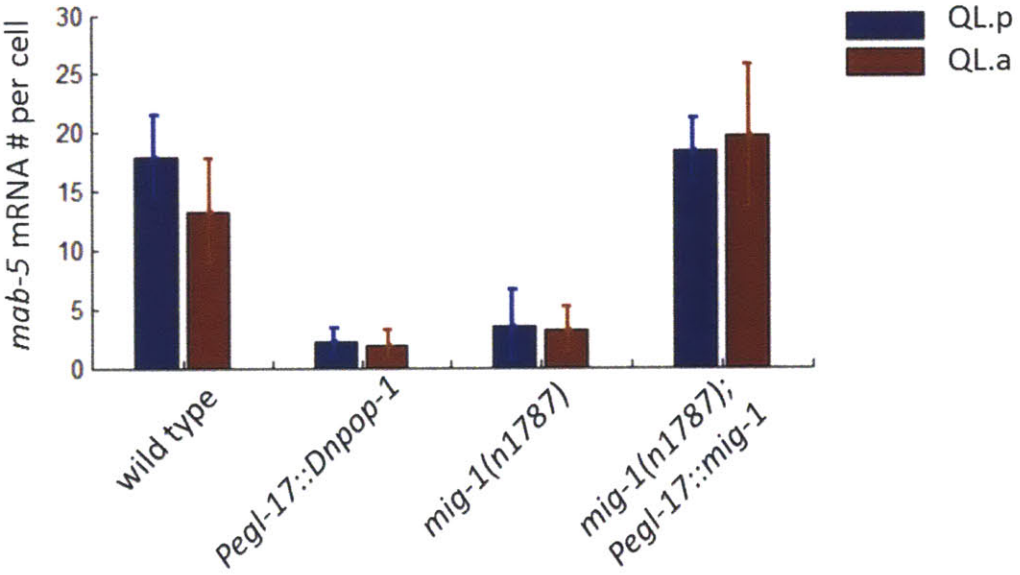



**Figure 3-7.** Loss or restoration of Wnt signaling in QL respectively abolishes or rescues *mab-5* expression in the QL daughters. Colored bars represent numbers of *mab-5* transcripts in QL.p (blue) and QL.a (brown). Genetic backgrounds of individual strains are indicated on the x-axis. Errorbars denote standard deviation of the mean.

## 3.4 *mab-5* Is Transcribed in the QL Daughters in a Position Dependent Manner

In Wnt signaling or *mab-5* mutants, the QL daughter cells start migrating towards the anterior shortly after their birth from cell division (Harris et al., 1996; Maloof et al., 1999). Since anterior migration quickly distances the Q daughter cell from the source of the EGL-20 ligand, there may be a critical time window for MAB-5 expression to reach a sufficient level to initiate posterior migration. In other words, sustained Wnt signaling and *mab-5* expression may critically depend on the posterior location of the QL daughters. To test this hypothesis, I examined the dynamics of *mab-5* transcription in the QL daughter cells in the *mab-5(e1239); smg-1(e1228)* mutants. As described in Chapter 2, this mutant fails to generate functional MAB-5 protein but spares the mutant *mab-5* transcripts from nonsense mediated transcript degradation. In this mutant, the QL daughters are fated to migrate to the anterior due to the lack of functional MAB-5. They however should in principle remain capable of Wnt signaling as no Wnt signaling components are directly perturbed. If the expression of *mab-5* in the QL daughters does not require the cells to remain in the posterior, continued *mab-5* transcription should be observed in QL daughter cells located anterior to their wild type positions. This was however not the case. As shown in Figure 3-8, while *mab-5* transcript levels continued to increase in newly generated QL daughters (lower panel, data points corresponding to $X > 0$), transcript levels began to decrease as QL daughters migrate anteriorly past their birth positions (marked by red and green arrows; note the initial posterior-ward displacement is likely not a result of active migration, but a result of changes in cell morphology, i.e. rounded to elongated, after cell division). In QL.p especially, there is a strong dependence of *mab-5* transcript levels on cell position (Pearson's R=-0.79), compared to lack of significant correlation in the wild type (Pearson's R=-0.01; note that QL.p

exhibits little active migration in the wild type). Taken together, the observation that *mab-5* transcription in the QL daughters depend strongly on cell position supports the model where *mab-5* activation early in the QL stage serves to ensure sufficient MAB-5 expression in newly divided QL daughters.

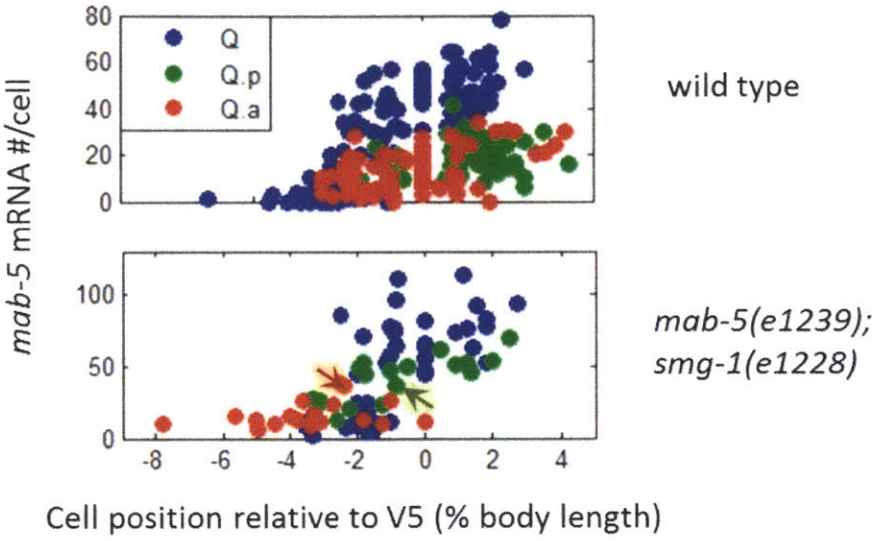



**Figure 3-8.** *mab-5* transcripts in QL.p exhibit strong position dependence in QL daughter cells in the *mab-5 (e1239)* mutant. (Top) *mab-5* transcript levels in wild type QL and QL daughter cells plotted against cell position relative to the stationary cell V5. (Bottom) mab-5 transcript levels in the left Q neuroblasts in a *mab-5(e1239);smg-1(e1228)* mutant background. Shaded arrows indicate typical positions at which newly generated QL .a (red arrow) and QL.p (green arrow) are found. Zero on the x-axis marks the position of the stationary cell V5. All positions are normalized to the body lengths of the animals. Left is anterior and right is posterior.

## 3.5 Multiple Genetic Interactions Found in the Q Mother Cells Are Shared by the Q Daughter Cells

As both similar and distinct transcriptional regulatory programs exist among the Q mother cells and their daughters (see results in section 3.2), I next questioned whether regulatory

interactions found in the Q mother cells continue to be active in the Q daughter cells. To address this question, I measured the transcript levels of the four Wnt receptor and two Hox genes in QL, QR and their daughters in both the wild type and strains harboring single mutations of the six genes (i.e. *mig-1/Fz, lin-17/Fz, mom-5/Fz, mab-5/Hox, lin-39/Hox, and cam-1/Ror*).

Figure 3-9 summarizes the transcriptional profiles for the above six genes in various Wnt signaling mutants. The expression level of each gene is normalized to its wild type level, and any significant increase (red) or decrease (blue) from the wild type is colored accordingly. Across the six cell types, differential expression in the above six genes was mainly observed in the left Q lineage. This is consistent with the observation that canonical Wnt signaling is strongly activated in QL but is inactive or only weakly active in QR (Korswagen, 2002). Perturbations that weaken the canonical Wnt signal, such as *mig-1(e1787)* and *lin-17(n671)* may have little effect on the level of canonical Wnt signaling in QR. Meanwhile, mutation in *mab-5*, a canonical Wnt signaling target already lowly expressed in QR, would have relatively little effect on downstream genes. Thus, the difference between the left and right Q lineages in their transcriptional response can be explained by their asymmetrical Wnt signaling level.

Within the QL lineage, many of the transcriptional changes found in QL were also detected in the QL daughters (Figure 3-9). A number of these changes appear more prominent in the daughter cells. One example is the increase of *lin-39* and *cam-1* expression following mutations in *mig-1*, *mab-5* or *egl-20*. This increase in transcriptional response may reflect a change in regulatory environment in the daughter cells. For example, MAB-5 and/or Wnt signaling may assume a dominant role in regulating some of their targets (e.g. *lin-39* and *cam-1*) in the QL daughters, whereas additional regulators may act in parallel to MAB-5 and Wnt signaling in the QL cell. Similar explanation can be applied to a number of changes observed only in QL or only

in its daughters (highlighted in yellow, Figure 3-9). For example, the invariant expression of *mab-5* in the *cam-1(gm122)* mutant may be attributed to the negative feedback of *mab-5* on its own transcription (see Chapter 2). The strong negative auto-regulation of *mab-5* may have effectively attenuated the effect of *cam-1(gm122)* mutation on *mab-5* transcription.

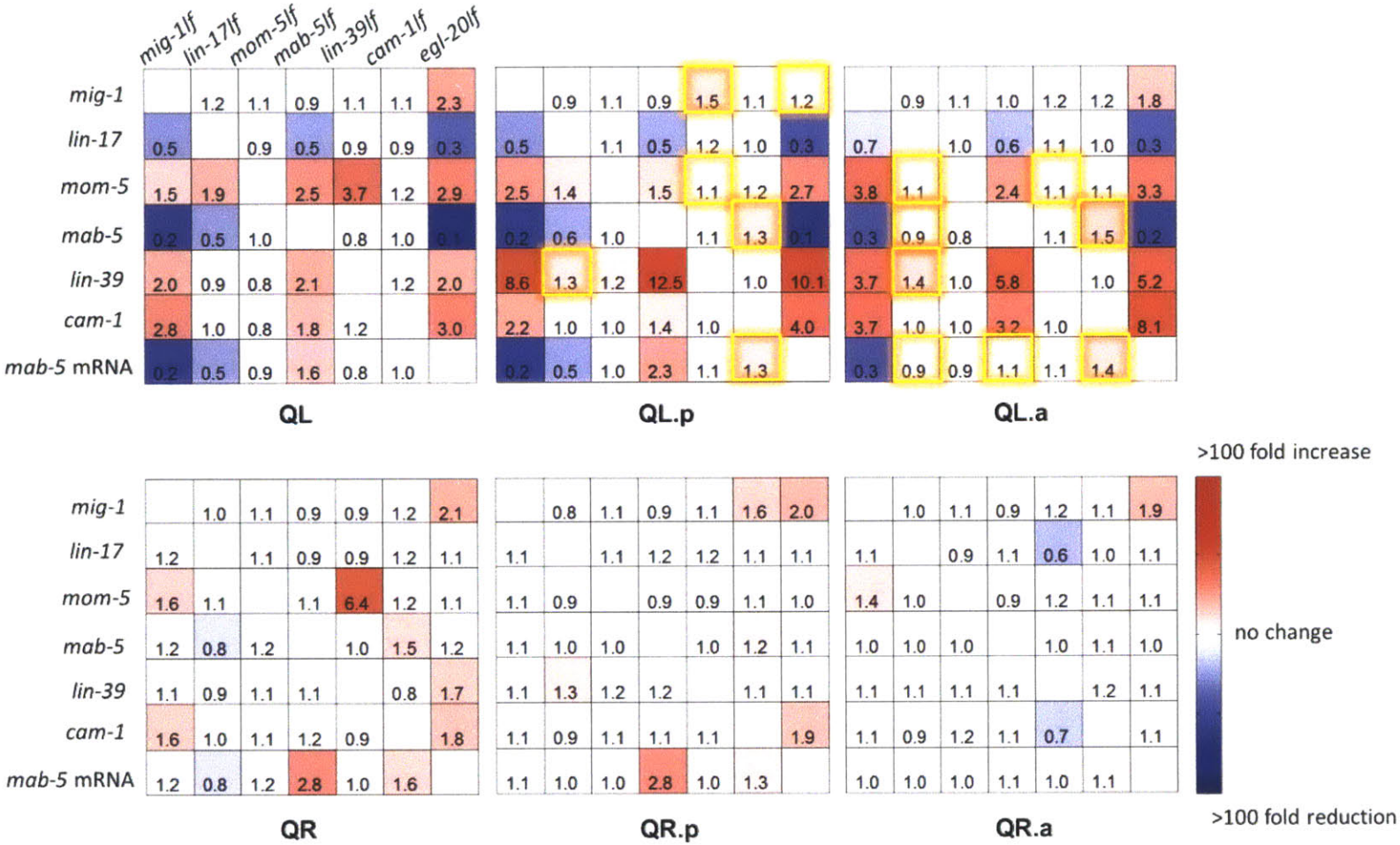



**Figure 3-9**: Summary of differential gene expression in various Wnt pathway mutants compared to the wild type. Each colored matrix correspond to the transcription profile of one cell type (e.g. QL, QL.a, QR). Each small square inside the block correspond to the abundance of a particular mRNA species (as labeled to the left) in a specific mutant background (as labeled above). The color of the square corresponds to the fold change in transcript abundance compared to the wild type. The color bar to the far right details the correspondence between color intensity and fold change values. Yellow squares denote differential gene expressions found only in the QL daughters, but not the QL mother cell.

## 3.6 Conclusions

In this Chapter, we report the expression pattern of various Wnt receptor and target genes in the Q daughter cells. Comparing the expression profiles between mother and daughter cells, I reveal a role of the mother cells in influencing the transcriptional profiles of the daughter cells. By assaying the transcriptional changes in the daughter cells in various Wnt pathway mutants, I show that some but not all of the genetic interactions observed in the mother cells were also present in the daughter cells. Additionally, new transcriptional regulatory programs become active in the Q daughters giving rise to distinct and asymmetric expression profiles in pairs of Q daughter cells. To the regulatory program that controls Wnt signaling activity and gene expression in the Q lineage, there is both history-dependence but also dynamic cell-specific modulation. A lesson from these findings is that controlling of gene expression in a developing multicellular tissue may be complicated by both the transcriptional history of the cell lineage as well as the dynamically evolving cellular context.

## 3.7 Experimental Procedures

### *C. elegans* strains and culture

*C. elegans* strains were grown at 20°C using standard culture conditions and techniques. The Bristol N2 strain was used as wild type. Mutant alleles and transgenes used in the study are:

Linkage Group I (LGI*): mig-1(e1787), lin-17(n671), smg-1(e1228), mom-5(gk812), mom-5(or57), mom-5(ne12);*

LGII: *cam-1(gm122);*

LGIII: *mab-5(e1239), lin-39(n1760);*

LGIV: *egl-20(n585);*

LGV: *heIs63[Pwrt-2::gfp-ph; Pwrt-2::h2b-gfp; Plin-48::tomato]*;

Extra-chromosomal array: *huEx278[Pegl-17::DNpop-1;Pmyo-2::tomato]*.

**Scoring QL descendent migration**

The precise positions of the Q descendants QL.pap/QL.paa were scored by DIC microscopy in late L1 stage larvae as described (Coudreuse et al., 2006).

**Single molecule fluorescence in situ hybridization**

SmFISH was performed as described (Raj et al. 2008). Manual segmentation of GFP-marked QL periphery was performed, followed by automated spot counting in MATLAB-based custom-written software. Total migratory distance (MD) was assayed by manually marking the nuclear position of QL and QR, tracing the A-P axis of the worm, and automatically computing the distance between QL and QR along the A-P axis.

**Statistical analysis**

The Mann-Whitney test was used to compare average gene expression levels between the wild type and the mutant strains. To control for multiple comparisons, the Benjamini-Hochberg procedure was used to achieve an FDR of less than 0.04 for comparison of transcript abundance. Corrected p-value of less than 0.05 was considered significant.

# Chapter 4

# EPILOGUE

## 4.1 Summary of Results

The work presented in this thesis stems from a central question: How is noise in gene expression handled by endogenous biological systems? To address this question, we focused on a stereotyped event in *C. elegans* development, the migration of the Q neuroblast. Building on previous study showing that the canonical Wnt pathway drives the posterior migration of the QL neuroblast by activating the Hox gene *mab-5*, we set out to uncover the mechanisms that ensure robust activation of *mab-5* in QL. We first quantified the level of *mab-5* transcripts in QL in the wild type. Unexpectedly, not only is *mab-5* consistently expressed at a high level across animals, but its steady state expression levels are tightly distributed within a narrow range. In comparison, multiple Wnt signaling mutants exhibited markedly more variable *mab-5* expression levels, reminiscent of their partially penetrant migratory phenotype,.

As the degree of transcriptional variability cannot be simply predicted by mean transcript level, we embarked on a search for the mechanistic cause of the increased variability. Building on the observation that the transcript levels of the three Frizzled receptors, *mig-1*, *lin-17*, and *mom-5*, are dynamically modulated during QL migration, we asked whether their levels are additionally affected by perturbation of the Wnt signaling pathway. By systematically profiling the levels of Frizzled receptor and *mab-5* expression in a series of Wnt pathway mutants, we found extensive feedback interaction between the Frizzled receptors, *mab-5* and the intracellular Wnt signal. Using computational network inference, we obtained a single most probable network

topology, which we then used to construct a mathematical model. Using theoretical approaches, we demonstrated a link between alterations in network topology and the observed increase in gene expression variability.

Extending our gene expression analysis to the Q daughter cells, we found that mRNA transcripts present at the point of Q cell division are effectively transferred to the daughter cells in a largely symmetric fashion. Thereafter, the Q daughter cells turn on transcriptional regulatory programs not entirely identical to their mothers. It is clear, however, that the transcription profiles and the regulatory interactions in the mother cells prominently influence the transcriptional profiles of the daughter cells. When a perturbation to the Wnt pathway alters the transcript abundances in the mother cells, similar changes tend to be observed in the daughter cells as well. These findings suggest that lineage history and cellular context are important factors that influence the transcription profile of a given cell.

Together, our results highlight the role of regulatory network topology in controlling the variability in gene expression. These results also provide some initial evidence that network topology may indeed be an adaptive trait employed endogenously to ensure the robust control of gene expression.

## 4.2 Implications and Future Directions

Controlling gene expression in a multicellular organism is a challenging task. Compared to single-cell organisms, the survival of a multicellular organism requires careful orchestration between many cells of different types and with different needs (Komili and Silver, 2008). In the system studied here, the genetic program that controls *mab-5* expression may have to ensure not only a sufficient level of *mab-5* expression for posterior migration, but also an accurate level of

*mab-5* expression to drive particular cell fates. The existence of multiple regulatory goals may explain why *mab-5* transcription is not simply activated to the max, but rather precisely kept within a narrow range. Since simple network motifs, such as single positive or negative feedback loops, may be insufficient to simultaneously satisfy multiple regulatory goals, more complex networks may have evolved to serve such purpose. Understanding the regulatory goals of a network may thus help us decipher design principles of complex biological networks.

The work in this thesis also highlights the emergent functions of complex networks. Many genome-wide network inference studies have revealed highly complex "interactomes" at essentially all levels of a living organism (Esposito et al., 2007). Understanding the complex interactions among genes, proteins, and cells is crucial both to understanding the fundamental principles of life, but also to develop ways to improve physiological function and battle disease. To understand the structure-function relationship of biological networks, many researchers have chosen to take the "modular" approach, namely to work with small sub-network, or motifs, that constitutes building blocks of larger networks (Shen-Orr et al., 2002; Doyle and Csete, 2005; Carlson et al., 2006; Davidson, 2010). The question then is whether larger networks are simply the sum of small sub-networks or possess emergent functions not attainable in the sub-networks. The work in Chapter 2 presents one such example where positive feedback and negative feedback, two well-characterized network motifs, synergize to more efficiently reduce variability in gene expression than negative feedback alone. There are likely more examples of this nature in the complex endogenous regulatory networks within living organisms. Discovering and dissecting endogenous regulatory networks may thus prove a fruitful avenue for generating new insights into the structure-function relationship of complex networks.

# APPENDIX

# Protocol of smFISH Optimized for Study in *C. elegans*

## Introduction

In *C. elegans*, the expression pattern of a gene provides important clues to understanding its biological function. To accurately depict endogenous transcriptional activity, a highly sensitive method is required to measure transcript levels in the intact tissue across various developmental stages. Conventional RNA in situ hybridization methods using hapten- (biotin or digoxygenin) labeled RNA probes rely on antibody binding for visualization, and are thus only semi-quantitative at best (Raap et al. 1995, Levsky et al. 2003). Additionally, hapten-labeled probes are prone to diffuse localization (when conjugated with alkaline phosphatase), low sensitivity (when conjugated with fluorescent molecules), and non-specific probe binding. Here, we introduce a recently developed mRNA *in situ* hybridization method (Raj et al. 2008) that circumvents the above difficulties to give single molecule resolution of transcript detection.

The single molecule fluorescent *in situ* hybridization (smFISH) method differs from conventional approaches by using many short (about 20 base pairs long) oligonucleotide probes to target different regions of the same mRNA transcript (Raj et al. 2008, Femino et al. 1998). Each oligo is conjugated with only one fluorophore and thus faintly visible by itself. Binding of multiple oligos to the same transcript yields a bright spot, indicative of a single mRNA transcript. Since mis-bound probes are unlikely to co-localize, this method effectively reduces false-positive signal from non-specific probe binding. The small oligo size allows the probes to efficiently penetrate through target tissue, yielding robust detection of even lowly abundant transcripts.

Subsequently, the total number of fluorescent spots within a single cell or region can be unambiguously counted and compared across different developmental stages and genetic backgrounds.

Given its many advantages, the smFISH method is a powerful tool to study transcriptional regulation. Its high sensitivity allows accurate characterization of the spatio-temporal patterns of endogenous gene expression. Its single-molecule resolution enables precise quantification of gene expression levels. Such quantitative information can in turn be used to assess, for example: 1) tissue-specific correlations in gene expression, 2) similarity and difference in gene expression across strains, 3) variability in gene expression, 4) tissue-specific signaling dynamics. To date, smFISH has been successfully applied to study a variety of questions in *C. elegans* biology (Raj et al. 2010, Harterink et al. 2011, Korzelius et al. 2011, Middelkoop et al. 2012, Saffer et al. 2011, Seidel et al. 2011, Topalidou et al. 2011). Here we provide the necessary technical details to set up and perform smFISH, from sample preparation to data analysis.

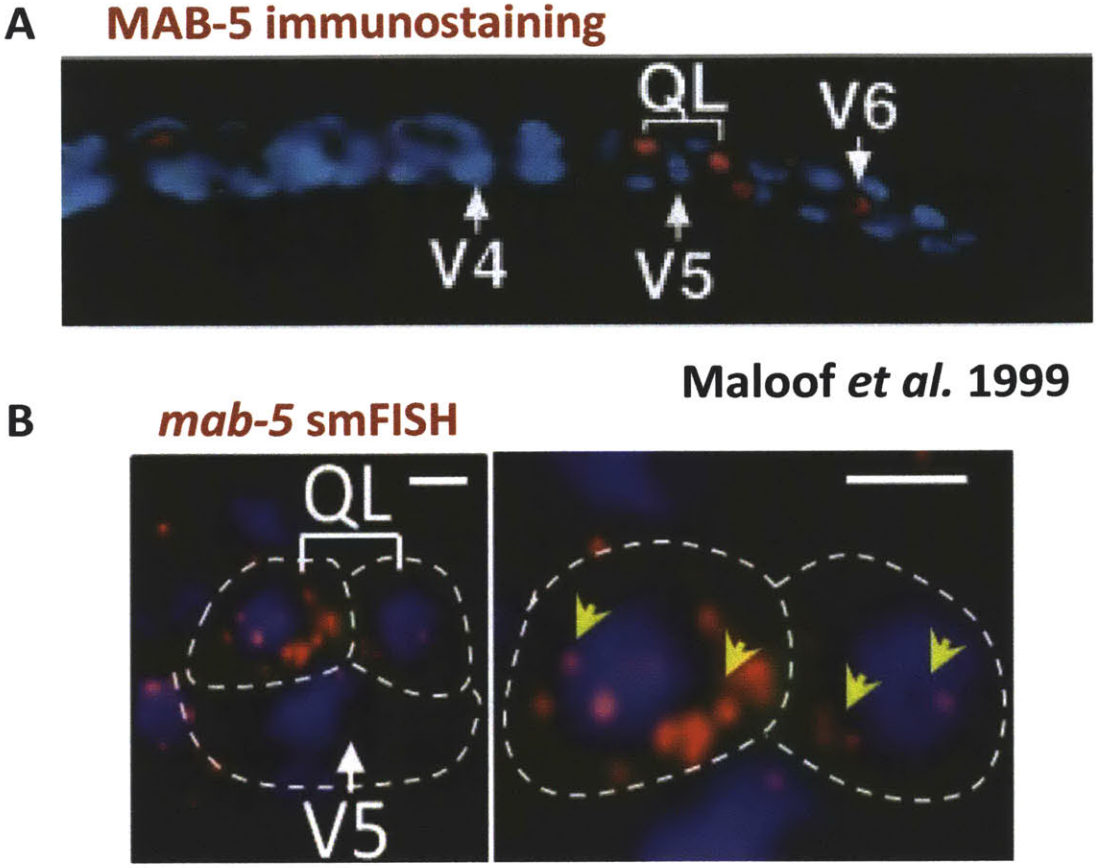



**Figure A-1**. Comparison of immunohistochemistry and smFISH results detecting *mab-5* expression in the QL neuroblasts. A. (Adapted from Maloof et al. 1999) MAB-5 immunostaining of a wild-type L1 larva. MAB-5 expression is seen in the two QL daughters above V5. B. *mab-5* mRNA detected by smFISH in L1 larva around the same age. Note the asymmetry in *mab-5* mRNA abundance between the two QL daughters.

## Protocols

The following protocol covers the 5 major steps of smFISH: **A.** Probe design and synthesis; **B.** Fixation of *C. elegans* worms and embryos; **C.** Hybridization; **D.** Image acquisition; **E.** Data analysis. This protocol is largely adapted from the general smFISH protocol detailed in Raj and Tyagi (2010), with notes and modifications specific to *C. elegans*. Unless otherwise noted, all reagents listed can be made in bulk ahead of time and stored at room temperature (RT).

## A. Probe design and synthesis

### 1. Design.

smFISH probes are DNA oligonucleotide probes designed according to the follow criteria:

1. Probe length: 17-22 base pairs
2. Probe spacing: no less than 2 base pairs
3. Probe GC content: ideally around 45% to ensure uniform binding efficiency
4. Number of probes: ideally between 30 and 96 depending on target transcript length. Typically 48 probes are used to ensure good signal quality, while as few as 20 have yielded satisfactory signal (e.g. Figure A-1 illustrates the result from 27 probes designed to target *mab-5* mRNA). To increase effective target length, part or all of the 3'UTR sequences can be combined with the cDNA sequence for probe design.

A web-based probe design software developed by Raj et al. (2008) is available free of charge at: http://www.biosearchtech.com/stellarisdesigner/. By supplying the target RNA sequence and the above criteria, one obtains a list of probe sequences with optimized GC contents. These sequences can then be directly submitted for probe synthesis.

### 2. Synthesis, coupling and purification.

**Synthesis:** Based on the designed oligonucleotide sequences, smFISH probes are generated en masse using a 96-well DNA synthesizer. Biosearch Technologies (Novato, CA) offers synthesizing service for custom-designed smFISH probes. The probes can be ordered in both

coupled and uncoupled forms. While ordering coupled probes minimizes the work involved in probe preparation, obtaining uncoupled probes allows the flexibility to choose fluorophores that are optimally compatible with the researcher's optical set-up and other experimental needs. Furthermore, coupled probes are delivered as a mix, which precludes the option of using subsets of the probe library to 1) selectively target different parts of the transcript, or 2) perform negative controls using deletion mutants (see additional comment in the following section and details of suggested controls in the Controls and Troubleshooting section). In the following section, we further suggest a list of factors to consider when deciding whether to pursue in-house coupling.

Given the small amount of probes needed for each hybridization experiment, synthesis can be carried out at a small scale (Biosearch now delivers 5nmol of each probe per custom order. This amount is typically sufficient for hundreds of experiments). When ordering uncoupled probes, the oligonucleotide probes should be synthesized with a 3'amine group to allow subsequent oligo-fluorophore conjugation. Additionally the oligos should be desalted and resuspended in water as opposed to Tris EDTA (TE).

**Coupling:**

** Note: When deciding whether to perform in-house coupling or order readily coupled probes, we suggest considering the following: 1) Is it clear which fluorophore to choose? This could vary depending on probe set and the optical setup. 2) Is it desirable to have the same probes coupled to multiple fluorophores? This is of concern when performing multiplex assays among three or more genes. 3) Is it necessary to split the probe sets to target different parts of a transcript? Pre-coupling require mixing of all probes in a library, thus excluding the possibility of alternative probe combinations. 4) Is in-house coupling feasible? Is there access to a nearby HPLC facility? 5) Does*

*experimental schedule allow an additional 3-4 days of probe preparation? 6) Is budget a concern? A single probe set can be coupled to multiple fluorophores (up to 3-4 coupling per probe set is possible), avoiding the need to purchase multiple libraries.*

Prior to coupling, one needs to decide on the desired type of fluorophore. The current protocol uses succinimidyl ester derivatives to couple to the 3' end of the oligonucleotide probes. Three types of commercially available fluorophores are commonly used: Cy5 (GE Amersham), Alexa 594 (Molecular Probes, Invitrogen), and tetramethylrhodamine (TMR) (Molecular Probes, Invitrogen). Biosearch also offers equivalent fluorophore (Quasar 670, and CAL Fluor 610 and 590) for their couple probe sets. In general we have found that using fluorophores with shorter emission wavelengths (such as Alexa 488) do not yield reliable signal due to high cellular autofluorescence. Table A-1 summarizes the strengths and weaknesses of each of the fluorophores as observed in our hands.

**Table A-1.** Strengths and weaknesses of fluorophores commonly used in smFISH.

| | **Peak Excitation/Emission Wavelength** | **Photostability** | **Autofluorescence level** | **RFP compatible** | **GFP compatible** |
|---|---|---|---|---|---|
| **Cy5** | 649/670 nm | Low | Low | Yes | Yes |
| **Alexa 594** | 590/617 nm | High | Medium | Affected | Yes |
| **TMR** | 564/570 nm | High | High | No | Yes |

For multiplexing assays, we have obtained good results with combinations of all three dyes (provided that each probe set works well when assayed on its own). We generally prefer to

pair Cy5 with Alexa 594, as the two fluorophores tend to give better signal quality than TMR. For initial testing, it is important to compare results from multiplex and single-fluorophore assays to ensure consistency. Also keep an eye out for "suspiciously similar" spot patterns, which is indicative of cross-talk between fluorescent channels.

Reagents for coupling:

DMSO (if coupling to TMR)

*0.1 M Sodium Bicarbonate (in RNase free water, pH 8.0)

*1 M Sodium Bicarbonate (in RNase free water, pH 8.0)

Ethanol (>95% pure)

3 M Sodium Acetate, pH 5.2

Fluorophore with succinimidyl ester group

* Note it is desirable to make sodium bicarbonate fresh, or check before use to make sure the pH level is correct.

Day 1:

1. From the uncoupled probe stock, combine 1 nmol (e.g. 10 ul from a 100 μM stock) of each probe into a single microcentrifuge tube.
2. Add 0.11 volumes 1 M sodium bicarbonate to give a final concentration of 0.1 M. If the total volume at this stage is < 0.3 mL, add some 0.1 M sodium bicarbonate to bring total volume to 0.3 mL.
3. Dissolve a small amount (roughly 0.2 mg) of dye into 50 μl of 0.1 M sodium bicarbonate.

a. Note: TMR can be hard to dissolve in aqueous solution, so one must first dissolve it in a small volume (<5 µl) of DMSO before adding 50 µl of 0.1 M sodium bicarbonate.

4. Add the dissolved fluorophore to the oligos.
5. Cover the tube in foil to prevent photo-bleaching and let the reaction proceed overnight at room temperature with gentle rocking.

Day 2:

6. In the morning, precipitate the oligos by adding 10% volume/volume of 3M sodium acetate and then adding 2.5 volumes of 100% EtOH. Store at –70°C for at least 1 hour (up to overnight).
7. Spin down the sample in a 4°C microcentrifuge for at least 15 minutes at maximum speed (~16K RCF).
8. After centrifugation, one should see a small colored pellet at the bottom of the tube. Aspirate away the fluorescent supernatant (containing uncoupled dye molecules) as completely as possible. If purification is not performed right away, the pellet is stable and can be stored at -20°C.

**Purification:** The pellet obtained from the coupling steps contains a mixture of coupled and uncoupled oligonucleotides. To separate the two species, we take advantage of the fact that coupled probes experience a large increase in hydrophobicity compared to the uncoupled ones. High-pressure liquid chromatography (HPLC) can thus be used to enrich for coupled probes.

Reagents and equipments for purification:

0.1 M Triethylammonium acetate, pH 6.5, filtered and degassed (Buffer A)

Acetonitrile for HPLC (Buffer B)

HPLC with a dual wavelength detector to measure both DNA and fluorophore absorbtion

C18 Column for HPLC, 218TP104

Speedvac rated for acetonitrile

Day 2 (if continued immediately after coupling step 8):

1. Resuspend pellet in appropriate volume (0.1→0.5 mL nuclease free water, depending on your HPLC)
2. Inject coupled probe into HPLC and run a program in which the percentage of buffer B varies from 7% to 30% over the course of around 30 minutes with a flow rate of 1 mL/minute.
    a. Note: Set the detector to monitor DNA absorption (260 nm) and the absorption of the coupled fluorophore (e.g., 555 nm for TMR).
3. One will observe two broad peaks. The first contains uncoupled probes and will only show a peak in the 260 nm channel. The second contains pure coupled probes and will show peaks in both channels. The two peaks will typically be separated by a few minutes of time or longer, with TMR having narrowly separated peaks and Cy5 having broadly separated peaks. With a series of microcentrifuge tubes, collect the entire second peak as soon as the signal begins rise and until it drops back to baseline.

Day 3:

4. Dry the purified probes in a speedvac (~ 3-5 hours for 0.5 mL).

a. Note: Be sure to prevent any light from hitting the probes during the drying process to prevent photobleaching, especially for photolabile dyes such as Cy5.

5. Resuspend all tubes together in a total of 50-100 µl of TE, pH 8.0 (equivalent to roughly 0.1-1 µM). This is the concentrated probe stock.
6. (Optional) Dilute this probe 1:10, 1:20, 1:50 and 1:100 in TE to make working stocks for testing probe concentration.

At this point, probe synthesis is complete. Probes can be stored in TE at -20°C for years.

## B. Fixation of *C. elegans* worms and embryos

Reagents for fixation:

1. Fixation solution: 4% paraformaldehyde (PFA) in 1x PBS
2. Bleaching solution for embryos (per 40 mL, store at 4°C):

   40 mL distilled water

   7.2 mL 5 N NaOH

   4.5 mL 6% NaHOCl
3. M9 buffer (per 1L):

   5.8 g $Na_2HPO4$

   3.0 g $KH_2PO4$

   0.5 g NaCl

   1.0 g $NH_4Cl$

   Dissolve in distilled water ($dH_2O$) to 1 L final volume

Fixation protocol for worms (larvae and adult):

1. Grow worms on plates seeded with OP50. Synchronize worms to the desired stage if needed.
    a. Note: The appropriate number of plates may vary depending on need, with the minimum requirements that 1) there will be enough worms to form a visible pellet in a microcentrifuge tube, 2) each plate is not too crowded.
2. Add 3-5 mL M9 buffer to the plate and swirl to release worms from surface, then transfer worms to a 15 mL conical centrifuge tube.
    a. Note: Distilled or deionized water may be used instead of M9 in this and subsequent steps.
3. Spin down to form a pellet and aspirate the supernatant.
4. Wash and spin down with 3-5 mL M9 buffer to rinse away bacteria and other debris.
5. Resuspend in 1 mL fixation solution and transfer to microcentrifuge tube. Keep rotating at room temperature for 45 min.
    a. Note: Since autofluorescence levels increase with fixation time, incubation in fixation solution should be kept short. This is especially important for older worms where tissue autofluorescence is relatively strong.
6. Wash twice with 1 mL 1x PBS.
7. Resuspend in 1 mL of 70% EtOH. Keep rotating for overnight (or longer) at 4°C. Store at 4°C for up to a month.

<u>Fixation protocol for embryos:</u>

1. Grow worms on OP50 plates till there are plenty of gravid hermaphrodites.
2. Wash worms off the plates with M9 into a 15 mL Falcon tube.

a. Note: Distilled or deionized water may be used instead of M9 in this and subsequent steps.

3. Spin down and resuspend in bleaching solution. Vortex or shake rigorously for 4-8 minutes until worms disappear and only embryos remain.
4. Spin down and aspirate. Then wash twice with M9 buffer.
5. Resuspend in 1 mL fixation solution and transfer to a microcentrifuge tube. Keep rotating at room temperature for 15 minutes.
6. Vortex and then immediately submerge tube in liquid nitrogen for 1 minute to freeze crack the eggshells.
7. Thaw in water at room temperature.
8. Once thawed, vortex and place on ice for 20 minutes.
9. Wash twice with 1 mL of 1x PBS.
10. Resuspend in 1 mL of 70% EtOH and keep rotating for overnight or longer at 4°C. Embryos can be hybridized up to a week following fixation.

## C. Hybridization protocol

Reagents for hybridization:

1. Hybridization buffer (per 10mL, store at -20°C freezer):

   *1 g dextran sulfate

   10 mg Escherichia coli tRNA

   100 µl 200 mM vanadyl ribonucleoside complex (NEB)

   40 µl 50 mg/mL RNase free BSA (Ambion)

   **Formamide (deionized, Ambion)

Add nuclease free (NF) water (Ambion) to 10 mL final volume

* Note: Dextran sulfate is viscous and hard to dissolve at room temperature. One solution is to mix dextran sulfate with 4 mL NF water, then sonicate till no clumps are visible.

** Note: Determine the amount based on the desired formamide concentration, e.g. 1 mL for 10% final concentration. Higher concentration yields higher probe binding stringency.

2. Wash buffer (per 50 mL):

   Formamide (deionized, Ambion), use same concentration as determined for hybridization buffer

   5 mL 20x SSC (RNase free, Ambion)

   Add NF water (Ambion) to 50 mL final volume

3. DAPI stain: Prepare working stock of 5 ng/mL in RNase free water, store in the freezer.
4. 2x SSC: Prepare from 20x SSC (RNase free, Ambion) in RNase free water (Ambion).
5. Antifade buffer (to be made fresh)
   1. Glox buffer (per 50 mL):

      2 mL 10% glucose in NF water

      250 µl 2M Tris-HCl, pH 8.0

      5 mL 20x SSC (RNase free, Ambion)

   2. Glycosidase stock: Dilute commercial Glox (Sigma) in 50 mM sodium acetate to working stock of 3.7 mg/mL, adjust to pH 5.0. Aliquot to 100 µl units and store at -20°C.
   3. Catalase (Sigma): store at 4°C.

Prior to imaging, mix 1 µl each of the glycosidase stock and catalase to 100 µl of Glox buffer. Each 100 µl antifade mix is enough for 1-2 hybridization samples.

Day 1:

1. Prepare the hybridization solution: to 100 µl of hybridization buffer, add 1 µl of each probe at the appropriate concentration, then vortex and centrifuge.
    a. Note: For the initial test of probe concentration, one may perform four parallel hybridizations with probes diluted at 1:10, 1:20, 1:50 and 1:100 in TE. In our hands, 1:20 works well for Cy5 coupled probes and 1:50 appears sufficient for Alexa594 coupled probes.
2. Centrifuge the fixed sample and aspirate away the ethanol.
3. Resuspend in 1 mL wash buffer that contains the same concentration of formamide as the hybridization buffer. Let stand for 2–5 min.
4. Centrifuge sample and aspirate away the wash buffer.
5. Add the hybridization solution prepared in step 1. Incubate overnight at 30°C in the dark.
    a. Note: The modified protocol on the Biosearch website now recommends 4 hour incubation at 37°C. We plan to try out this modification and compare carefully with our existing data sets. The result will be included in a future edition of this protocol.

Day 2:

1. In the morning, add 1 mL wash buffer to the sample. Vortex, spin down and aspirate.
2. Resuspend in another 1mL wash buffer and incubate at 30°C for 30 min.

3. Spin down and aspirate. Resuspend in 1mL wash buffer
4. Add 1.5 µl DAPI stain (5 ng/mL) for nuclear counterstaining. Vortex and incubate at 30°C for 30 min.
5. Spin down and aspirate the wash buffer. Resuspend in 2x SSC, vortex, spin down and aspirate.
6. If imaging without glox antifade solution (possible for Alexa594 and TMR), resuspend in a small amount (enough the cover the samples) of 2x SSC and proceed to imaging.
7. If imaging with the antifade solution, resuspend in a small amount of glox buffer and proceed to imaging.
    a. Note: While hybridized samples can be stored temporarily in 2x SSC at 4°C. Prolonged storage increases the risk of signal degradation. It is thus advisable to image immediately following hybridization.

* Note: The hybridization protocol posted by Biosearch (http://www.biosearchtech.com/assets/bti_custom_stellaris_celegans_protocol.pdf) differs from ours in the following: 1) omission of the blocking reagents (tRNA and BSA) from the hybridization buffer; 2) hybridize at 37°C (instead of 30°C) for 4 hours (instead of overnight). At the moment, we have not tried or systematically compared their protocol with our results. We plan to do so and will report our results in a future edition of this protocol.

## D. Imaging

**Microscopy equipment:**

1. Microscope: Standard wide-field fluorescence microscope (e.g., Nikon TE2000 or Ti, Zeiss Axiovert).

a. Note: Confocal microscopes, while excellent in spatial precision, use high light intensity and cause rapid bleaching of smFISH signals. They are especially problematic when taking multiple z sections and are thus not recommended for smFISH imaging.

2. Light source: A strong light source is essential for spot detection. Mercury or metal-halide lamp (e.g., ExFo Excite, Prior Lumen 200) are both good candidates. The metal-halide lamps are generally brighter and thus ideally suited for far red dyes such as Cy5.
    a. Note: With a metal-halide lamp, the exposure time we use for smFISH signal detection is generally around 1-2s, while only 100-200 ms is needed for GFP and DAPI.
3. Filter sets: Three filters, excitation, dichroic, and emission filters, are needed for each fluorescent channel. Choice of filters should be made based on the optical features of the corresponding fluorophores used (see Table 2 for a list of the filters we use).
4. Camera: Standard cooled CCD camera optimized for low-light level imaging rather than speed (preferably with 13 mm pixel size or less; e.g. CoolSNAP HQ from Pixis, Princeton Instruments).
5. High numerical aperture (NA> 1.3) 100x DIC objective (be sure to check transmission properties when using far red dyes such as Cy5 or Cy5.5). We have also seen spots using an oil-immersion 60x objective, but the reduced spatial resolution makes the spots somewhat more difficult to identify computationally.

**Table A-2.** Examples of optical filter sets compatible with multiplex mRNA detection

| | **Excitation** | **Dichroic** | **Emission** | **Supplier** |
|---|---|---|---|---|
| **Cy5** | HQ620/60x | Q660LP | HQ700/75m | Chroma |
| **Alexa 594** | 590DF10 | 610DRLP | 620DF30 | Omega |
| **TMR** | 546DF10 | 555DRLP | 580DF30 | Omega |

**Software:** Standard microscopy software capable of memorizing sample positions and imaging through z optical stacks. We currently use MetaMorph (Molecular Devices), which offers easily programmable user interface.

**Imaging chamber**: There are two purposes of the imaging chamber: 1. To affix the sample in a small region and prevent drying; 2. To minimize the thickness of the sample by flattening, thereby reducing out of focus light. To assemble the chamber:

1. Pipette 2-5 µl of the sample in antifade solution onto a clean 8mm round cover glass (Electron Microscopy Sciences, #1.5 thickness).
2. Gently tap a clean 22x22 mm square cover glass (VWR, #1 thickness) onto the drop of sample solution. This will cause the round cover glass to quickly adhere to the square glass.
    a. Optional: Clean cover glasses before hand by rinsing with 70% EtOH and let dry.
3. Immediately flip the square glass so that the round glass is on top. Let sit for half a minute or so while covered in dark.

4. While waiting, adhere a square Silicone Isolator (Grace Biolabs, 20 mm diameter x 0.5 depth) to a regular microscope slide.
5. Gently remove excess antifade solution from the rim of the round cover glass with Kimwipe.
6. Adhere the square cover glass to the Silicone Isolator with the round cover glass facing towards the microscope slide. Press down on the edges of the square glass to create a tight seal. This constitutes the imaging chamber.
7. Affix the imaging chamber to the microscope stage. Make sure to position the microscope slide correctly so that the imaging chamber faces the objective. Proceed to locate individual worms or embryos.

**What to expect:** A successful hybridization and sample preparation should yield clear fluorescent spots roughly 200-500 nm in size. These spots are so called "diffraction limited spots" based on the fact that the size of an mRNA molecule (nm) is far below the optical limit of the widefield microscope (μm). On a digitally acquired smFISH image, spot intensities usually vary within 2 fold of one another (for the same gene detected in the same fluorescent channel), with the exception of transcription centers in the nuclei (where large amounts of nascent transcripts accumulate), which can be much brighter. Optimal smFISH signal should be more than 1.5 times in intensity above the tissue background. When viewed on an image, the spots should be distinct in shape (i.e. not blurred) and readily identifiable by eye. Figure A-2 illustrates two smFISH experiments, one probing *elt-2* (a transcription factor involved in specifying the future intestine, Fukushige et al. 1998) in embryos (from Raj et al. 2010), the other staining three Wnt ligands expressed in the posterior body of the L1 larvae (from Harterink et al. 2011).

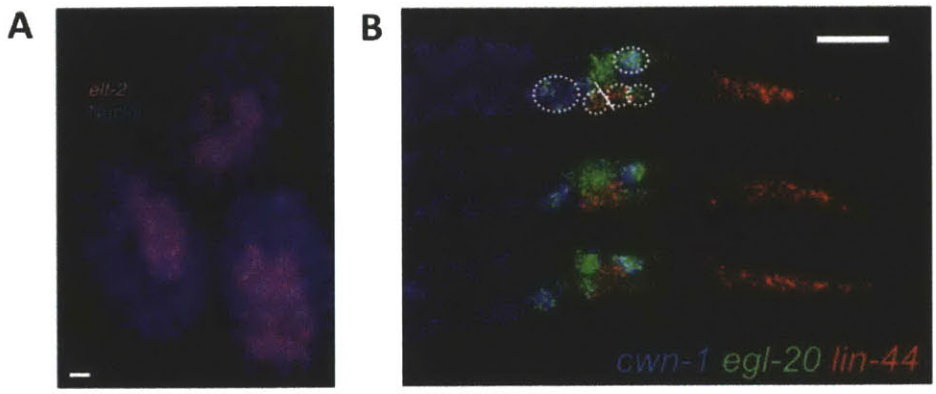



**Figure A-2.** Examples of smFISH staining from embryos and L1 larvae. **A.** (Adapted from Raj et al. 2010) Detection of *elt-2* mRNA (cyan spots) in mixed stage embryos. Scale bar is 5 μm long. **B.** (Adapted from Harterink et al. 2011) Detection of three Wnt ligands, *cwn-1, egl-20* and *lin-44,* in the posterior body of staged L1 larvae. Scale bar is 10 μm long.

## E. Data analysis

Successful smFISH experiments can simultaneously provide high quality information on the location, abundance, and transcriptional states of multiple mRNA species. To extract this wealth of information in an efficient and unbiased way, we recommend analyzing smFISH data using custom written computer software. While the exact analysis procedure may vary depending on experimental goals, we suggest the following analysis streamline (**Fig. A-3**) and briefly outline its implementation.

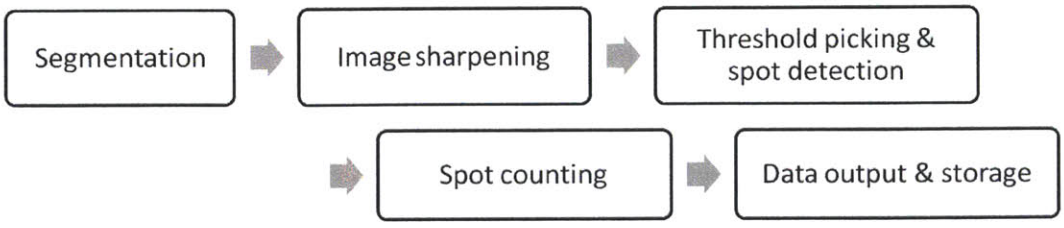



**Figure A-3.** Recommended steps to computationally analyze smFISH data.

1. **Image conversion:** Common imaging softwares such as MetaMorph and NIH-image save images in .tiff or .stk format. These files are not directly analyzable in computing software such as MATLAB, and must be converted to a readable format.

    a. Note: A MATLAB code that reads data from tiff files is available online at MATLAB Central: http://www.mathworks.com/matlabcentral/fileexchange/10298

2. **Image segmentation:** Image segmentation is a useful step in localizing mRNA expression to specific cell or tissue-types. Oftentimes, it is possible to take advantage of DAPI staining (for demarcating the body axes and major tissue groups, i.e. body wall muscles, ventral nerve cord, intestinal cells, etc.), GFP reporter systems, as well as smFISH signals (by staining for genes previously known to be expressed in a given cell or tissue) to accurately assign smFISH spots to its tissue of origin. Computational software can facilitate this process by overlaying images from multiple channels (e.g. Cy5, Alexa594, GFP and DAPI), followed by manual or automatic annotation of the relevant spatial landmarks. These annotations can later be overlaid onto the smFISH images to allow tissue-specific transcript quantification. Additional analysis, such as measurement of tissue length or size, can also be performed computationally at this step. If performing segmentation using computational software (e.g. MATLAB), the resulting traces and annotations can be saved as a standalone data file and used for region-specific spot counting (step 5) later.

3. **Image sharpening:** To facilitate subsequent automatic spot detection, it is useful to first enhance the signal-to-noise contrast of the smFISH images. This can be done by convolving each image file (i.e. a matrix consisting of individual pixel values) with two

mathematical functions (kernels), a Gaussian followed by a Laplacian. In practice the two functions can be convolved first to form a computational filter called the Laplacian of Gaussian (LoG). As shown in **Figure A-4**, the effect of convolving the original image with the LoG filter is twofold: a. the Gaussian filter smoothens the background to reduce small speckles that could be confused with real spots; b. the Laplacian filter then performs edge detection by amplifying the contrast between adjacent dim and bright pixels. The width of the Gaussian filter should be picked to resemble the size of a typical spot. Alternatively this can be done by trial and error (we recommend starting around 1.5) by visually comparing the filtered and original images. The following is an example of MATLAB code used for image sharpening:

```
H = -fspecial('log',15,1.2);    % H is the LoG filter with a
                                        standard deviation of 1.2
                                        and width around 5
g = imfilter(im,H);             % im is the converted smFISH image
```

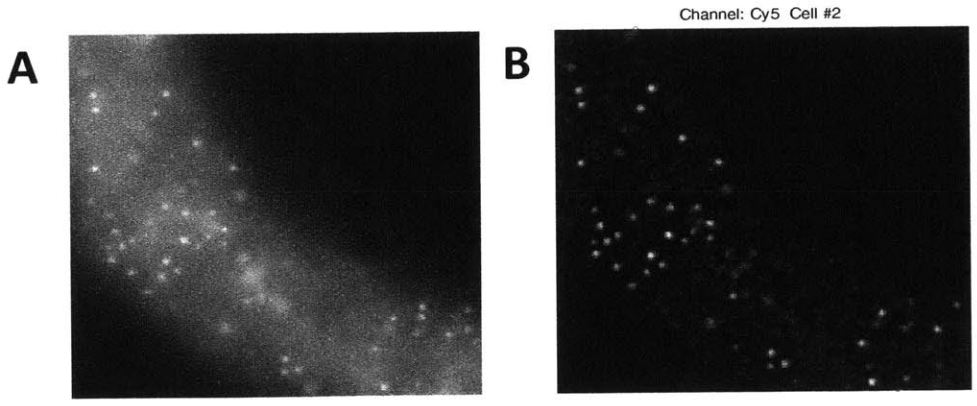



**Figure A-4.** Effects of LoG filtering on smFISH images. **A.** Raw image data staining *lin-17* mRNA in an L1 larva (shown is a single image from a stack). **B.** The same image after processing by a LoG filter with the above MATLAB commands. Note dramatic reduction in background and enhancement in spot signal.

4. **Spot detection and semi-automated threshold picking:** Since smFISH signals are significantly brighter than the background, individual spots are detected by identifying regions in the image with pixel values higher than a pre-specified threshold. One way of determining this threshold is to systematically screen threshold values in incremental steps from the minimum to the maximum pixel intensities. For each threshold value tested, the software identifies filled regions (or "connected objects") wherein all pixel values are above threshold. The total number of filled regions is then summed up to yield the total spot count corresponding to the given threshold.

   After spot detection and counting has been carried out for all thresholds, a plot can be generated showing total spot counts as a function of different threshold values. With high quality imaging data, this curve is expected to start high and drop rapidly to reach a brief plateau, before it finally approaches zero (**Fig. A-5**). The observed plateau corresponds to all threshold values above the background auto-fluorescence and below the real signal intensity. The total spot count is approximately constant in this range and corresponds reliably to the number of actual smFISH spots. Seemingly convenient, automatic detection of this plateau is sometimes difficult (since the size and flatness of this plateau varies from image to image). Especially during initial testing of the software, we recommend to manually pick the threshold, and plot the computationally detected spots over the original image to ensure that no over- or under-counting has occurred.

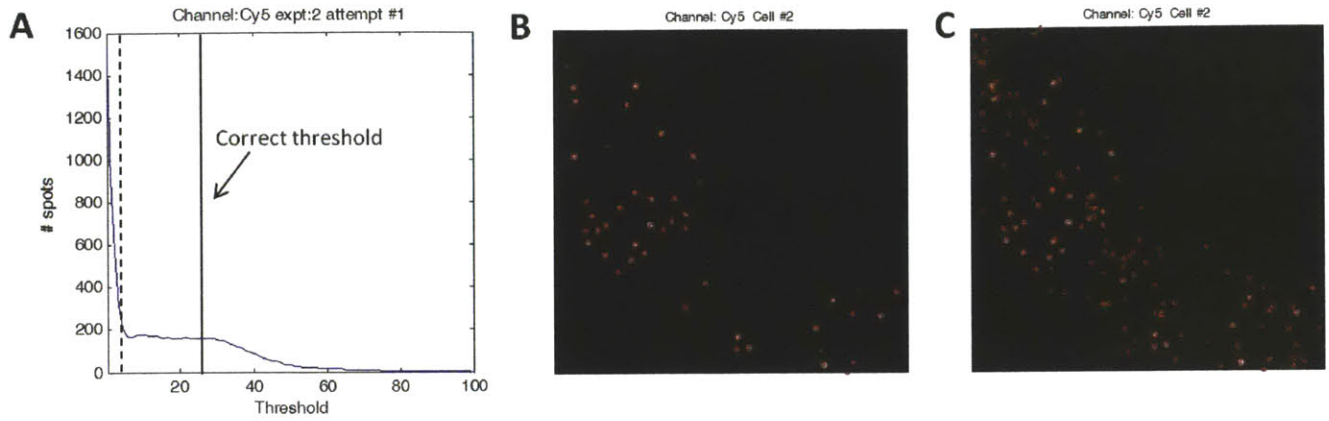



**Figure A-5.** Threshold picking and automatic detection of smFISH spots. **A.** Total spot counts as a function of pixel intensity threshold. The correct threshold should be placed where the spot count is insensitive to threshold value (where the "plateau" is). **B.** Spot detection when correct threshold (solid vertical line in A) is chosen. Red circles indicate computationally identified spots for a single image from a stack. Note out-of-focus spots are not counted. **C.** Spot detection when threshold is chosen too low (dotted line in A). Note many out-of-focus spots and background speckles are mistakenly included as spots.

5. **Region-specific spot counting:** Spot quantification within a specific cell or tissue can be conveniently achieved by aligning the results from image segmentation and spot detection. Computationally this can be done by generating a binary map for the region of interest (ROI) and using this map to filter out detected spots that are outside the ROI. MATLAB's Image Processing Toolbox provides many built-in functions that can greatly facilitate the image analysis steps described here. A sample of such MATLAB based software is available at the Raj lab website http://rajlab.seas.upenn.edu/pdfs/raj_nat_meth_2008_software.zip. More details of using MATLAB to perform image processing and single-molecule identification can be found in a recent paper by Scott Rifkin (2011). Other programming platforms, e.g. Python and Image J, also hold great promise in generating fast and accurate analysis software.

## Controls and troubleshooting

As the power of smFISH method lies in the accurate detection of single mRNA molecules, it is important to perform proper controls for each new smFISH probe library generated. Here we outline a number of control experiments that address common concerns regarding smFISH results. Additionally, while the smFISH protocols described here worked robustly for many *C. elegans* genes (including Wnt and Notch pathway genes, endoderm specification genes, etc. in both worms and embryos), we did notice a number of technical factors that can affect data quality. We thus provide a list of troubleshooting tips below to facilitate further optimization of the protocols.

**Recommended controls:**

1. *Does a single smFISH spot represent a single mRNA molecule?*

   Solution 1: To confirm single molecule resolution, one may randomly sample the peak pixel intensities of individual smFISH molecules within a given series of images (a process automatable by computer software). The spot intensities shall then be plotted on a histogram. The resulting distribution is expected to be unimodal with one narrow peak (this would not be the case if mRNA molecules frequently occur as groups of two or more) (Vargas et al. 2005). Large mRNA aggregates in the nucleus or specialized organelles (as in the case of transcription centers, P bodies and stress granules) are often many fold brighter and can be easily excluded from analysis by pixel thresholding. To control for the rare case where mRNA of the same species dimmerize (an example being

the Drosophila *bicoid*, Wagner et al. 2004), one may proceed to perform quantitative reverse-transcription PCR (qRT-PCR) as described below.

Solution 2: To independently check mRNA level, a control qRT-PCR experiment can be performed on bulk collected *C. elegans* tissue. As it is currently difficult to enrich for particular tissue or cell types in a bulk collection (with the exception of dissected gonads), this approach would only apply to whole worm or embryo measurement. qRT-PCR measurement from bulk collected tissue then needs to be normalized to by the estimated total number of worms or embryos. The obtained average expression level per animal can then be compared to average smFISH measurements, and the two are expected to be largely consistent.

2. *Are we robustly detecting all mRNA transcripts of a given species?* (Positive control)

Solution 1: Separate the uncoupled probe library into two halves containing non-overlapping sets of probes. Couple these two probe sets respectively with two different fluorophores and hybridize both probe sets to the same sample. If the majority of mRNA transcripts are reliably detected, most spots should appear double-labeled, and the overall degree of co-localization should be high. We have consistently detected around 85% co-localization, suggesting the majority of the mRNA transcripts are being detected by smFISH. Additionally, it is recommended to partition the probe library in such a way that allows probes with one fluorophore to interleave probes with the other fluorophore. Compared with having the two sets of probes targeting two different regions of the endogenous mRNA, this approach avoids difference in probe binding efficiency along different parts of the mRNA. A potential weakness is that this approach is non-robust to

mRNAs that are inaccessible by the smFISH protocol (e.g. mRNAs that are localized in specialized organelles or embodied by protein complexes). The following approach provides an independent and complementary assessment of total mRNA quantity.

Solution 2: To determine whether the number of transcripts detected by smFISH is accurate, one may perform, in parallel, quantitative reverse-transcription PCR (qRT-PCR) on a known number of synchronized worms. The total and average number of transcripts should agree between the two approaches. However this approach may not work well when transcript number within a specific tissue or region is of interest. qRT-PCR measurement is also subject to imperfect worm synchronization and may yield average counts that are lower or higher than the smFISH result (Vargas et al. 2005, Raj et al. 2008).

3. *Are we detecting mRNA transcripts other than the species of interest?* (Negative control)

Solution: If possible, obtain a mutant strain with either a large deletion in the coding sequence. Design probes that specifically target the deleted region, and hybridize the same probe set to both the wild-type strain and the mutant. While the probe should yield detectable signal in wild-type worms, no spots should be detected for the mutant. If a deletion allele is not available or the deletion is too short, consider using a strain harboring nonsense mutation. Confirm first by RT-PCR that the mRNA is greatly reduced compared to the wild-type (likely resulting from nonsense-mediated decay). Then perform smFISH to confirm that similar reduction in transcripts is observed in the mutant.

General note: When designing probes, it is good practice to blast the probe library against the *C. elegans* genome to ensure that majority of the targeted sequences are unique (esp. important when multiple paralogs exist in the genome). It is also important to keep the GC content of the probes around below 70% as GC rich probes are more prone to non-specific binding.

4. *How do we differentiate smFISH spots from auto-fluorescent speckles?* (Negative control)

   Solution 1: Auto-fluorescent speckles tend to show up in multiple fluorescent channels while smFISH spots light up only in the channel determined by the coupled fluorophore. Overlay images from different fluorescent channels to differentiate auto-fluorescence from the real signal.

   Solution 2: If auto-fluorescence is a strong concern, one may perform the hybridization procedure without adding the coupled probes. In the unlikely case that many fluorescent spots show up in the image, we know for sure that there is significant interfering signal from auto-fluorescence.

5. *Are there bleed-through between channels (in the case of multi-color smFISH)?*

   Solution: Bleed through between channels can happen when a signal in one channel is extremely strong. This is the case with intra-nuclear transcription centers, densely-packed transcripts (sometimes localized in organelles), and highly-over-expressed transgenes. To check for bleed-through, one may check if the spots detected in different channels (expected to represent two different transcript species) are highly co-localized. Alternatively, one may perform single-color FISH to see if spots are detected outside the expected channel. To avoid bleed-through choose channels that are as far apart in

emission wavelength as possible and choose optical filters that separate well between the channels (Table 2).

**Troubleshooting tips:**

1. *Low signal intensity or no signal.*

   Whenever possible, use a previously tested probe set (coupled to the same fluorophore) as positive control to make sure all equipments (light source, camera, software, etc.) are working properly. For issues specific to the particular probe set, consider checking the following: 1) *Probe concentration.* Test higher concentrations (e.g. start with one order of magnitude higher) to see if the signal quality improves. Based on experience, we recommend diluting in TE 1:20 for Cy5 coupled probes, 1:50 for Alexa, and 1:10 for TMR, followed by 1:100 dilution in hybridization buffer. 2) *Tissue penetration.* Make sure the sample has been incubated in EtOH for ample amount of time (e.g. 24 hours). 3) *Hybridization stringency.* Formamide concentration of the hybridization buffer directly controls binding efficiency. While 10% formamide generally works well for all *C. elegans* stages, one may try decreasing the concentration to see if signal improves. One should be cautious with this approach as low formamide concentration also ups the chance of non-specific binding.

2. *High fluorescent background*

   Multiple causes are likely: 1) *Unhealthy worms.* Starvation and over-crowded culture lead to high auto-fluorescence in the worm tissue. This is the most common cause of high background and should be avoided by proper worm culture maintenance. 2) *Over-fixation.* The second common cause and especially problematic for older worms. Keep

fixation time under an hour and adjust fixation duration to troubleshoot. Also make sure to thoroughly wash off the fixative with 1x PBS to terminate fixation. 3) *High probe concentration or insufficient washing*. This can be checked by testing lower dilutions and incubate longer during the two wash steps. 4) *Low probe binding stringency*. This could be a result of either low formamide concentration or high probe GC content. In both cases, probes bind non-specifically leading to increased background fluorescence. Unlike tissue auto-fluorescence, this phenomenon should be specific to the channel of the fluorophore whereas auto-fluorescence affects all channels. 5) *Out-of-focus light*. Real FISH signals that are outside the focal plane appear as diffuse background fluorescence in the focal plane. This problem is often dramatically improved by sucking away any excess fluid in the imaging chamber and maximally flattening the sample.

3. *Nonspecific signal (due to auto-fluorescence or bleed-through)*

   First make sure that the sample is collected from a clean, healthy batch of worms to reduce overall background. Sometimes, signal in the Alexa594 channel can bleed into the Cy5 channel if the two optical filters are not optimally separated in spectrum. Extremely high signal intensity (often in the case of over-expressed transgenes) in one channel can also broadly show up as "pseudo-spots" in multiple other channels. To check, test each probe set in a separate hybridization to see if problem persists.

4. *Irregular or diffuse spot morphology interfering with spot identification and counting*

   Successful smFISH experiment should yield spots circular in 2D and with sizes similar to one another. However liquid between the sample and the objective can obscure the expected optical properties of the spots. In such case, try to maximally flatten the sample

to reduce sample thickness and out-of-focus light. Additionally, highly local spot density (which occurs with highly expressed genes or transgenes) can lead to inevitable difficulty in resolving individual spots. When processing through computational software, try to identify local maxima in addition to "connected regions". This allows sub-diffraction detection of the (in principal) exact location of the mRNA transcript and thereby enhances the resolution of the smFISH signal.

# BIBLIOGRAPHY


Acar, M., Becskei, A., and Van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. Nature *435*, 228–232.

Acar, M., Mettetal, J.T., and Van Oudenaarden, A. (2008). Stochastic switching as a survival strategy in fluctuating environments. Nature Genetics *40*, 471–475.

Akaike, H.A.I. (1974). A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control *ac-19*,.

Van Amerongen, R., and Nusse, R. (2009). Towards an integrated view of Wnt signaling in development. Development (Cambridge, England) *136*, 3205–3214.

Andrec, M., Kholodenko, B.N., Levy, R.M., and Sontag, E. (2005). Inference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy. Journal of Theoretical Biology *232*, 427–441.

Aranda-Anzaldo, A., and Dent, M. a. R. (2003). Developmental noise, ageing and cancer. Mechanisms of Ageing and Development *124*, 711–720.

Austin, D.W., Allen, M.S., McCollum, J.M., Dar, R.D., Wilgus, J.R., Sayler, G.S., Samatova, N.F., Cox, C.D., and Simpson, M.L. (2006). Gene network shaping of inherent noise spectra. Nature *439*, 608–611.

Balázsi, G., Van Oudenaarden, A., and Collins, J.J. (2011). Cellular decision making and biological noise: from microbes to mammals. Cell *144*, 910–925.

Barabási, A.-L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nature Reviews. Genetics *5*, 101–113.

Becskei, a, Séraphin, B., and Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. The EMBO Journal *20*, 2528–2535.

Becskei, a, and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. Nature *405*, 590–593.

Blake, W.J., Balázsi, G., Kohanski, M. a, Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., and Collins, J.J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. Molecular Cell *24*, 853–865.

Blelloch, R., Newman, C., and Kimble, J. (1999). Control of cell migration during Caenorhabditis elegans development. Current Opinion in Cell Biology *11*, 608–613.

Boeger, H., Griesenbeck, J., and Kornberg, R.D. (2008). Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. Cell *133*, 716–726.

Brandman, O., Ferrell, J.E., Li, R., and Meyer, T. (2005). Interlinked fast and slow positive feedback loops drive reliable cell decisions. Science (New York, N.Y.) *310*, 496–498.

Burnham, K.P., and Anderson, D.R. (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (Springer).

Cadigan, K.M., Fish, M.P., Rulifson, E.J., and Nusse, R. (1998). Wingless repression of Drosophila frizzled 2 expression shapes the Wingless morphogen gradient in the wing. Cell *93*, 767–777.

Cağatay, T., Turcotte, M., Elowitz, M.B., Garcia-Ojalvo, J., and Süel, G.M. (2009). Architecture-dependent noise discriminates functionally analogous differentiation circuits. Cell *139*, 512–522.

Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. Nature *440*, 358–362.

Cao, Y., Lu, H.-M., and Liang, J. (2010). Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda. Proceedings of the National Academy of Sciences of the United States of America *107*, 18445–18450.

Carlson, M.R.J., Zhang, B., Fang, Z., Mischel, P.S., Horvath, S., and Nelson, S.F. (2006). Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. BMC Genomics *7*, 40.

Chalancon, G., Ravarani, C.N.J., Balaji, S., Martinez-Arias, A., Aravind, L., Jothi, R., and Babu, M.M. (2012). Interplay between gene expression noise and regulatory network architecture. Trends in Genetics : TIG *28*, 221–232.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544–547.

Chung, H., and Levens, D. (2005). Minireview Molecules and c-myc Expression : Keep the Noise Down ! *20*, 157–166.

Coudreuse, D.Y.M., Roël, G., Betist, M.C., Destrée, O., and Korswagen, H.C. (2006). Wnt gradient formation requires retromer function in Wnt-producing cells. Science (New York, N.Y.) *312*, 921–924.

Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. Nature *468*, 911–920.

Denning, G., Jamieson, L., Maquat, L.E., Thompson, E. a, and Fields, a P. (2001). Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein. The Journal of Biological Chemistry *276*, 22709–22714.

Doyle, J., and Csete, M. (2005). Motifs, control, and stability. PLoS Biology *3*, e392.

Dunlop, M.J., Cox, R.S., Levine, J.H., Murray, R.M., and Elowitz, M.B. (2008). Regulatory activity revealed by dynamic correlations in gene expression noise. Nature Genetics *40*, 1493–1498.

Eldar, A., Chary, V.K., Xenopoulos, P., Fontes, M.E., Losón, O.C., Dworkin, J., Piggot, P.J., and Elowitz, M.B. (2009). Partial penetrance facilitates developmental evolution in bacteria. Nature *460*, 510–514.

Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. Nature *467*, 167–173.

Elf, J., and Ehrenberg, M. (2003). Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. Genome Research *13*, 2475–2484.

Esposito, A., Dohm, C.P., Bähr, M., and Wouters, F.S. (2007). Unsupervised fluorescence lifetime imaging microscopy for high content and high throughput screening. Molecular & Cellular Proteomics : MCP *6*, 1446–1454.

Félix, M., and Wagner, a (2008). Robustness and evolution: concepts, insights and challenges from a developmental model system. Heredity *100*, 132–140.

Ferrell, J.E., Tsai, T.Y.-C., and Yang, Q. (2011). Modeling the cell cycle: why do certain circuits oscillate? Cell *144*, 874–885.

Forrester, W.C., Kim, C., and Garriga, G. (2004). The Caenorhabditis elegans Ror RTK CAM-1 inhibits EGL-20/Wnt signaling in cell migration. Genetics *168*, 1951–1962.

Friedman, N., Cai, L., and Xie, X. (2006). Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression. Physical Review Letters *97*, 1–4.

Geisen, M.J., Di Meglio, T., Pasqualetti, M., Ducret, S., Brunet, J.-F., Chedotal, A., and Rijli, F.M. (2008). Hox paralog group 2 genes control the migration of mouse pontine neurons through slit-robo signaling. PLoS Biology *6*, e142.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. Cell *123*, 1025–1036.

Green, J., Inoue, T., and Sternberg, P.W. (2008). Opposing Wnt pathways orient cell polarity during organogenesis. Cell *134*, 646–656.

Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007). Probing the limits to positional information. Cell *130*, 153–164.

Grimson, A., Connor, S.O., Newman, C.L., and Anderson, P. (2004). SMG-1 Is a Phosphatidylinositol Kinase-Related Protein Kinase Required for Nonsense-Mediated mRNA Decay in Caenorhabditis elegans. *24*, 7483–7490.

Harris, J., Honigberg, L., Robinson, N., and Kenyon, C. (1996). Neuronal cell migration in C. elegans: regulation of Hox gene expression and cell position. Development (Cambridge, England) *122*, 3117–3131.

Hasty, J., McMillen, D., Isaacs, F., and Collins, J.J. (2001). Computational studies of gene regulatory networks: in numero molecular biology. Nature Reviews. Genetics *2*, 268–279.

Hasty, J., Pradines, J., Dolnik, M., and Collins, J.J. (2000). Noise-based switches and amplifiers for gene expression. Proceedings of the National Academy of Sciences of the United States of America *97*, 2075–2080.

Henrichsen, C.N., Chaignat, E., and Reymond, A. (2009). Copy number variants, diseases and gene expression. Human Molecular Genetics *18*, R1–8.

Herman, M.A. (2003). Wnt Signaling in C. elegans. In Wnt Signaling in Development, M. Kuhl, ed. (Eurekah.com and Kluwer Academic/Plenum Publishers),.

Hirsch, H. a, Iliopoulos, D., Joshi, A., Zhang, Y., Jaeger, S. a, Bulyk, M., Tsichlis, P.N., Shirley Liu, X., and Struhl, K. (2010). A transcriptional signature and common gene networks link cancer with lipid metabolism and diverse human diseases. Cancer Cell *17*, 348–361.

Hooshangi, S., Thiberge, S., and Weiss, R. (2005). Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. Proceedings of the National Academy of Sciences of the United States of America *102*, 3581–3586.

Hornstein, E., and Shomron, N. (2006). Canalization of development by microRNAs. Nature Genetics *38 Suppl*, S20–4.

Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. Nature Reviews. Genetics *6*, 451–464.

Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biology *7*, e1000149.

Van Kampen, N.G. (1983). Stochastic Processes in physics and chemistry (North-Holland).

Katoh, M., and Katoh, M. (2007). Comparative integromics on FZD7 orthologs: conserved binding sites for PU.1, SP1, CCAAT-box and TCF/LEF/SOX transcription factors within 5'-promoter region of mammalian FZD7 orthologs. International Journal of Molecular Medicine *19*, 529–533.

Kenyon, C.J., Austin, J., Costa, M., Cowing, D.W., Harris, J.M., and Honigberg, L. (1997). The Dance of the Hox Genes : Patterning the Anteroposterior Body Axis of Caenorhabditis elegans The Dance of the Hox Genes : Patterning the Anteroposterior Body Axis of Caenorhabditis elegans. Spring 293–305.

Kepler, T.B., and Elston, T.C. (2001). Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophysical Journal *81*, 3116–3136.

Kholodenko, B.N., Kiyatkin, A., Bruggeman, F.J., Sontag, E., Westerhoff, H. V, and Hoek, J.B. (2002). Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. Proceedings of the National Academy of Sciences of the United States of America *99*, 12841–12846.

Komili, S., and Silver, P. a (2008). Coupling and coordination in gene expression processes: a systems biology view. Nature Reviews. Genetics *9*, 38–48.

Korswagen, H.C. (2002). Canonical and non-canonical Wnt signaling pathways in Caenorhabditis elegans: variations on a common signaling theme. BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology *24*, 801–810.

Kussell, E., and Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. Science (New York, N.Y.) *309*, 2075–2078.

Lagha, M., Bothma, J.P., and Levine, M. (2012). Mechanisms of transcriptional precision in animal development. Trends in Genetics : TIG *28*, 409–416.

Levin, M.D., Morton-Firth, C.J., Abouhamad, W.N., Bourret, R.B., and Bray, D. (1998). Origins of individual swimming behavior in bacteria. Biophysical Journal *74*, 175–181.

Li, G.-W., and Xie, X.S. (2011). Central dogma at the single-molecule level in living cells. Nature *475*, 308–315.

Lidstrom, M.E., and Konopka, M.C. (2010). The role of physiological heterogeneity in microbial population behavior. Nature Chemical Biology *6*, 705–712.

Ling, L., Nurcombe, V., and Cool, S.M. (2009). Wnt signaling controls the fate of mesenchymal stem cells. Gene *433*, 1–7.

Liu, C., Wang, Y., Smallwood, P.M., and Nathans, J. (2008). An essential role for Frizzled5 in neuronal survival in the parafascicular nucleus of the thalamus. The Journal of Neuroscience : the Official Journal of the Society for Neuroscience *28*, 5641–5653.

Maloof, J.N., Whangbo, J., Harris, J.M., Jongeward, G.D., and Kenyon, C. (1999). A Wnt signaling pathway controls hox gene expression and neuroblast migration in C. elegans. Development (Cambridge, England) *126*, 37–49.

Mantzaris, N. V (2007). From single-cell genetic architecture to cell population dynamics: quantitatively decomposing the effects of different population heterogeneity sources for a genetic network with positive feedback architecture. Biophysical Journal *92*, 4271–4288.

Middelkoop, T.C., Williams, L., Yang, P.-T., Luchtenberg, J., Betist, M.C., Ji, N., Van Oudenaarden, A., Kenyon, C., and Korswagen, H.C. (2012). The thrombospondin repeat containing protein MIG-21 controls a left-right asymmetric Wnt signaling response in migrating C. elegans neuroblasts. Developmental Biology *361*, 338–348.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science (New York, N.Y.) *298*, 824–827.

Munsky, B., Neuert, G., and Van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. Science (New York, N.Y.) *336*, 183–187.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840–846.

Nijhout, H.F. (2002). The nature of robustness in development. BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology *24*, 553–563.

Novick, A., and Weiner, M. (1957). Enzyme induction as an all-or-none phenomenon. PNAS *43*, 553–566.

Ou, G., and Vale, R.D. (2009). Molecular signatures of cell migration in C. elegans Q neuroblasts. The Journal of Cell Biology *185*, 77–85.

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of Escherichia coli. Nature *427*, 737–740.

Pan, C.-L., Howell, J.E., Clark, S.G., Hilliard, M., Cordes, S., Bargmann, C.I., and Garriga, G. (2006). Multiple Wnts and frizzled receptors regulate anteriorly directed cell and growth cone migrations in Caenorhabditis elegans. Developmental Cell *10*, 367–377.

Pedraza, J.M., and Van Oudenaarden, A. (2005). Noise propagation in gene networks. Science (New York, N.Y.) *307*, 1965–1969.

Pulak, R., and Anderson, P. (1993). mRNA surveillance by the Caenorhabditis elegans smg genes. Genes & Development *7*, 1885–1897.

Raj, A., Bogaard, P. Van Den, Rifkin, S.A., Oudenaarden, A. Van, and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nature Methods *5*, 877–879.

Raj, A., and Van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biology *4*, e309.

Raj, A., Rifkin, S.A., Andersen, E., and Oudenaarden, A. Van (2010). Variability in gene expression underlies incomplete penetrance. Nature *463*, 913–918.

Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. Science (New York, N.Y.) *304*, 1811–1814.

Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., and Elowitz, M.B. (2005). Gene regulation at the single-cell level. Science (New York, N.Y.) *307*, 1962–1965.

Rulifson, E.J., Wu, C., and Nusse, R. (2000). Pathway Specificity by the Bifunctional Receptor Frizzled Is Determined by Affinity for Wingless. *6*, 117–126.

Salser, S.J., and Kenyon, C. (1992). Activation of a C. elegans Antennapedia homologue in migrating cells controls their direction of migration. Nature.

Sanchez, A., Garcia, H.G., Jones, D., Phillips, R., and Kondev, J. (2011). Effect of promoter architecture on the cell-to-cell variability in gene expression. PLoS Computational Biology *7*, e1001100.

Sato, a, Kojima, T., Ui-Tei, K., Miyata, Y., and Saigo, K. (1999). Dfrizzled-3, a new Drosophila Wnt receptor, acting as an attenuator of Wingless signaling in wingless hypomorphic mutants. Development (Cambridge, England) *126*, 4421–4430.

Schulte, G. (2010). International Union of Basic and Clinical. LXXX. The Class Frizzled Receptors. Pharmacological Reviews *62*, 632–667.

Shahrezaei, V., and Swain, P.S. (2008a). Analytical distributions for stochastic gene expression. Proceedings of the National Academy of Sciences of the United States of America *105*, 17256–17261.

Shahrezaei, V., and Swain, P.S. (2008b). The stochastic nature of biochemical networks. Current Opinion in Biotechnology *19*, 369–374.

Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nature Genetics *31*, 64–68.

Silhankova, M., and Korswagen, H.C. (2007). Migration of neuronal cells along the anterior-posterior body axis of C. elegans: Wnts are in control. Current Opinion in Genetics & Development *17*, 320–325.

Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature *459*, 428–432.

Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Developmental Biology *56*, 110–156.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science (New York, N.Y.) *329*, 533–538.

Thattai, M., and Van Oudenaarden, A. (2004). Stochastic gene expression in fluctuating environments. Genetics *167*, 523–530.

To, T.-L., and Maheshri, N. (2010). Noise can induce bimodality in positive transcriptional feedback loops without bistability. Science (New York, N.Y.) *327*, 1142–1145.

Walczak, A.M., Mugler, A., and Wiggins, C.H. (2011). Analytic methods for modeling stochastic regulatory networks. 1–37.

Wang, B.B., Müller-Immergluck, M.M., Austin, J., Robinson, N.T., Chisholm, a, and Kenyon, C. (1993). A homeotic gene cluster patterns the anteroposterior body axis of C. elegans. Cell *74*, 29–42.

Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D. V (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell *122*, 169–182.

Wernet, M.F., Mazzoni, E.O., Celik, A., Duncan, D.M., Duncan, I., and Desplan, C. (2006). Stochastic spineless expression creates the retinal mosaic for colour vision. Nature *440*, 174–180.

Whangbo, J., and Kenyon, C. (1999). A Wnt signaling system that specifies two patterns of cell migration in C. elegans. Molecular Cell *4*, 851–858.

Willert, J., Epping, M., Pollack, J.R., Brown, P.O., and Nusse, R. (2002). A transcriptional response to Wnt protein in human embryonic carcinoma cells. BMC Developmental Biology *2*, 8.

Xiong, W., and Ferrell Jr, J.E. (2003). A positive-feedback-based bistable " memory module " that governs a cell fate decision. *426*, 460–465.

Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X.S. (2006). Probing gene expression in live cells, one protein molecule at a time. Science (New York, N.Y.) *311*, 1600–1603.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nature Structural & Molecular Biology *15*, 1263–1271.

Zinovyeva, A.Y., Yamamoto, Y., Sawa, H., and Forrester, W.C. (2008). Complex network of Wnt signaling regulates neuronal migrations during Caenorhabditis elegans development. Genetics *179*, 1357–1371.

Femino A.M., Fay F.S., Fogarty K., and Singer R.H. (1998). Visualization of single RNA transcripts *in situ*. Science *280*, 585-590.

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

Harterink M., Kim D., Middelkoop T.C., Doan T.D., van Oudenaarden A., and Korswagen H.C. (2011). Neuroblast migration along the anteroposterior axis of C. elegans is controlled by opposing gradients of Wnts and a secreted Frizzled-related protein. Development *138*, 2915-2924.

Korzelius J., The I., Ruijtenberg S., Prinsen M.B., Portegijs V., Middelkoop T.C., Groot Koerkamp M.J., Holstege F.C., Boxem M., and van den Heuvel S. (2011). *Caenorhabditis elegans* cyclin D/CDK4 and cyclin E/CDK2 induce distinct cell cycle re-entry programs in differentiated muscle cells. PloS Genetics *7*, e1002362.

Levsky J.M. & Singer R.H. (2003). Fluorescence in situ hybridization: past, present and future. Journal of Cell Science. *116*, 2833–2838.

Maloof, J.N., Whangbo J., Harris J.M., Jongeward G.D., and Kenyon C. (1999). A Wnt signaling pathway controls Hox gene expression and neuroblast migration in *C. elegans*. Development *126*, 37-49.

Middelkoop T.C., Williams L., Yang P.T., Betist M.C., Ji N., van Oudenaarden A., Kenyon C., Korswagen H.C. (2012). The thrombospondin repeat containing protein MIG-21 controls a left-right asymmetric Wnt signaling response in migrating *C. elegans* neuroblasts. Developmental Biology *361*, 338-348.

Raap A.K., van de Corput M.P., Vervenne R.A., van Gijlswijk R.P., Tanke H.J., and Wiegant J. (1995). Ultra-sensitive FISH using peroxidase-mediated deposition of biotin- or fluorochrome tyramides. Human Molecular Genetics *4*, 529-534.

Raj A., Rifkin S.A., Anderson E., and van Oudenaarden A. (2010). Variability in gene expression underlies incomplete penetrance. Nature *463*, 913-918.

Raj A. and Tyagi S. (2010). Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. Methods in Enzymology *472*, 365-386.

Raj A., van den Bogaard P., Rifkin SA, van Oudenaarden A., and Tyagi S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nature Methods *5*, 877–879.

Rifkin S.A. (2011). Identifying fluorescently labeled single molecules in image stacks using machine learning. Molecular Methods for Evolutionary Genetics. (Orgogozo & Rockman, eds.). Methods in Molecular Biology. *772*: 329-348.

Saffer A.M., Kim D., van Oudenaarden A., and Horvitz H.R. (2011). The *Caenorhabditis elegans* synthetic multivulva genes prevent Ras pathway activation by tightly repressing global ectopic expression of lin-3 EGF. PloS Genetics. *7*, e1002418.

Seidel H.S., Ailion M., Li J., van Oudenaarden A., Rockman, M.V., and Kruglyak L. (2011). A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in C. elegans. PloS Biology *9*, e1001115.

Topalidou I., van Oudenaarden A.,and Chalfie M. (2011). Caenorhabditis elegans aristaless/Arx gene alr-1 restricts variable gene expression. PNAS *108*, 4063-4068.

Vargas D.Y. Raj A., Marras S.A., Kramer F.R. and Tyagi S. (2005). Mechanism of mRNA transport in the nucleus. PNAS *102*, 17008-17013.

Wagner C., Ehresmann C., Ehresmann B., and Brunel C. (2004). Mechanism of dimerization of Bicoid mRNA: Initiation and stabilization. Journal of Biological Chemistry. *279,* 4560-4569.