ABSTRACT

JERMUSYK, ASHLEY ANN. Novel Methods for Elucidating Signaling Pathways by Using *Drosophila melanogaster* Embryos. (Under the direction of Dr. Gregory T. Reeves).

In multicellular organisms, cellular signaling events are crucial for patterning tissues, as well as for maintaining healthy adult tissues, while improper signaling can lead to disease states, such as cancer. Therefore, cellular signaling processes must be tightly regulated. Complex systems of gene regulatory circuits control these signaling processes and act to buffer these systems against noise, thereby minimizing mistakes in gene expression and preventing patterning defects or disease states. There are two different approaches that can be carried out to gain further understanding of these signaling processes: (1) studying endogenous networks and (2) studying synthetic networks. The main difficulties associated with studying endogenous networks are due to the high connectivity and complexities of these networks which make hypothesis testing difficult. However, this difficulty was overcome in our work by exploring differences in natural gene expression due to subtle genomic differences to discover novel binding sites and connections within the anterior-posterior patterning system of the Drosophila melanogaster embryo. Alternatively to overcome these issues with endogenous networks, synthetic networks can be used. These synthetic studies are usually done in single cell systems and it can be difficult to extrapolate these effects to multi-cellular organisms. This difficulty was overcome in our work by using synthetic gene networks in a multi-cellular system, namely the Drosophila embryo.

Using synthetic gene networks, negative feedback and cross-repression were studied, both of which have been proposed to provide robustness to gene networks. In our negative feedback study, we were able to create both an attenuation and a shuttling system depending on the amount of activator to inhibitor in the system. Cross-repression networks were

created such that scaling and sharpness could be studied. While conducting these studies it is important that gene levels within the network must be similar to each other to exhibit functionality. To do this it is advantageous to introduce new methods for tuning gene expression levels which can be applied to these network studies, but also to future endogenous and synthetic studies in Drosophila and other multi-cellular systems. Two methods that were explored are self-cleaving hammerhead ribozymes and short upstream open reading frames.

The anterior-posterior patterning system in the Drosophila embryo was examined by quantifying differences in positioning of gene expression boundaries of Krüppel and Even-skipped domains in a selection of wild-caught lines. The differences in the positions of Krüppel and Even-skipped were then correlated to genomic differences. These genomic differences were then explored to determine whether they affected transcription factor binding sites and, if so, the identity of these novel transcription factors.

Novel Methods for Elucidating Signaling Pathways by Using *Drosophila melanogaster* Embryos

by
Ashley Ann Jermusyk

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Chemical and Biomolecular Engineering

Raleigh, North Carolina

2016

APPROVED BY:

Gregory T. Reeves Committee Chair	Chase Beisel
James Mahaffey	Balaji Rao

BIOGRAPHY

Ashley Jermusyk was born in Southampton, NY and graduated from Miller Place High School. She attended Lafayette College, earning a B.S. in Chemical Engineering. She will (hopefully) shortly have a PhD in Chemical Engineering from North Carolina State University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction and Background

CHAPTER 2: Transcription Factor Networks

CHAPTER 4: Determination of Novel Players in the *Drosophila melanogaster* Anterior-Posterior Patterning System Using Natural Variation

CHAPTER 5: Creation and Modeling of Synthetic Sharpness and Scaling Gene Networks in Drosophila Embryos

CHAPTER 6: New Tools for Controlling Gene Expression in *Drosophila melanogaster* Embryos Using Hammerhead Ribozymes and Short Upstream Open Reading Frames

CHAPTER 1

Introduction and Background

1.1 - Gene Regulation and Tissue Patterning

Methods of gene regulation are conserved throughout nature (Reeves, Stathopoulos 2009, Milo et al. 2002, Thieffry, Sanchez 2003, Peter, Davidson 2009). Therefore we can use the information about regulation in one system (in this case the development of the *Drosophila melanogaster* embryo) and apply it to systems in other organisms and throughout the life cycle of an organism. It is important to understand how this regulation occurs because errors during development or tissue maintenance results in disease. By gaining a better understanding of the interactions that occur, we can gain insights into mis-regulation that causes diseases and therefore design treatments for these diseases.

During development, gene regulation occurs through a cascade of signals. A protein known as a morphogen is expressed in a concentration gradient, which triggers gene expression of downstream genes based on concentration thresholds (Wolpert 1969). This concentration gradient is formed by a localized source of protein which then diffuses outward and is degraded, resulting in an exponentially decaying concentration gradient. While the morphogen is usually used to trigger expression of the downstream patterning genes, many methods of control exist between these genes to allow for robust patterning. These recurring, functional gene interactions are known as network motifs (Alon 2007).

These motifs can exhibit different outputs, known as emergent properties, which represent the different patterning results of the motifs depending on the specific conditions within the system. These emergent properties are new properties of these small networks, which the individual parts do not possess. Common motifs and how they are combined to form gene regulatory networks and result in these emergent properties

are described in Chapter 2. Simple gene circuits of particular note are negative feedback and cross-repression, which are discussed in Chapter 2. The emergent properties of particular note are shuttling, attenuation, scaling, and sharpness. These networks were examined using experimental synthetic gene networks, for more details on this work see in Chapter 3 (negative feedback and associated attenuation and shuttling) and Chapter 5 (cross-repression and associated sharpness and scaling).

1.1.1 - Shuttling

One emergent property that has recently been identified is shuttling. In shuttling an initial broad morphogen signal is refined to produce a sharp, robust gradient. In shuttling there must be a second molecule which forms a complex with the morphogen and acts as the shuttling molecule. A third molecule can also be used which functions to break up this morphogen-shuttler complex. In the developing *Drosophila melanogaster* embryo a shuttling mechanism is able to produce a sharp gradient of Toll by the action of its inhibitor Spaetzle (Haskel-Ittah et al. 2012). This mechanism has also been proposed to refine the Dorsal gradient in the *Drosophila* embryo through its inhibitor Cactus (Carrell and O'Connell, unpublished data). These are just two of the identified instances of shuttling occurring in the *Drosophila* embryo. However, this shuttling mechanism is most likely present in many different systems and in a diverse set of organisms.

1.1.2 – Scaling

In a scaling network the system is able to reproducibly express a gene over a domain (as a given fraction of the tissue) regardless of tissue size. One way this can be

accomplished by using inputs from two opposing morphogen gradients (Houchmandzadeh, Wieschaus & Leibler 2005, McHale, Rappel & Levine 2006). These two opposing morphogen gradients can provide inputs for two different downstream genes that are expressed in domains on opposite ends of the tissue. Interactions between these downstream genes causes positioning of the intersection of these genes to be dependent on the initial morphogen gradients and therefore tissue size. It has been proposed and demonstrated through mathematical models, these two opposing morphogen gradients can be used to reproducibly position Hunchback expression in the early *Drosophila* embryo regardless of changes to the Bicoid morphogen gradient (Houchmandzadeh, Wieschaus & Leibler 2005, McHale, Rappel & Levine 2006). This mechanism of scaling is proposed to occur, but has been difficult to validate experimentally. To investigate the ability of a simple scaling network to achieve reproducible positioning of downstream genes we created a synthetic scaling network in the *Drosophila* embryo, see Chapter 5.

1.2 - Anterior-Posterior Patterning in *Drosophila melanogaster* embryo

Gene regulatory motifs come together to form gene regulatory networks (see Chapter 2 for more details). One well studied gene regulatory network is the anterior-posterior patterning system in the *Drosophila melanogaster embryo*. The morphogen gradient, Bicoid, initiates the patterning of the anterior-posterior axis in the *Drosophila melanogaster* embryo. Bicoid signals downstream gap genes (*hunchback*, *Krüppel*, *giant*, and *knirps*), which are expressed in broad domains (Perkins et al. 2006, Jaeger 2011). The gap genes activate pair-rule genes, expressed in repeating stripes in the

embryo. The pair-rule genes activate segment polarity genes, and these segment polarity genes together with the pair rule genes regulate the Hox genes (McGinnis, Krumlauf 1992, von Dassow et al. 2000).

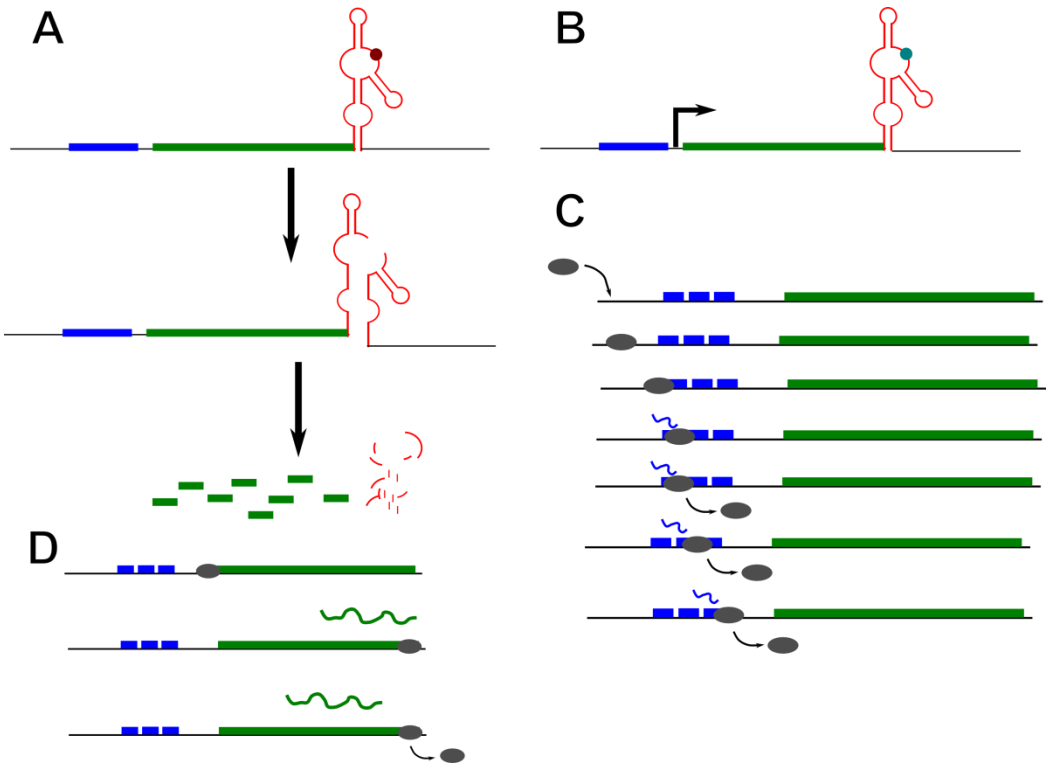

Figure 1.1: Ribozymes and upstream open reading frames can control gene expression. (A) The active ribozyme is downstream of the gene of interest on the same mRNA strand. Self-cleavage occurs in the hammerhead, removing the poly-(A) tail, destabilizing the mRNA, leading to self-degradation. (B) A single base pair substitution inactivates the ribozyme, preventing self-cleavage and allowing for translation. (C) Short open reading frames (blue) are translated by the ribosome as it scans the mRNA, preventing translation of the gene of interest (green). (D) A small number of ribosomes will "leak" past the short uORFs and translated the gene of interest.

1.3 - Tuning Gene Expression

For studies of both endogenous networks and synthetic networks it is advantageous to accurately tune gene expression levels within these networks. This has historically been done in endogenous networks by introducing mutations into the extended gene region. However, the effects of these mutations are not predictable, so it is

difficult to alter expression to a desired level. RNAi techniques have been used for both endogenous and synthetic systems to knock-down gene expression (Deans, Cantor & Collins 2007, Horn et al. 2011). In addition, a number of methods have been used to alter gene expression in single-cell organisms that show promise for use in synthetic networks. These methods include: protein tags for targeted protein degradation, special promoters systems, self-cleaving hammerhead ribozymes, short upstream open reading frames, and synthetic ribosome binding sites (Salis, Mirsky & Voigt 2009, Guzman et al. 1995, McGinness, Baker & Sauer 2006, Yen et al. 2004). We investigated self-cleaving hammerhead ribozymes and short upstream open reading frames for use in *Drosophila* embryos in Chapter 6.

1.3.1 - Self-Cleaving Hammerhead Ribozymes

The hammerhead ribozyme sequence is inserted such that it is expressed in the same mRNA as the gene you wish to control, see Fig. 1.1A. The ribozyme can then undergo self-cleavage (Fig. 1.1A) (Yen et al. 2004, Auslander, Ketzer & Hartig 2010). Following this self-cleavage of the mRNA, the mRNA is degraded resulting in lower mRNA levels and therefore lower protein expression (Yen et al. 2004, Auslander, Ketzer & Hartig 2010). The ability of the ribozyme to undergo self-cleavage can be altered by changing the sequences surrounding the ribozyme, modifying its ability to fold into an active conformation. Additionally the ribozyme can be inserted at different locations relative to the open reading frame (5' or 3') to alter its functionality (Yen et al. 2004, Auslander, Ketzer & Hartig 2010). This system has been designed to tune protein

expression in mammalian cells (Auslaender et al. 2014, Wieland, Auslaender & Fussenegger 2012).

1.3.2 - Short Upstream Open Reading Frames

By inserting short open reading frames upstream of the open reading frame for a gene of interest, the level of protein expression for the gene of interest can be altered (Fig. 1.1C-D) (Ferreira, Overton & Wang 2013). Only a fraction of ribosomes will "leak" past the upstream open reading frames to translate the gene of interest (Ferreira, Overton & Wang 2013). The level of protein expression can be adjusted by changing the number of upstream open reading frames and the ability of the ribosome to recognize these open reading frames by altering the sequences surrounding the start codon (Ferreira, Overton & Wang 2013).

1.4 - Synthetic Gene Networks

In many cases it is desirable to study a gene motif in isolation. This can be accomplished using a synthetic gene network. Simple synthetic networks have been created which demonstrate the ability to reproduce a functional gene network in isolation. This includes a gene oscillator, toggle switch, and counting devices (Elowitz, Leibler 2000, Gardner, Cantor & Collins 2000, Friedland et al. 2009). Simple patterns have been created using simple synthetic gene networks in *E. coli* clonal populations (Payne et al. 2013, Mitarai, Jensen & Semsey 2015, Schaerli et al. 2014, Basu et al. 2004). These devices have used computational models to determine networks that can express these patterns (Payne et al. 2013, Mitarai, Jensen & Semsey 2015, Schaerli et al. 2014, Basu et

al. 2004). These networks demonstrate synthetic gene networks in 2D systems; however work has not been carried out to bring these synthetic networks into true multicellular tissue patterning systems. We were able to create and express synthetic gene networks in *Drosophila* embryos, which are discussed in Chapters 3 and 5. Since these motifs are present in a number of different multi-cellular systems, including many patterning networks as described above, we can apply the information gained from these synthetic gene networks in *Drosophila* to many different organisms and systems.

References

Alon, U. 2007, "Network motifs: theory and experimental approaches.", *Nat Rev Genet,* vol. 8, no. 6, pp. 450-461.

Auslaender, S., Stuecheli, P., Rehm, C., Auslaender, D., Hartig, J.S. & Fussenegger, M. 2014, "A general design strategy for protein-responsive riboswitches in mammalian cells", *Nature Methods,* vol. 11, no. 11, pp. 1154-1160.

Auslander, S., Ketzer, P. & Hartig, J.S. 2010, "A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression.", *Mol Biosyst,* vol. 6, no. 5, pp. 807-814.

Basu, S., Mehreja, R., Thiberge, S., Chen, M. & Weiss, R. 2004, "Spatiotemporal control of gene expression with pulse-generating networks", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 101, no. 17, pp. 6355-6360.

Deans, T.L., Cantor, C.R. & Collins, J.J. 2007, "A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells", *Cell,* vol. 130, no. 2, pp. 363-372.

Elowitz, M.B. & Leibler, S. 2000, "A synthetic oscillatory network of transcriptional regulators.", *Nature,* vol. 403, no. 6767, pp. 335-338.

Ferreira, J.P., Overton, K.W. & Wang, C.L. 2013, "Tuning gene expression with synthetic upstream open reading frames", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 110, no. 28, pp. 11284-11289.

Friedland, A.E., Lu, T.K., Wang, X., Shi, D., Church, G. & Collins, J.J. 2009, "Synthetic Gene Networks That Count", *Science,* vol. 324, no. 5931, pp. 1199-1202.

Gardner, T., Cantor, C. & Collins, J. 2000, "Construction of a genetic toggle switch in Escherichia coli", *Nature,* vol. 403, no. 6767, pp. 339-342.

Guzman, L., Belin, D., Carson, M. & Beckwith, J. 1995, "Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P-Bad Promoter", *Journal of Bacteriology,* vol. 177, no. 14, pp. 4121-4130.

Haskel-Ittah, M., Ben-Zvi, D., Branski-Arieli, M., Schejter, E.D., Shilo, B. & Barkai, N. 2012, "Self-Organized Shuttling: Generating Sharp Dorsoventral Polarity in the Early Drosophila Embryo", *Cell,* vol. 150, no. 5, pp. 1016-1028.

Horn, T., Sandmann, T., Fischer, B., Axelsson, E., Huber, W. & Boutros, M. 2011, "Mapping of signaling networks through synthetic genetic interaction analysis by RNAi", *Nature Methods,* vol. 8, no. 4, pp. 341-U91.

Houchmandzadeh, B., Wieschaus, E. & Leibler, S. 2005, "Precise domain specification in the developing Drosophila embryo.", *Phys Rev E Stat Nonlin Soft Matter Phys,* vol. 72, no. 6 Pt 1, pp. 061920.

Jaeger, J. 2011, "The gap gene network", *Cellular and Molecular Life Sciences,* vol. 68, no. 2, pp. 243-274.

McGinness, K., Baker, T. & Sauer, R. 2006, "Engineering controllable protein degradation", *Molecular cell,* vol. 22, no. 5, pp. 701-707.

McGinnis, W. & Krumlauf, R. 1992, "Homeobox genes and axial patterning", *Cell,* vol. 68, no. 2, pp. 283.

McHale, P., Rappel, W. & Levine, H. 2006, "Embryonic pattern scaling achieved by oppositely directed morphogen gradients", *Physical Biology,* vol. 3, no. 2, pp. 107-120.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. 2002, "Network motifs: Simple building blocks of complex networks", Science, vol. 298, no. 5594, pp. 824-827.

Mitarai, N., Jensen, M.H. & Semsey, S. 2015, "Coupled Positive and Negative Feedbacks Produce Diverse Gene Expression Patterns in Colonies", *Mbio,* vol. 6, no. 2, pp. e00059-15.

Payne, S., Li, B., Cao, Y., Schaeffer, D., Ryser, M.D. & You, L. 2013, "Temporal control of self-organized pattern formation without morphogen gradients in bacteria", *Molecular Systems Biology,* vol. 9, pp. UNSP 697.

Perkins, T.J., Jaeger, J., Reinitz, J. & Glass, L. 2006, "Reverse engineering the gap gene network of Drosophila melanogaster", *Plos Computational Biology,* vol. 2, no. 5, pp. 417-428.

Peter, I.S. & Davidson, E.H. 2009, "Genomic control of patterning", International Journal of Developmental Biology, vol. 53, no. 5-6, pp. 707-716.

Reeves, G.T. & Stathopoulos, A. 2009, "Graded Dorsal and Differential Gene Regulation in the Drosophila Embryo", Cold Spring Harbor Perspectives in Biology, vol. 1, no. 4, pp. a000836.

Salis, H.M., Mirsky, E.A. & Voigt, C.A. 2009, "Automated design of synthetic ribosome binding sites to control protein expression", *Nature biotechnology,* vol. 27, no. 10, pp. 946-U112.

Schaerli, Y., Munteanu, A., Gili, M., Cotterell, J., Sharpe, J. & Isalan, M. 2014, "A unified design space of synthetic stripe-forming networks", *Nature Communications,* vol. 5, pp. 4905.

Thieffry, D. & Sanchez, L. 2003, "Dynamical modelling of pattern formation during embryonic development", Current opinion in genetics & development, vol. 13, no. 4, pp. 326-330.

von Dassow, G., Meir, E., Munro, E. & Odell, G. 2000, "The segment polarity network is a robust development module", *Nature,* vol. 406, no. 6792, pp. 188-192.

Wieland, M., Auslaender, D. & Fussenegger, M. 2012, "Engineering of ribozyme-based riboswitches for mammalian cells", *Methods,* vol. 56, no. 3, pp. 351-357.

Wolpert, L. 1969, "Positional information and the spatial pattern of cellular differentiation.", *J Theor Biol,* vol. 25, no. 1, pp. 1-47.

Yen, L., Svendsen, J., Lee, J., Gray, J., Magnier, M., Baba, T., D'Amato, R. & Mulligan, R. 2004, "Exogenous control of mammalian gene expression through modulation of RNA self-cleavage", *Nature,* vol. 431, no. 7007, pp. 471-476.

CHAPTER 2

Transcription Factor Networks

Ashley A. Jermusyk and Gregory T. Reeves

*This chapter is published in *Encyclopedia of Cell Biology*

2.1 - Introduction

The animal body is a complex system which is regulated by a series of interconnecting networks set-up during the initial stages of development. These large transcription factor networks, or gene regulatory networks (GRNs), are composed of sets of smaller networks, or modules, responsible for specific subroutines in the patterning process. For example, in the fruit fly, *Drosophila melanogaster,* two cascading gene regulatory networks serve to pattern the axes of the embryo setting up gene expression domains in the adult fly (Reeves, Stathopoulos 2009, Perkins et al. 2006, Jaeger et al. 2004). Different networks are responsible for correct differentiation later in development. As an example, the GRN map for the dorsal-ventral axis is described in Figure 2.1. This network features many interactions between different transcription factors to properly pattern the major tissues along this axis in the developing embryo.

The final results of these networks can be seen in the correct differentiation of organisms ranging from fruit flies to sea urchins to humans. While these organisms may seem varied, each uses multiple GRNs consisting of similar wiring strategies. At the mechanistic level, all GRNs are built from the ground-up on the same principles: transcription factors binding to specific DNA sequences and regulating proximate genes. These genes in-turn may code for transcription factors that regulate still other genes.

At a higher level of abstraction, collections of a handful of such genetic interactions may be over-represented patterns of functional interactions called motifs. While the identity of the genes themselves differ, network motifs with the same wiring have roughly the same function, and make up the complex gene regulatory networks seen in all of these animals (Reeves, Stathopoulos 2009, Milo et al. 2002, Thieffry, Sanchez

2003, Peter, Davidson 2009a). Some common motifs are: repression, feedforward loops, positive feedback, negative feedback, and cross-repression. These motifs serve as building blocks, with each interaction serving a specific functional role in gene regulation. Together these basic interactions serve to create a properly patterned organism.

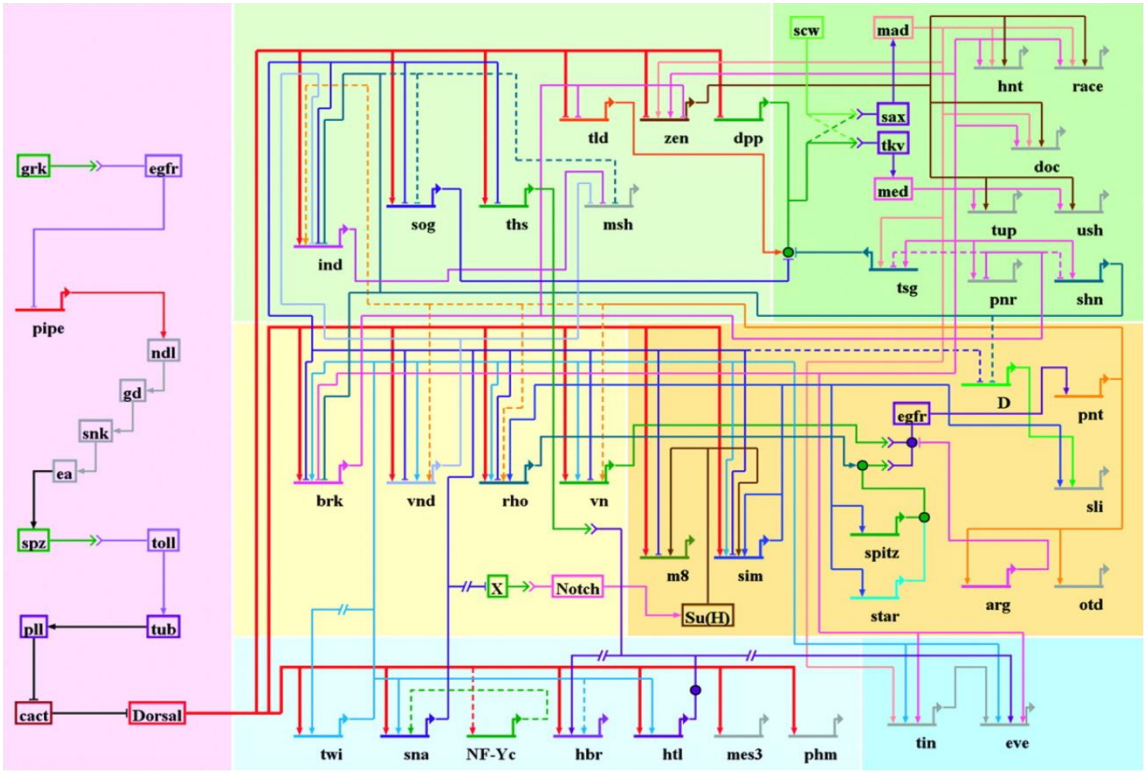

Figure 2.1: GRN map of the system patterning the dorsal-ventral axis in the *Drosophila melanogaster* early embryo from 2 to 5 hr after fertilization. The activity of each transcription factor on the other members of the system is shown. The green, yellow, and blue boxes represent the three types of embryonic tissue. In addition to signaling within a specific tissue type, activity also occurs between tissues. (Reprinted from (Levine, Davidson 2005), ©National Academy of Sciences)

In addition to studying these components of GRNs, more can also be learned by understanding features of the networks as a whole. Global properties of GRNs, such as the small-world property, the scale-free property, and modularity, are responsible for many advantageous design features of GRNs. For example, the GRN depicted in Figure

2.1 is modular in that it is divided into parts that perform separate biological tasks (patterning the individual tissue systems). In this chapter, we will review the common properties of GRNs at each level of abstraction: local, intermediate (motifs), and global. Our focus is on GRNs acting in development, however many of the same features can be applied to other biological contexts, ranging from metabolism to homeostasis.

2.2 - Local Properties of GRNs

At the most basic level, GRNs are composed of transcription factors specifically binding to DNA sequences to regulate nearby genes. Typically, gene regulation is depicted as occurring 5' to a gene's transcriptional start site, but these regulatory elements could be downstream of the gene, as well as within an intron of a gene, or even within a neighboring gene. *cis*-regulatory elements (i.e., the DNA sequences to which transcription factors bind) often contain clusters of several binding sites for distinct transcription factors that each converge to determine the transcriptional state of a gene (Fig. 2.2A). In this section, we'll discuss how these interactions are experimentally verified, and how they are modeled.

2.2.1 - Constructing GRN maps

To construct a GRN map, one must proceed at the local level, with the verification of direct interactions between transcription factor and the regulated gene (Peter, Davidson 2009b). If an organism's genome is sequenced, part of this approach may include a computational search for known binding site sequences. However, a computational approach is rarely sufficient by itself, due to the fact that most binding sites are short, with significant ambiguity at several positions, making the frequency of

purely sequence-based hits prohibitively high, with many false positives. Furthermore, as mentioned above, a gene may be regulated by *cis*-elements that may be a significant distance from the transcriptional start site.

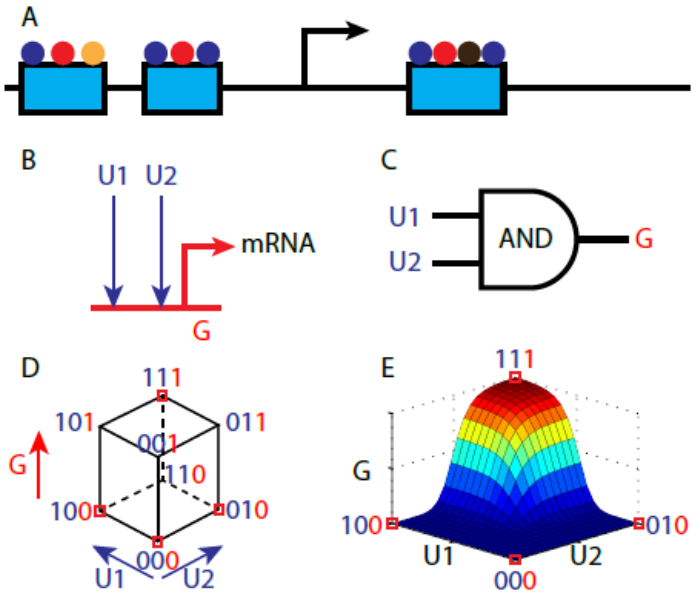

Figure 2.2: Local properties of transcriptional regulation. (A) Structure of *cis*-regulation. *cis*-regulatory elements, represented by blue boxes, contain clusters of binding sites for several transcription factors (colored circles). These clusters can be close to, far from, or even downstream of the transcriptional start site (represented by arrow). (B) Example of gene, G, with two inputs, U1 and U2. (C) Regulation of G can be represented by boolean logic. In this example, both inputs are necessary for gene expression. The logic is represented by an AND gate. (D) Cube representation of the eight possible states for the variables U1, U2, G. Each vertex represents a state in which the variables are either 1 or 0 (inputs in blue, output in red). The only allowable states are indicated by red squares. (E) Continuous representation of gene expresision output. The two inputs can vary continuously from zero to maximal, resulting in a continous variation in G. Often, the input/output relationship is taken to be ultrasensitive, which approximates a logical on/off system. (B-E) modified from Modeling DNA sequence-based cis-regulatory gene networks, Hamid Bolouri and Eric H Davidson, *BioEssays*. (Bolouri, Davidson 2002b) Copyright (c) 2002, Wiley periodicals, Inc.

Several lines of experimental evidence can be employed in verifying a gene regulatory interaction. Direct evidence of transcription factor binding is valuable. One way to find regions bound by transcription factors is ChIP-based experiments. However,

while these data reveal bona-fide physical interactions, it is not always clear which of these interactions are functional, resulting in false positives.

Correlative evidence in gene expression provides a solid foundation for determining the veracity of a genetic interaction. It should be verified the transcription factor is present in the same set of cells as the effect (either transcription or repression of the target gene, depending on the character of the putative regulation). In a similar manner, perturbing transcription factor expression (through knockout mutation, ectopic expression, or RNAi-type knockdown) should have the expected effect on gene expression.

Enhancer trap analysis, which tests the sufficiency of a stretch of DNA to control a gene expression pattern, is one of the most important tools for verifying the role of the DNA sequence in driving gene expression. In this method, the *cis*-regulatory element in question is cloned upstream of a minimal promoter and a marker gene (commonly bacterial *lacZ*, as this gene is not natively present in metazoans). If transgenic expression of this DNA cassette results in the expression of *lacZ* in the pattern of the endogenous gene from which the putative enhancer was derived, this is strong positive evidence the enhancer in question regulates gene expression. Further evidence for the action of a specific transcription factor on this enhancer would include mutating the trancription factor binding sites within the transgenic enhancer and monitoring the resulting change in *lacZ* expression.

In all, the identification of the regulatory interactions of a single gene can be laborious and time-consuming (Yuh, Bolouri & Davidson 2001). In this manner, developmental GRN maps, such as for the sea urchin endomesodermal network, and the

Drosophila dorsal-ventral and anterior-posterior patterning networks, have resulted from many hours of painstaking experiments.

2.2.2 - Modeling cis-regulatory interactions

Once a set of *cis*-regulatory interactions have been experimentally identified for a given gene, how should one predict the effect transcription factor expression has on transcriptional output? Often, these interactions are modeled using boolean on/off logic (Bolouri, Davidson 2002b). From a digital point of view, gene regulation can be seen as a type of logic gate. For example, take a gene, G, regulated by two transcription factors, U1 and U2 (Fig. 2.2B). If this gene requires both activators to be present in order to be activated, this is taken as an AND gate (Fig. 2.2C). The presence (or absence) of the transcription factors sets U1 or U2 to a value of one (or zero), resulting in the output being either on or off (Fig. 2.2D). Another common example is an OR gate, in which the presence of either transcription factor is sufficient for transcriptional activation. If there are competing inputs (if U1 were an activator and U2 a repressor), this could be modeled by logic such as "U1 AND NOT U2".

The digital approach has the advantage of needing to know only the wiring topology and logic of the network; binding affinities and other biophysical parameters are not needed. As such, it is a very coarse-grained approach. When the purely digital logic approach does not have sufficient detail, a continuous approach may be taken (Bolouri, Davidson 2002b). For example, in the case of a series of morphogen-mediated gene patterns, a concentration gradient (analog signal) of a protein is read-out by cells in the tissue to be patterned. In such cases, often the relationship between transcription

factor(s) and transcriptional output may be modeled as a highly-sigmoidal (ultrasensitive) function (Fig. 2.2E). Such models seek only to capture the phenomenology of transcriptional regulation, and most of the biophysical interactions are grouped into a handful of parameters. If the response is sensitive enough, these models approach the switch-like digital case.

If more mechanistic detail is desired, thermodynamic approaches have also been employed, in which the probability of transcription is related to the equilibrium of transcription factor binding site occupancy (Reeves et al. 2006, Zinzen et al. 2006). These models tend to be less switch-like, and require detailed knowledge of the binding affinities and interactions among transcription factors at the *cis*-regulatory locus. Each of these approaches (digital/topological, phenomenological, and mechanistic/thermodynamic) has advantages and should be employed based on the level of input detail known, the level of output detail desired, and computing power available.

2.3 – Motifs

At an intermediate level of abstraction, the interactions among three or more genes can serve a basic functional purpose, regardless of the identity of the genes involved. These network interactions, or motifs, do not often act as functionally separable from the rest of the network, yet nevertheless have well-defined kinetic behavior (Babu et al. 2004). Several motifs have been discovered to be over-represented sets of connections (Milo et al. 2002), implying the functional subroutines carried out by these motifs may have an important, general purpose in regulation of gene expression.

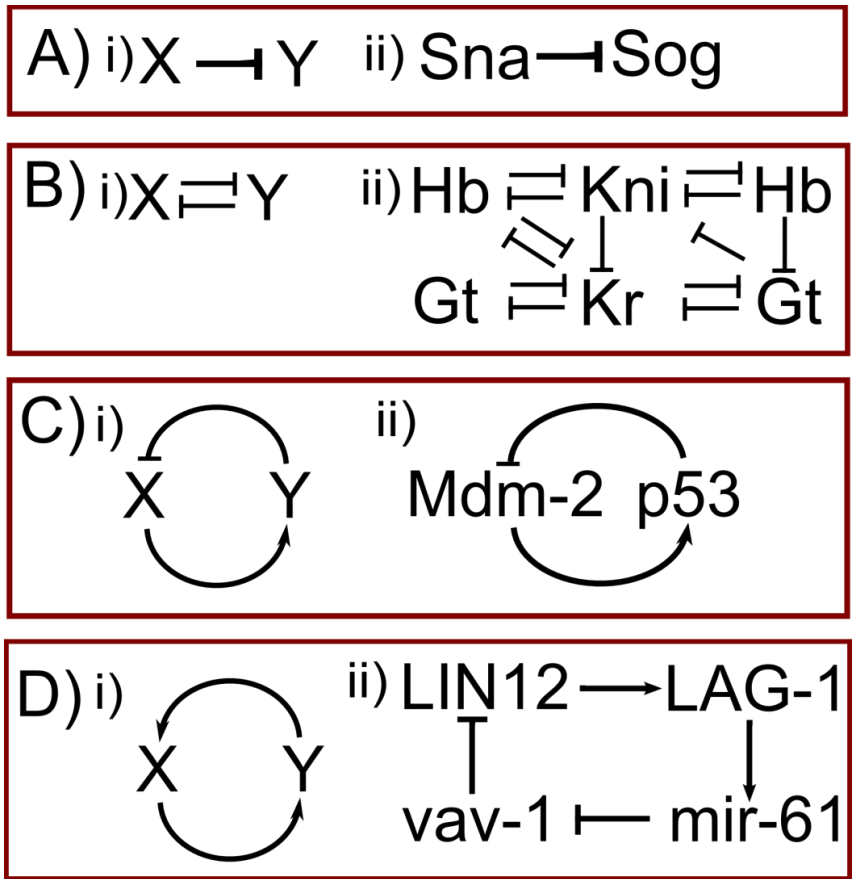

Figure 2.3: Some of the most common motifs including: repression (A), cross-repression (B), negative feedback (C), and positive feedback (D). (A) Repression motif, one protein (X) represses the repression of a second protein (Y), this can be seen in dorsal-ventral patterning in the *Drosophila* embryo between Sna and Sog (Stathopoulos, Levine 2002, Maduro, Rothman 2002). (B) The cross-repression motif features mutual inhibition between two proteins (X and Y) and occurs between Hb, Kni, Gt, and Kr in the anterior-posterior patterning system in the *Drosophila* embryo (Jaeger 2011). (C) Negative feedback features one protein (X) activating a second protein (Y) where the second protein represses the expression or inhibits the function of the first protein; this motif is seen in the tumor regulatory pathway between Mdm-2 and p53 (Oren 1999). (D) Positive feedback is similar to negative feedback except the second protein (Y) activates or facilitates the first protein (X); this can be seen in the vulval formation pathway in *C. elegans* where LIN-12 activates LAG-1, which in turn activates LIN-12 through activation of mir-61 followed by repression of vav-1, which in turn represses LIN-12.

Some common motifs are repression, cross repression, negative feedback, positive feedback, and feedforward loops. These motifs may seem very different, and in fact can be used in very different ways; they all function to increase the ability of an organism to create reproducible patterns despite perturbations and obstacles to proper development.

In each of these systems, transcription factors bind to other transcription factors or to *cis*-regulatory elements to activate or repress expression of other transcription factors. These interactions are seen throughout development in a variety of systems. Our understanding of how these interactions can be grouped into different motifs to produce a given result in one system can be applied more broadly to a similar group of interactions, or the same motif, in another system.

2.3.1 – Repression

Perhaps the simplest motif found during development is repression. During patterning, morphogen gradients activate expression in large domains based on concentration thresholds. In many instances, expression domains must be restricted to limited regions, which can be achieved through repression by a neighboring protein (Stathopoulos, Levine 2002, Maduro, Rothman 2002). For example, along the dorsal-ventral axis of the *Drosophila* embryo, Snail is expressed in a narrow domain at the ventral midline and represses *short gastrulation* (*sog*) (Stathopoulos, Levine 2002). This repression by Snail eliminates *sog* expression in the ventral-most cells, where Snail is present, dividing *sog* expression into two broad lateral stripes (Stathopoulos, Levine 2002). Just like in the *Drosophila* embryo, *C. elegans* need to control the spatial domain of gene expression in endoderm development (Maduro, Rothman 2002). In this system, SKN-1 activates expression, whereas POP-1 represses expression, creating a switch that serves to direct tissue to either an endoderm or mesoderm fate, thereby limiting the boundaries of the mesoderm (Maduro, Rothman 2002). Similar regulation occurs in patterning the *Drosohila* egg and heart just to name a few of the diverse uses for

repression (Cripps, Olson 2002, Wasserman, Freeman 1998). In each of these instances an input signal activates expression over a large region, and repression is used to eliminate this expression in certain region or after a certain period of time.

2.3.2 - Cross-repression

Mutual inhibition, or cross repression, can be used to produce stripes or clear borders between gene expression regions in many different organisms. In the developing *Drosophila* embryo, one of the first steps to patterning of the anterior-posterior axis is the proper positioning of the so-called gap genes, namely Knirps, Hunchback, Giant, and Krüppel (Jaeger 2011). The initial expression of these gap genes are established by the morphogen Bicoid in large overlapping domains (Jaeger 2011). Cross repression between neighboring genes refines these domains and creates sharp borders between the domains (Jaeger 2011). Since each region of transcription factor inhibits the neighboring transcription factor, two stable regions of expression are produced, which easily respond to perturbations near the interface of expression (Jagla et al. 2002). Similar cross repression activity occurs in the specification of cardiac and muscle cell fates in the *Drosophila* embryo (Jagla et al. 2002). These sharp domain borders are needed to create clear limits on downstream gene expression and therefore delineate positioning of tissue differentiation.

Mutual inhibition also regulates the branching of cells in mammalian lungs, mammary epithelial cells, trachea in flies, and vasculature in flies (Lu, Werb 2008). In the fly trachea, mutual inhibition regulates the number of branches by determining the leading cell through competition for branch-inducing factors (Lu, Werb 2008). Similarly,

in mammary epithelial cells, mutual inhibition functions as a mechanism for cells to recognize surrounding cells and avoid other branches; this results in new branches turning away from the current line of growth and branches to end as they approach another branch (Lu, Werb 2008). The use of mutual inhibition by branching cells functions similarly to the earlier examples of defining borders of stripes of gene expression in the developing embryo. In these diverse systems, mutual inhibition or cross repression provides a method for cells to recognize neighboring cells and to stabilize the interactions between these cells by controlling the position and number of branches in a tissue or by controlling the downstream signal a region will receive prior to tissue differentiation.

2.3.2 - Negative Feedback

Negative feedback can be used to buffer a system against perturbations, therefore making it more robust (Ma et al. 2009). In mice, a negative feedback loop featuring Mdm-2 regulates the degradation of *p53*, a tumor suppressor gene (Oren 1999). In this loop, p53 binds to the *mdm-2* gene, stimulating the production of Mdm-2 protein, which then binds to p53 to inactivate and stimulate p53 degradation (Oren 1999). The result of this contradictory relationship is a tight control over p53 expression, which is important since improper control of *p53* can result in cancer (Oren 1999). A similar negative feedback loop exists in embryonic stem cells to control pluripotency and self-renewal (Pan et al. 2006). In these systems negative feedback serves to limit gene expression thereby adding a level of control to the system; the importance of this control is apparent in the prevention of cancer, but is also important in a number of different systems.

Alternatively, oscillations can also be produced using negative feedback. Oscillations result in periodic expression of genes in time or space. This can serve as a method to produce evenly spaced stripes in a field as the tissue is created from replicating cells. During vertebrate somitogenesis, multiple negative feedback loops produce oscillations in Fgf, Wnt, and Notch which control the segmentation clock mechanism (Aulehla, Herrmann 2004, Gibb, Maroto & Dale 2010). This creates periodic gene expression, which results in each segment forming at specific time intervals and therefore in specific positions in the tissue (Aulehla, Herrmann 2004, Gibb, Maroto & Dale 2010). This oscillatory behavior is especially important in tissue that experiences directional development, creating stripes of expression as the tissue is created.

In some instances it can be desirable to limit the length of time or range of a signal; this effect can also be accomplished using negative feedback. In this case when a threshold is reached the signal can be turned off using a negative regulator; this type of regulation occurs in the JAK/STAT pathway (Janus kinase/signal transducers and activators of transcription) in vertebrates (Rawlings, Rosler & Harrison 2004). This provides a valuable method to control signaling by turning off expression as a new stage in development is reached, it can also be used in regulation of tissue to turn-off a signal after the perturbation to the system has been removed.

Negative feedback can be used to achieve multiple goals. Most notably is to buffer the system against perturbations by interlocking the expression of two genes and thereby preventing fluctuations from being carried over into downstream gene expression. To create properly patterning tissues, negative feedback, through

oscillations, can create reproducible bands of gene expression. These are just a few of the many ways negative feedback is used to produce properly differentiated tissues.

2.3.3 - Positive Feedback

Positive feedback can also provide robustness to perturbations to the system. This robustness does not sacrifice sensitivity to the input signal (Hornung, Barkai 2008, Kadelka, Murrugarra & Laubenbacher 2013). In many systems switch-like behavior is used to achieve this robustness and produce precise borders, by creating bistable systems (Shah, Sarkar 2011, Siegal-Gaskins et al. 2011, Graham et al. 2010). In bistable systems two equilibrium states can be achieved, with a small perturbation, the system will settle into a different state. This can be seen in endomesoderm specification in sea urchins (Minokawa, Wikramanayake & Davidson 2005, Yuh et al. 2004). In this system, initial β-catenin activates *wnt8* and *blimp1/krox* expression; this in turn causes nuclearization of β-catenin and therefore more *wnt8* and *blimp1/krox* to be expressed (Minokawa, Wikramanayake & Davidson 2005). In this case, relatively low levels of β-catenin are needed to create a stable domain of Wnt8 expression (Minokawa, Wikramanayake & Davidson 2005). Positive feedback is used to create bistable systems in other systems such as the mouse hindbrain, *Drosophila* wing, and *Xenopus* oocyte (Barrow, Stadler & Capecchi 2000, Nonchev et al. 1996a, Nonchev et al. 1996b, Vesque et al. 1996, Yan et al. 2009, Xiong, Ferrell 2003).

In a variety of organisms, similar systems exist where positive feedback produces precise regions of expression following an input signal. This can be seen, for instance in the control of vulval formation in *C. elegans* development (Yoo, Greenwald 2005). In

this system LIN-12 activates LAG-1 expression which goes on to activate mir-61 expression (Yoo, Greenwald 2005). In turn, mir-61 represses *vav-1* which is a repressor of *lin-12* (Yoo, Greenwald 2005). While varying in construction, positive feedback loops occur in many different systems in nature, including *C. elegans* neuron differentiation, *Drosophila* gut patterning, *Drosophila* eye development, and the interferon pathway in animal immune response (Johnston et al. 2005, BIENZ, TREMML 1988, Johnston et al. 2011, Taniguchi, Takaoka 2001). In each of these cases positive feedback is able to use relatively small initial signals to effectively pattern tissue from stable domains of gene expression.

2.3.4 - Feedforward Loops

Feedforward loops can help a system to respond to perturbations. Feedforward loops come in two flavors: coherent and incoherent. For coherent feedforward loops, the direct effect gene X has on gene Z is the same as the effect of gene X on gene Z indirectly through gene Y, as shown in Figure 2.4. With incoherent feedforward loops, it is the opposite. Incoherent feedforward loops can be very useful to reset a system after it responds to a perturbation (Ma et al. 2009). These incoherent feedforward loops are found to regulate multiple systems in yeast and *E. coli* (Milo et al. 2002, Shen-Orr et al. 2002).

Incoherent feedforward loops can also be used to produce stripes in many different systems including the *Drosophila* notum and *Drosophila* eggshell (Hironaka, Iwasa & Morishita 2012, Yakoby et al. 2008). The feedforward loop allows for gene expression far from the source of the activation signal, at low morphogen levels

(Hironaka, Iwasa & Morishita 2012). In the *Drosophila* notum, the morphogen Decapentaplegic (Dpp) activates Pannier (Pnr) and U-shaped (Ush) production (Hironaka, Iwasa & Morishita 2012). Pnr and Ush can form a complex to repress *wingless* (*wg*) expression or unbound Pnr can activate *wg* expression (Hironaka, Iwasa & Morishita 2012). Together these two paths combine to form a stripe of Wg at low levels of Dpp (Hironaka, Iwasa & Morishita 2012). In each of these cases, the feedforward loop allows for gene expression at low levels of the input signal at the periphery of the differentiating field.

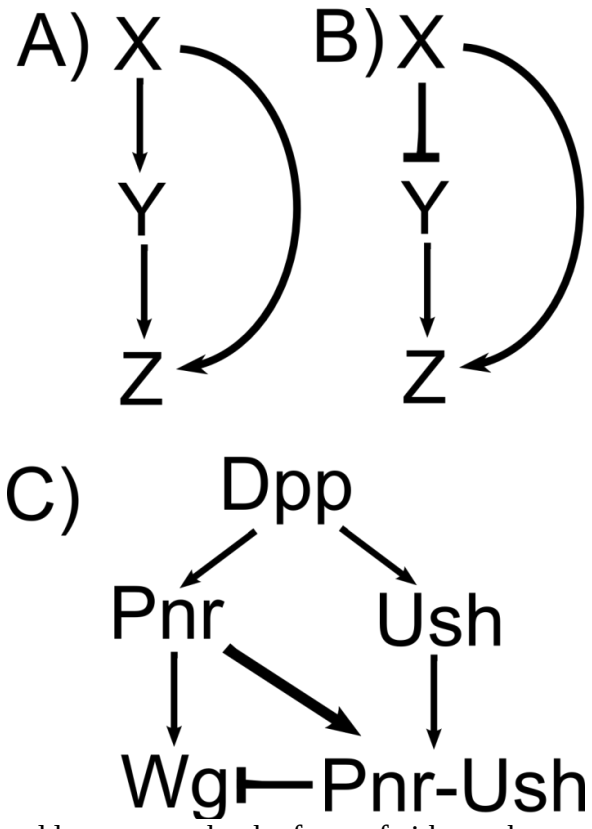

Figure 2.4: Feedforward loops can take the form of either coherent (A) or incorehent (B). In a coherent feedforward loop, the direct effect gene X has on gene Z is the same as the effect of gene X on gene Z indirectly through gene Y. This can be accomplished in a simple loop where X activates Y and Z, and Y goes on to activate Z (A). However in the incoherent feedforward loop, the effect gene X has on gene Z indirectly through gene Y is opposite the effect it has on gene Z directly. This can be accomplished by altering the network such that X represses Y (B). (C) An incoherent feedforward loop regulates Decapentaplegic regulation of Wingless in the *Drosophila* notum

In some instances an initial signal is only available at the start of development. In these cases it is desirable to amplify this signal and provide a sustained signal output. This effective signal duration can be achieved using a feedforward loop. In patterning the dorsal-ventral axis of the developing *Drosophila* embryo, a combination of coherent and incoherent feedforward loops signal from Dorsal through either Snail or Twist to control expression of multiple genes in the neurogenic ectoderm (Zartman, Shvartsman 2007). In each of these cases the feedforward signal is used because it creates a mechanism to utilize a small input signal and transform it.

2.4 - Global Properties of GRNs

The global properties exhibited by GRNs are largely dictated by the topology of the wiring of the network. Studies of the global properties of GRNs have often visualized GRNs as "directed graphs," with genes being the nodes, and regulation of one gene on another being a directed connection. By and large, GRNs have been found to adhere to several general topological features, meaning that GRNs studied to date display many common properties. In this section, we will review three global topological features -- the small-world property, the scale-free property, and modularity -- and discuss what each these features imply.

2.4.1- Small-world networks

Many biological networks – including metabolic networks, neural networks, as well as GRNs – have been found to be of the "small-world" character (Albert 2005). Several types of human networks also exhibit this property, such as social networks,

power grids, and connections among actors in Hollywood films (Watts, Strogatz 1998b). Small-world networks are ones in which any two nodes are connected by a relatively short series of paths, yet the probability of any two given nodes being directly connected to each other is relatively high. In other words, small-world networks achieve a balance between maintaining a low degree of separation (commonly referred to as "six degrees" of separation) and having a high degree of clustering (or "cliquishness").

In contrast to small-world networks, one may have a completely regular network, or lattice network, in which all nodes are connected only to their ℓ nearest neighbors. This network has a high degree of clustering, but to move from one node to another on the other side of the network requires many local steps. One simple way to construct a small-world network is to begin with a lattice network, and make a few "short circuit" connections across the network (Watts, Strogatz 1998b) (see Fig. 2.5A,B).

On the other hand, one may have a completely random network, where the probability of two nodes being directly connected is uniform across all nodes. In such a network, the idea of clustering is completely lost, yet the shortest distance between any two nodes is small, allowing one to cross from one node to another in a relatively short time.

Small-world networks have the advantage of keeping both qualities of local clustering and global connectivity. This allows for rapid propagation of information among all nodes in the network, yet also allows for groups of nodes to act together in a similar, local task.

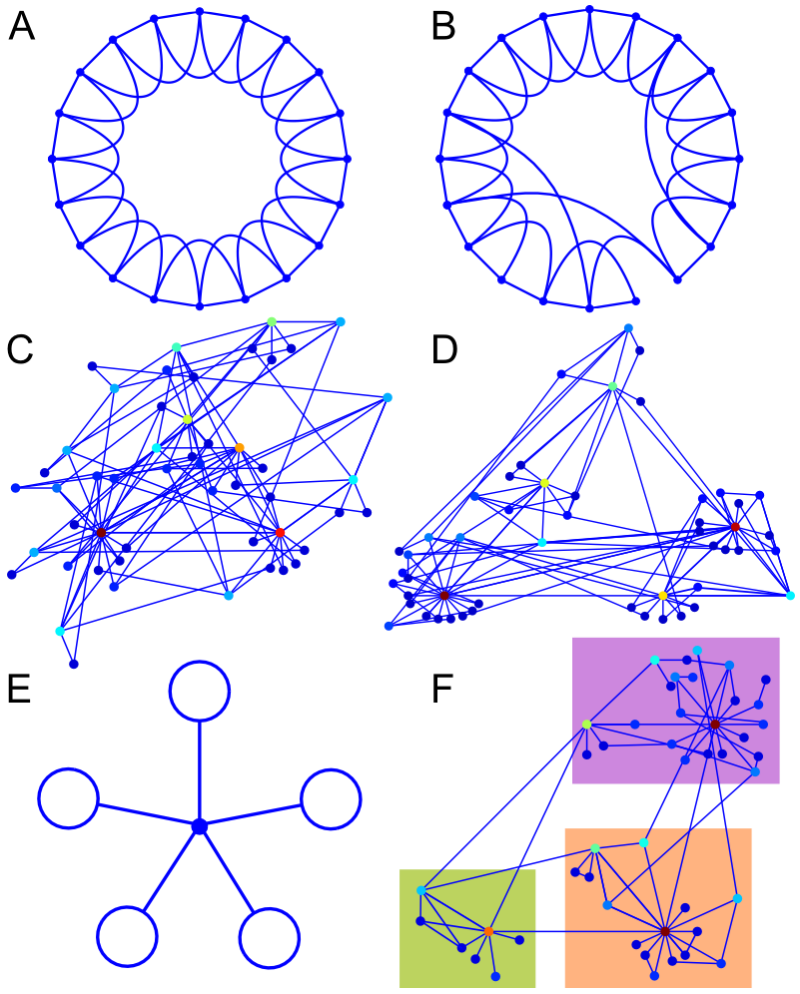

Figure 2.5: Illustrations of the small-world, scale-free, and modular properties of a network. (A) Lattice network. Each node is connected to its nearest and next-nearest neighbors. (B) Small-world network, made from randomly rewiring three connections in the lattice network. Because the vast majority of nodes have four connections, this is not a scale-free network. (C) Scale-free network. Most nodes have very few connections, but there is a significant number of nodes with many connections. However, this scale-free network was randomly wired, so has very low clustering, and thus is not small-world. Node color code: From dark red (most connections), through red, orange, yellow, green, blue, and finally to dark blue (fewest connections). (D) Scale-free and small-world network. Connections in this network were made preferentially so that clustering is high (a node's neighbors will also likely be connected to each other). However, cross-connections between local groups are too high for this network to be considered modular. Color code same as in (C). (E) Modular, lattice network. Each open circle at the end of a spoke represents a lattice network as in (A). The regular lattice structure rules out this network from being small-world or scale-free, even though it is modular. (F) Small-world, scale-free, modular network. Clustered groups of nodes can be grouped into modules (light green, light orange, and light purple) that have relatively few connections between them. Color code same as in (C). (A) and (B) adapted by permission from Macmillan Publishers Ltd: *Nature* (Watts, Strogatz 1998a), Copyright (c) 1998.

2.4.2 - Scale-free networks

Besides focusing on the paths and connections between nodes, another way to characterize a network is to focus on how many connections each node has. Early network studies assumed the number of node connections conformed to a Poisson distribution, and thus was strongly peaked at an average value <*k*>. This implied that the number of nodes with more than <*k*> connections decayed exponentially and that there were few if any nodes with many more than <*k*> connections. In that sense, the "scale" of the network was <*k*> connections per node.

However, as storage and computing power increased, and topological data of large networks became more available, it was understood that few if any naturally-occurring or man-made networks followed a Poisson distribution in node connectivity. Instead, the probability that a node has *k* connections decays as a power law: $P(k) = k^{-n}$ (Barabasi, Albert & Jeong 1999). This result implied that there were a comparatively large number of nodes with higher-than-average connectivity. In this sense, there was no strongly preferred number of node connections, and thus such networks have been dubbed "scale-free" (Barabasi, Albert & Jeong 1999).

Intuitively, a scale-free network is one in which most nodes are connected, on average, to a small number of other nodes. However, there is a small but significant number of nodes that are highly connected (see Fig. 2.5C). These "hubs" hold most of the network together, and at a higher level of connectivity, there are still larger hubs that connect the smaller hubs together. In this sense, the scale-free property has much overlap with the small-world property, but they are not identical. For example, both types of networks display a relatively low degree of separation. In scale-free networks, this

property is provided by hubs, or master regulators, that connect parts of the network together, whereas in small-world networks, this property could be achieved simply by making a few random connections across the network. (Watts, Strogatz 1998b, Barabasi, Albert & Jeong 1999). On the other hand, a necessary component of a small-world network is local clustering, yet this property is not required to be a scale-free network. For example, compare the two scale-free networks in Fig. 2.5C and D .

As with the small-world property, there is a long list of both man-made and biological networks that exhibit the scale-free property, including GRNs. The scale-free property is advantageous for these networks, as it imparts a high degree of fault- and failure-tolerance (Albert and Jeong and Barabasi 2000, Jeong et al. 2000). This is because if a gene is randomly disabled by mutation, it is likely to have been a peripheral gene rather than a highly-connected hub. On the other hand, even mutations to single hubs can sometimes be tolerated, due to the connectedness of the network through the rest of the hubs. This property is also advantageous from a medical perspective. For example, if several hubs can be intelligently targeted, such as in cancer therapy, it is possible to completely destroy network operation. None of this is to say that mutating a single essential gene, whether peripheral or highly-connected, will not have disastrous consequences; indeed, that is the natural consequence of the fact that genes have important functions in the cell. Instead, the point is a scale-free network is designed in such a way as to avoid cascading failures, from the network point of view.

2.4.3 – Modularity

As we have discussed before, sets of interactions between a handful of genes are typically characterized as "motifs", and the topology of the motif determines its kinetic behavior. At a higher level of abstraction, several motifs combine together to form systems of motifs called modules. The modularity of GRNs means that small "communities" of genes have a high level of interconnectedness and together perform a related function, such cellular differentiation. These modules are further connected (sparsely) to each other to compose the entire GRN.

It is clear that the modularity of a GRN is related to the scale-free property, as the hubs in a scale-free sense serve as master regulators connecting several modules. However, these properties are not identical. It is possible to have a scale-free network in which local clustering is not maintained, resulting in a non-modular (and non-small-world) network (Fig. 2.5C). Even in scale-free networks with high degrees of clustering, it may be difficult to identify separate modules if there is a significant degree of cross-connectivity among groups (Fig. 2.5D). Modular networks may also violate the small-world property. The network in Fig. 2.5E is modular, but is neither scale-free nor small-world. In contrast, the network depicted in Fig. 2.5F is a good example of a scale-free, small-world, modular network: each highlighted group of nodes is tightly clustered, largely separate from the others, and is centered on a few highly-connected hubs, yet there are enough connections among the modules to preserve a low degree of separation.

It is also important to note the modularity of a network depends not only on its connectivity, but also on the specific function of the genes in the local cluster. Local clusters of genes must act together in a specific function for a network to be considered

modular. For example, in sea urchin development, there are distinct modules, composed of groups of interacting genes, that separately control endoderm, mesoderm, and ectoderm development (Peter, Davidson 2009c). It is evident that genes within a developmental module must be co-expressed in a subset of cells at a particular stage of development.

In the realm of human-designed systems, building objects with modularity is a well-established engineering principle. As such, the modularity of GRNs reveals an analogy to engineered systems. For example, human engineers design hardware (or software) subsystems, with self-contained purposes, that can be "swapped out" for different applications, such as the modular bay on a laptop, which can hold a DVD drive, an extra hard drive, or a floppy disk drive. Thus, for GRNs, modularity provides adaptability, in which a subset of the GRN can be slightly modified without largely affecting the rest of the network. As a side benefit, the modularity of GRNs allows for more facile study by curious scientists.

2.5 – Conclusions

During development organisms utilize networks of interacting genes, known as GRNs, to properly pattern the organism. Several different properties can be used to describe the overall construction of these GRNs, such as modularity, scale-free, and small-world. In addition to looking at these GRNs as a whole, more can be learned about these networks by looking at the smaller functional elements, known as motifs, which make up these GRNs. These motifs accomplish a diverse range of specific goals within the larger system. These goals include: responding to perturbations, amplifying a short-

term signal, creating sharp borders between expression regions, and producing oscillations. These differing goals can be accomplished using motifs such as repression, cross-repression negative feedback, positive feedback, and feedforward loops. Each motif can be used in differing ways to meet different goals depending on the specific set-up of the individual network. In addition to be used for multiple goals, the same motifs are used by many different systems and occur reiteratively throughout nature to pattern tissues in many different organisms. Because of this, the information learned about these motifs in one organism can be applied to other animals. This provides a valuable tool in understanding gene regulation in general as well as in our particular organism of interest. Thereby allowing us to better utilize and extrapolate findings from previous research and inform our new studies.

References

Albert and Jeong and Barabasi 2000, "Error and attack tolerance of complex networks", Nature, vol. 406, no. 6794, pp. 378-382.

Albert, R. 2005, "Scale-free networks in cell biology.", J Cell Sci, vol. 118, no. Pt 21, pp. 4947-4957.

Aulehla, A. & Herrmann, B. 2004, "Segmentation in vertebrates: clock and gradient finally joined", Genes & development, vol. 18, no. 17, pp. 2060-2067.

Babu, M., Luscombe, N., Aravind, L., Gerstein, M. & Teichmann, S. 2004, "Structure and evolution of transcriptional regulatory networks", Current opinion in structural biology, vol. 14, no. 3, pp. 283-291.

Barabasi, A., Albert, R. & Jeong, H. 1999, "Mean-field theory for scale-free random networks", Physica a, vol. 272, no. 1-2, pp. 173-187.

Barrow, J., Stadler, H. & Capecchi, M. 2000, "Roles of Hoxa1 and Hoxa2 in patterning the early hindbrain of the mouse", Development, vol. 127, no. 5, pp. 933-944.

Bienz, M. & Tremml, G. 1988, "Domain of Ultrabithorax Expression in Drosophila Visceral Mesoderm from Auto-Regulation and Exclusion", Nature, vol. 333, no. 6173, pp. 576-578.

Bolouri, H. & Davidson, E.H. 2002a, "Modeling DNA sequence-based cis-regulatory gene networks.", Dev Biol, vol. 246, no. 1, pp. 2-13.

Bolouri, H. & Davidson, E.H. 2002b, "Modeling transcriptional regulatory networks.", Bioessays, vol. 24, no. 12, pp. 1118-1129.

Cotterell, J. & Sharpe, J. 2010, "An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients", Molecular Systems Biology, vol. 6, pp. 425.

Cripps, R. & Olson, E. 2002, "Control of cardiac development by an evolutionarily conserved transcriptional network", Developmental biology, vol. 246, no. 1, pp. 14-28.

Davidson, E.H. & Levine, M.S. 2008, "Properties of developmental gene regulatory networks.", Proc Natl Acad Sci U S A, vol. 105, no. 51, pp. 20063-20066.

Davidson, E.H., McClay, D.R. & Hood, L. 2003, "Regulatory gene networks and the properties of the developmental process.", Proc Natl Acad Sci U S A, vol. 100, no. 4, pp. 1475-1480.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C., Amore, T.M.a.G., Hinman, V., Otim, C.A.a.O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z.j., Schilstra, M.J., Clarke, P.J.C., Arnone, M.I., Cameron, L.R.a.R.A., McClay, D.R., Hood, L. & Bolouri, H. 2002, "A genomic regulatory network for development.", Science, vol. 295, no. 5560, pp. 1669-1678.

Gibb, S., Maroto, M. & Dale, J.K. 2010, "The segmentation clock mechanism moves up a notch", Trends in cell biology, vol. 20, no. 10, pp. 593-600.

Graham, T.G.W., Tabei, S.M.A., Dinner, A.R. & Rebay, I. 2010, "Modeling bistable cell-fate choices in the Drosophila eye: qualitative and quantitative perspectives", Development, vol. 137, no. 14, pp. 2265-2278.

Hironaka, K., Iwasa, Y. & Morishita, Y. 2012, "Multiple feedback loops achieve robust localization of wingless expression in Drosophila notum development", Journal of theoretical biology, vol. 292, pp. 18-29.

Holme, P. & Kim, B. 2002, "Growing scale-free networks with tunable clustering", Phys.Rev.E, vol. 65, no. 2, pp. 0261071-0261074.

Hornung, G. & Barkai, N. 2008, "Noise propagation and signaling sensitivity in biological networks: A role for positive feedback", Plos Computational Biology, vol. 4, no. 1, pp. e8.

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K., Manu, Myasnikova, E., Vanario-Alonso, C., Samsonova, M., Sharp, D. & Reinitz, J. 2004, "Dynamic control of positional information in the early Drosophila embryo", Nature, vol. 430, no. 6997, pp. 368-371.

Jaeger, J. 2011, "The gap gene network", Cellular and Molecular Life Sciences, vol. 68, no. 2, pp. 243-274.

Jagla, T., Bidet, Y., Da Ponte, J., Dastugue, B. & Jagla, K. 2002, "Cross-repressive interactions of identity genes are essential for proper specification of cardiac and muscular fates in Drosophila", Development, vol. 129, no. 4, pp. 1037-1047.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. & BarabÃ¡si, A.L. 2000, "The large-scale organization of metabolic networks.", Nature, vol. 407, no. 6804, pp. 651-654.

Johnston, R., Chang, S., Etchberger, J., Ortiz, C. & Hobert, O. 2005, "MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision", Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 35, pp. 12449-12454.

Johnston, R.J., Jr., Otake, Y., Sood, P., Vogt, N., Behnia, R., Vasiliauskas, D., McDonald, E., Xie, B., Koenig, S., Wolf, R., Cook, T., Gebelein, B., Kussell, E., Nakagoshi, H. & Desplan, C. 2011, "Interlocked Feedforward Loops Control Cell-Type-Specific Rhodopsin Expression in the Drosophila Eye", Cell, vol. 145, no. 6, pp. 956-968.

Kadelka, C., Murrugarra, D. & Laubenbacher, R. 2013, "Stabilizing gene regulatory networks through feedforward loops", Chaos, vol. 23, no. 2, pp. 025107.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J., Volkert, T.L., Fraenkel, E., Gifford, D.K. & Young, R.A. 2002, "Transcriptional regulatory networks in Saccharomyces cerevisiae.", Science, vol. 298, no. 5594, pp. 799-804.

Levine, M. & Davidson, E. 2005, "Gene regulatory networks for development", Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 14, pp. 4936-4942.

Lu, P. & Werb, Z. 2008, "Patterning Mechanisms of Branched Organs", Science, vol. 322, no. 5907, pp. 1506-1509.

Ma, W., Trusina, A., El-Samad, H., Lim, W.A. & Tang, C. 2009, "Defining Network Topologies that Can Achieve Biochemical Adaptation", Cell, vol. 138, no. 4, pp. 760-773.

Maduro, M. & Rothman, J. 2002, "Making worm guts: The gene regulatory network of the Caenorhabditis elegans endoderm", Developmental biology, vol. 246, no. 1, pp. 68-85.

Mangan, S. & Alon, U. 2003, "Structure and function of the feed-forward loop network motif.", Proc Natl Acad Sci U S A, vol. 100, no. 21, pp. 11980-11985.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. 2002, "Network motifs: Simple building blocks of complex networks", Science, vol. 298, no. 5594, pp. 824-827.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R.a.S.,Shai, Ayzenshtat, I., Sheffer, M. & Alon, U. 2004, "Superfamilies of evolved and designed networks.", Science, vol. 303, no. 5663, pp. 1538-1542.

Minokawa, T., Wikramanayake, A. & Davidson, E. 2005, "Cis-Regulatory Inputs of the Wnt8 Gene in the Sea Urchin Endomesoderm Network", Developmental biology, vol. 288, no. 2, pp. 545-558.

Nonchev, S., Maconochie, M., Vesque, C., Aparicio, S., ArizaMcNaughton, L., Manzanares, M., Maruthainar, K., Kuroiwa, A., Brenner, S., Charnay, P. & Krumlauf, R. 1996a, "The conserved role of Krox-20 in directing Hox gene expression during vertebrate hindbrain segmentation", Proceedings of the National Academy of Sciences of the United States of America, vol. 93, no. 18, pp. 9339-9345.

Nonchev, S., Vesque, C., Maconochie, M., Seitanidou, T., ArizaMcNaughton, L., Frain, M., Marshall, H., Sham, M., Krumlauf, R. & Charnay, P. 1996b, "Segmental

expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20", Development, vol. 122, no. 2, pp. 543-554.

Oltvai, Z.N. & Barab\'asi, A. 2002, "Systems biology. Life's complexity pyramid.", Science, vol. 298, no. 5594, pp. 763-764.

Oren, M. 1999, "Regulation of the p53 tumor suppressor protein", Journal of Biological Chemistry, vol. 274, no. 51, pp. 36031-36034.

Pan, G., Li, J., Zhou, Y., Zheng, H. & Pei, D. 2006, "A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal", Faseb Journal, vol. 20, no. 10, pp. 1730-+.

Perkins, T.J., Jaeger, J., Reinitz, J. & Glass, L. 2006, "Reverse engineering the gap gene network of Drosophila melanogaster", Plos Computational Biology, vol. 2, no. 5, pp. 417-428.

Peter, I.S. & Davidson, E.H. 2009a, "Genomic control of patterning", International Journal of Developmental Biology, vol. 53, no. 5-6, pp. 707-716.

Peter, I.S. & Davidson, E.H. 2009b, "Modularity and design principles in the sea urchin embryo gene regulatory network", FEBS letters, vol. 583, no. 24, pp. 3948-3958.

Peter, I.S. & Davidson, E.H. 2009c, "Modularity and design principles in the sea urchin embryo gene regulatory network.", FEBS Lett, vol. 583, no. 24, pp. 3948-3958.

Rawlings, J., Rosler, K. & Harrison, D. 2004, "The JAK/STAT signaling pathway", Journal of cell science, vol. 117, no. 8, pp. 1281-1283.

Reeves, G.T., Muratov, C.B., Schupbach, T. & Shvartsman, S.Y. 2006, "Quantitative models of developmental pattern formation", Developmental Cell, vol. 11, no. 3, pp. 289-300.

Reeves, G.T. & Stathopoulos, A. 2009, "Graded Dorsal and Differential Gene Regulation in the Drosophila Embryo", Cold Spring Harbor Perspectives in Biology, vol. 1, no. 4, pp. a000836.

Shah, N.A. & Sarkar, C.A. 2011, "Robust Network Topologies for Generating Switch-Like Cellular Responses", Plos Computational Biology, vol. 7, no. 6, pp. e1002085.

Shen-Orr, S., Milo, R., Mangan, S. & Alon, U. 2002, "Network motifs in the transcriptional regulation network of Escherichia coli", Nature genetics, vol. 31, no. 1, pp. 64-68.

Siegal-Gaskins, D., Mejia-Guerra, M.K., Smith, G.D. & Grotewold, E. 2011, "Emergence of Switch-Like Behavior in a Large Family of Simple Biochemical Networks", Plos Computational Biology, vol. 7, no. 5, pp. e1002039.

Stathopoulos, A. & Levine, M. 2002, "Dorsal gradient networks in the Drosophila embryo", Developmental biology, vol. 246, no. 1, pp. 57-67.

Stathopoulos, A. & Levine, M. 2005, "Genomic regulatory networks and animal development.", Dev Cell, vol. 9, pp. 449-462.

Taniguchi, T. & Takaoka, A. 2001, "A weak signal for strong responses: Interferon-alpha/beta revisited", Nature Reviews Molecular Cell Biology, vol. 2, no. 5, pp. 378-386.

Telesford, Q.K., Joyce, K.E., Hayasaka, S., Burdette, J.H. & Laurienti, P.J. 2011, "The ubiquity of small-world networks.", Brain Connect, vol. 1, no. 5, pp. 367-375.

Thieffry, D. & Sanchez, L. 2003, "Dynamical modelling of pattern formation during embryonic development", Current opinion in genetics & development, vol. 13, no. 4, pp. 326-330.

Vesque, C., Maconochie, M., Nonchev, S., ArizaMcNaughton, L., Kuroiwa, A., Charnay, P. & Krumlauf, R. 1996, "Hoxb-2 transcriptional activation in rhombomeres 3 and 5 requires an evolutionarily conserved cis-acting element in addition to the Krox-20 binding site", Embo Journal, vol. 15, no. 19, pp. 5383-5396.

Wasserman, J. & Freeman, M. 1998, "An autoregulatory cascade of EGF receptor signaling patterns the Drosophila egg", Cell, vol. 95, no. 3, pp. 355-364.

Watts, D.J. & Strogatz, S.H. 1998a, "Collective dynamics of 'small-world' networks.", Nature, vol. 393, no. 6684, pp. 440-442.

Watts, D. & Strogatz, S. 1998b, "Collective dynamics of 'small-world' networks", Nature, vol. 393, no. 6684, pp. 440-442.

Xiong, W. & Ferrell, J. 2003, "A positive-feedback-based bistable 'memory module' that governs a cell fate decision", Nature, vol. 426, no. 6965, pp. 460-465.

Yakoby, N., Lembong, J., Schupbach, T. & Shvartsman, S.Y. 2008, "Drosophila eggshell is patterned by sequential action of feedforward and feedback loops", Development, vol. 135, no. 2, pp. 343-351.

Yan, S., Zartman, J.J., Zhang, M., Scott, A., Shvartsman, S.Y. & Li, W.X. 2009, "Bistability coordinates activation of the EGFR and DPP pathways in Drosophila vein differentiation", Molecular Systems Biology, vol. 5, pp. 278.

Yoo, A. & Greenwald, I. 2005, "LIN-12/Notch activation leads to micro RNA-mediated down-regulation of vav in C-elegans", Science, vol. 310, no. 5752, pp. 1330-1333.

Yuh, C.H., Bolouri, H. & Davidson, E.H. 2001, "Cis-regulatory logic in the endo16 gene: switching from a specification to a differentiation mode of control.", Development, vol. 128, no. 5, pp. 617-629.

Yuh, C., Dorman, E., Howard, M. & Davidson, E. 2004, "An otx cis-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network", Developmental biology, vol. 269, no. 2, pp. 536-551.

Zartman, J.J. & Shvartsman, S.Y. 2007, "Enhancer organization: Transistor with a twist or something in a different vein?", Current Biology, vol. 17, no. 24, pp. R1048-R1050.

Zinzen, R.P., Senger, K., Levine, M. & Papatsenko, D. 2006, "Computational models for neurogenic gene expression in the Drosophila embryo", Current Biology, vol. 16, no. 13, pp. 1358-1365.

CHAPTER 3

Analyzing Negative Feedback and Shuttling Using a Synthetic Gene Network Expressed in the *Drosophila melanogaster* Embryo

Ashley A. Jermusyk, Nicholas P. Murphy, and Gregory T. Reeves

*This chapter will be submitted to *Molecular Systems Biology*

3.1 - Introduction

Regulation of gene expression through genetic interactions, interconnected into complex networks, is crucial to the fitness of all organisms. These genetic regulatory networks are composed of several over-represented sets of interactions, called "motifs", which are individually amenable to study (Milo et al. 2002). Many such studies are currently being conducted using synthetic gene network motifs in single-cell systems (Elowitz, Leibler 2000b, Gardner, Cantor & Collins 2000, Friedland et al. 2009, Payne et al. 2013, Mitarai, Jensen & Semsey 2015, Schaerli et al. 2014, Basu et al. 2004). Such systems are highly advantageous from a practical point of view and often shed light on the dynamic behavior of network motifs. However, this research is unable to address the question of how these networks translate into inherently multi-cellular systems such as tissue patterning, stem-cell differentiation, cancer, and wound healing systems, each of which has a spatial component. This study seeks to address how a synthetic negative feedback network behaves in space in the developing *Drosophila melanogaster* embryo.

Negative feedback loops in biology can result in a rich diversity of phenomenological behavior (reviewed in Jermusyk, Reeves 2016). Under some conditions, negative feedback can destabilize the output of a system and create oscillations (Elowitz, Leibler 2000a, Hoffmann et al. 2002, Holley et al. 2002). Under other conditions, it may instead serve to stabilize the system against perturbations in the input signal. Negative feedback acts in this manner to control tumor suppression genes in mice as well as pluripotency and self-renewal in embryonic stem cells (Oren 1999, Pan et al. 2006). In a spatially-distributed system, the negative feedback that occurs when a morphogen activates its own inhibitor (the "self-enhanced ligand degradation" paradigm)

may add robustness to downstream gene expression patterns (Eldar et al. 2003, Reeves et al. 2005, Lander 2007, Lander et al. 2009). Negative feedback can also be used to limit the range or length scale of a signal. This is occurs in the JAK/STAT pathway in vertebrates (Rawlings, Rosler & Harrison 2004).

Here we create a spatially distributed synthetic gene network in the early *Drosophila* embryo. We use the *bcd* 3' UTR to express the yeast activator Gal4 in an anterior-posterior gradient (Macdonald, Struhl 1988, Weil, Forrest & Gavis 2006, Frigerio et al. 1986, Driever, Nusslein-Volhard 1988, Spirov et al. 2009, Huang, Rusch & Levine 1997, Janody et al. 2001). To create negative feedback, we engineered a *gal80* construct to contain three or five *UAS* sites (Upstream Activating Sequences), which are activated by Gal4 (Giniger, Ptashne 1988, Lohr, Venkov & Zlatanova 1995, Elliott, Brand 2008). Gal80 binds to and inactivates Gal4 (Janody et al. 2001, Carrozza et al. 2002).

We found that, depending on the amount of *gal4* and *gal80* present in the embryo, this negative feedback system can exhibit either an attenuation or a shuttling phenotype, in which Gal4/Gal80 binding and diffusion can extend the spatial range of Gal4 signaling. Both mathematical modeling, as well as expression of the Gal80-binder Gal3, validate our findings (Egriboz et al. 2013). This work demonstrates how in spatial systems, gene networks can produce very different outputs depending on the relative spatial domains of inputs.

3.2 - Results

3.2.1 - Gal4-driven lacZ expression has a graded border

A negative feedback network was created consisting of *gal4*, *gal80*, and *lacZ* (see Fig. 3.1D). We used a previously-published Gal4 construct (Gal4-GCN4:Bcd 3'UTR (Janody et al. 2001)) that mimics the Bicoid anterior-posterior concentration gradient. For baseline measurements, we first imaged embryos containing only Gal4 (four copies of this construct) and *UAS-lacZ* (no *UAS-gal80*).

In these embryos, the synthetic gradient in Gal4 activates the expression of the *UAS-lacZ* construct in a spatially-dependent fashion. Using fluorescent *in situ* hybridization, together with image analysis protocols (see Materials and Methods) we were able to quantify the expression domain of *lacZ* (Fig. 3.1E and Fig. A.1).

We found that the expression boundary of *lacZ*, resulting from the Gal4 gradient, is not sharp, in contrast to previous work using this Gal4 construct (Crauk, Dostatni 2005). This difference may be due to the differences between in situ hybridization procedures using alkaline-phosphatase staining versus fluorescent detection.

3.2.2 - Gal80 expression attenuates lacZ expression

Next, to measure the effect of the negative feedback loop, we analyzed embryos containing all three constructs: *gal4* (four copies of the *gal4-bcd 3'UTR), UAS-lacZ*, as well as *UAS-gal80* (one copy). We tested two different promoter strengths for *gal80*: three or five UAS sites were used. The expression profile for *lacZ* with and without *gal80* was analyzed to determine the effect of Gal80 mediated negative feedback on *lacZ*

production due to Gal4. We found that the expression pattern of *lacZ* is qualitatively unchanged (Fig. 3.1F). Furthermore, *gal80* expression is similar to that of *lacZ*.

To determine the extent to which *gal80* affects the *lacZ* profile, we first compared the normalized intensity of *lacZ* at each point along the AP axis when there is no *gal80* present (control) to when there is *gal80* present (in either the three or five *UAS* sites scenario) (Fig. 3.1G). There was no statistically significant difference between the normalized intensity of *lacZ* without *gal80* and with *gal80* linked to three UAS sites along the entire AP axis. However, there is a difference (p-value < 0.05) between the profiles for *lacZ* without *gal80* and with *gal80* linked to five UAS sites from 21 to 38% embryo length and from 41 to 44% embryo length; with the maximum difference at 29% embryo length.

To compare the curves using a single summary statistic, we evaluated the AP position, x_L, at which the *lacZ* profile fell to 27% maximal intensity. This level was chosen because it corresponds to the normalized intensity of the *lacZ* profile with no *gal80* at 29% embryo length. As we found previously, with only 3 UAS sites driving gal80 expression, no statistically significant effect on the *lacZ* profile was observed, as compared to the system without gal80 (Fig 3.1H).

From this analysis we are able to characterize the nature of the shift in lacZ when *UAS*x5:*gal80* is present in the system. In this case, attenuation is observed and *lacZ* expression was shifted toward the anterior pole (x = 0.262 ± 0.045 with UASx5:*gal80* vs. x = 0.291 ± 0.071 with no gal80, p = 0.038) (Fig. 3.1H). At the same time a decrease in the standard deviation was observed (F-test for variance, p = 0.002). These two observations are indicative of negative feedback and demonstrate the ability of a simple

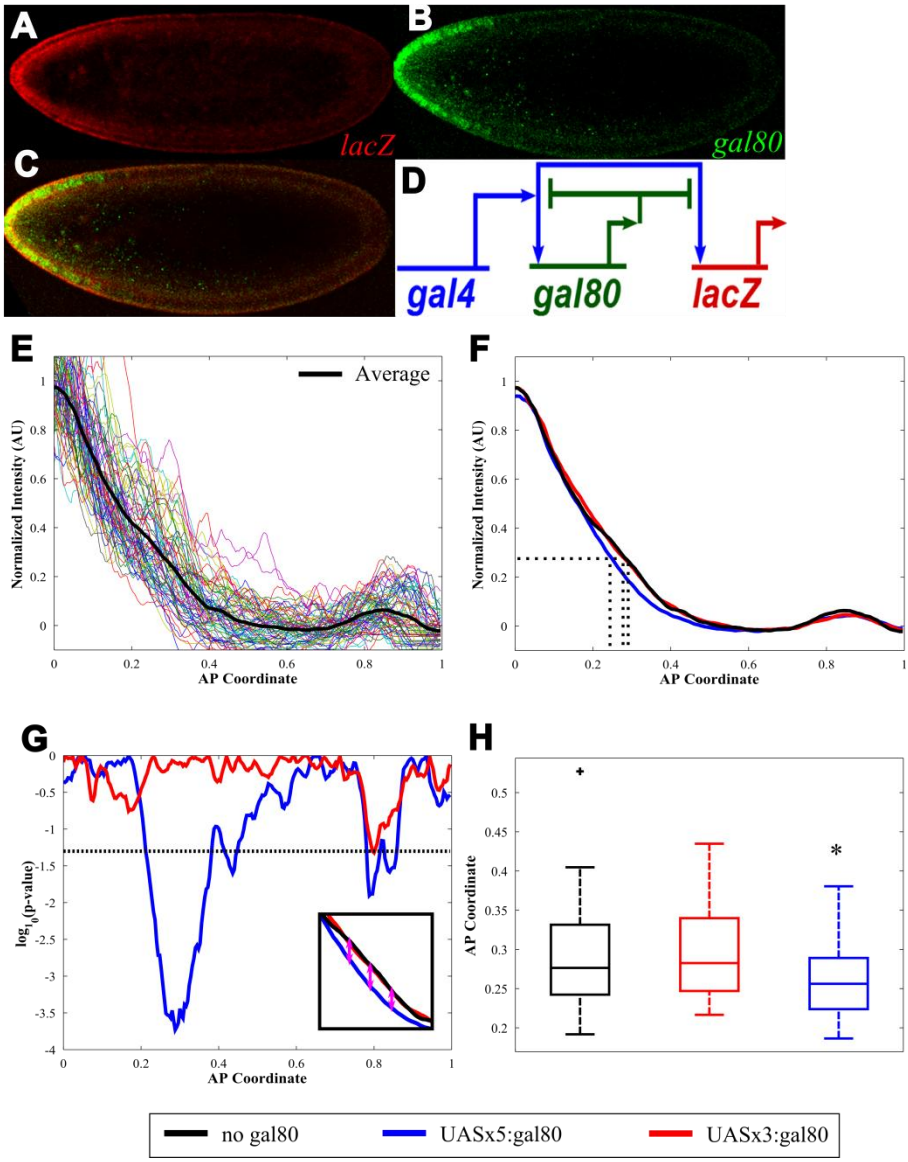

Figure 3.1: Effect of Gal80 on *lacZ* expression in attenuation situation. **(A)** *lacZ* mRNA expression at the mid-saggittal plane in an embryo expressing *UAS*x5:*gal80*, from mothers with four copies of Gal4GCN4. **(B)** *gal80* mRNA expression in the same embryo as (A). **(C)** Merged image of expression in (A) and (B). **(D)** Network diagram, Gal4 activated *gal80* and *lacZ* expression. Gal80 binds to Gal4, repressing *gal80* and *lacZ* activation. **(E)** Quantification of *lacZ* mRNA expression in embryos without *gal80*, each colored curve represents the dorsal or ventral side of a single embryo. The average for all embryos is in black. **(F)** Average curves for *lacZ* expression in embryos without *gal80* (n = 35), with *UAS*x3:*gal80* (n = 36), and *UAS*x5:*gal80* (n = 51). **(G)** Difference between the normalized intensity of *lacZ* without *gal80* versus with *UAS*x5:*gal80* or *UAS*x3:*gal80* at a given position along the anterior-posterior axis, dashed line denotes p = 0.05. **(H)** Box plots of AP coordinate where normalized intensity is 0.27 (see dashed lines in [F],maximum difference between no *gal80* control and *UAS*x5:*gal80* [G]). Asterisk denote statistical significance ($p < 0.05$).

negative feedback loop to reproducibly give rise to gene expression in a given spatial domain by buffering against minor biological and environmental fluctuations.

3.2.3 - Increasing abundance of Gal80 creates a shuttling system

In order to increase the strength of negative feedback through Gal80 we altered the amounts of *gal4* and *gal80* in this system. By increasing the amount of Gal80 relative to Gal4 we should see a greater effect of Gal80 and enhanced control due to negative feedback on the system. Therefore, we analyzed the lacZ profiles in embryos carrying two copies of *gal4-bcd 3'UTR* (half the amount of *gal4* as used previously) and either one or two copies of *UAS-gal80* (previously only one copy of *UAS-gal80* was used). As before, the *UAS-gal80* construct contained either three or five UAS sites. To quantify the differences between these curves, we determined the point along the AP axis at which each curve passes 31% maximal intensity (Fig. 3.2A). This corresponds to the normalized intensity of the *lacZ* profile in embryos without Gal80 at 26% embryo length, or the position along the AP axis with the maximum difference between embryos without Gal80 and with Gal80. At the highest Gal80 to Gal4 ratio, the *lacZ* profile shifted away from the anterior pole ($x = 0.321 \pm 0.073$ vs. $x = 0.265 \pm 0.046$ with no gal80, $p = 5 \times 10^{-4}$), contrary to expectation (see Fig. 3.2B).

One phenomenon that could be responsible for epansion of the lacZ profile is facilitated diffusion or "shuttling". This "shuttling" would occur if Gal80 binding to Gal4 increases the effective diffusion of Gal4. This can be accomplished by an increase in the effective diffusion of free Gal4 when Gal80 is present. If Gal4/Gal80 complex

preferrentially remains unbound to the DNA and allows for Gal4 to diffuse out of the nuclei and further away from the anterior pole, it would expand the signaling range of Gal4. The existence of this shuttling phenomenon was validated in a number of ways: using a model to demonstrate it is biophysically possible to switch between attenuation and shuttling in our system, adding a molecule to break-up this Gal4/Gal80 complex and create an effective sink for Gal4, and for this increase in signaling using this molecule to break-up the Gal4/Gal80 complex not to have an effect in the attenuation system.

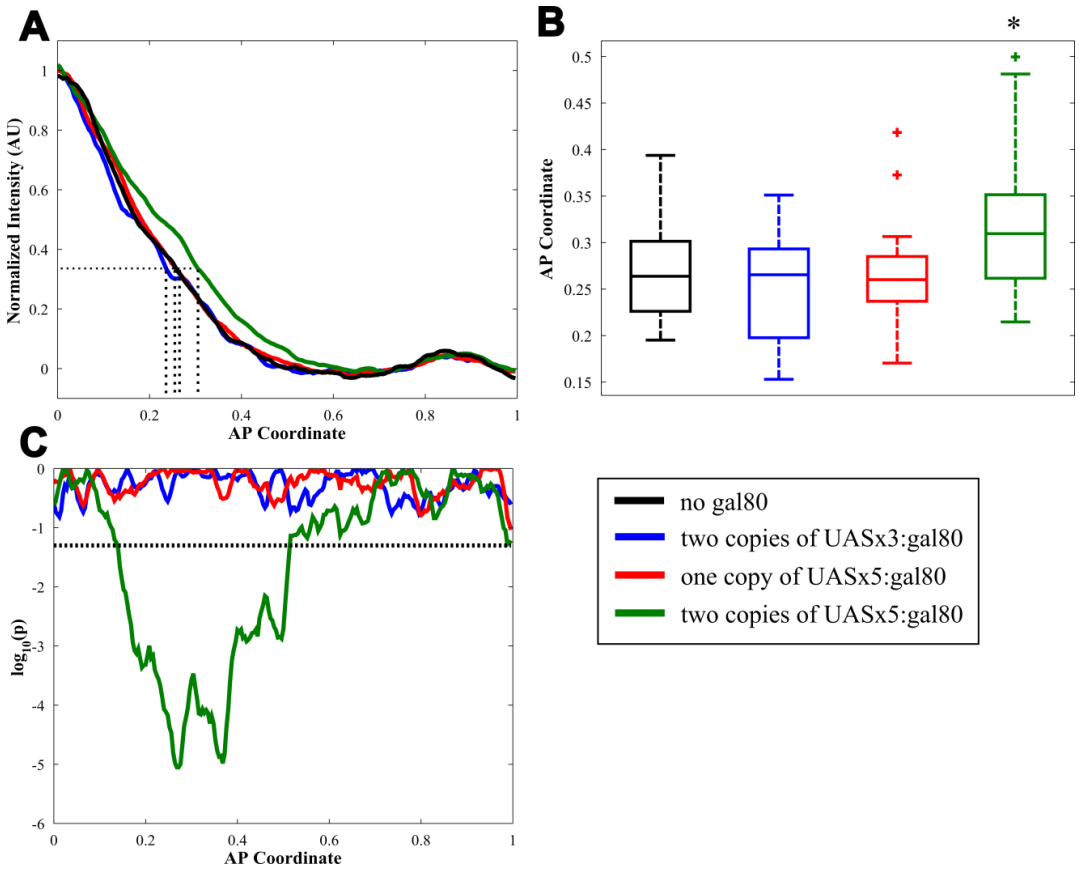

Figure 3.2: Gal80 is able to create a shuttling system. **(A)** Average curves of *lacZ* mRNA expression at the mid-saggittal plane in embryos with no *gal80* (n = 27), two copies of *UAS*x5:*gal80* (n = 36), two copies of *UAS*x3:*gal80* (n = 13), or one copy of *UAS*x5:*gal80* (n = 22), from mothers with two copies of Gal4GCN4. **(B)** Box plots of AP coordinate where normalized intensity is 0.31 (see dashed lines in [A], maximum difference between no *gal80* control and two copies of *UAS*x5:*gal80* [C]). Asterisk denotes statistical significance ($p < 0.005$). **(C)** Difference between normalized intensity of *lacZ* without *gal80* versus with varying amounts of *gal80*, dashed line denotes $p = 0.05$.

3.2.4 - A model of Gal4/Gal80 interactions predicts both attenuation and shuttling regimes

One question that arises is whether it would be biophysically possible to have a system in which Gal80 attenuates the signaling range of Gal4 at low Gal80:Gal4 ratios, but extends the signaling range of Gal4 at high Gal80:Gal4 ratios. To answer this question, we built a mechanistic model of Gal4/Gal80 interactions, as shown below.

$$0 = \lambda_g^2 g_{xx} - g - \mu(gr - \nu c)$$

$$0 = \lambda_r^2 r_{xx} - r - \beta\mu(gr - \nu c) + q_r f_r(g)$$

$$0 = \lambda_c^2 c_{xx} - \rho_c c + \mu(gr - \nu c)$$

In these equations, g represents the concentration of Gal4, r the concentration of the repressor Gal80, and c the Gal4/Gal80 complex. At steady state, each component diffuses, is degraded, and participates in a reversible binding reaction with forward rate μ and affinity ν. Gal80 is produced by nuclei in which Gal4 signaling, represented by the function $f_r(g)$ (see Methods), is sufficiently high. We assume no-flux boundary conditions for all species at both $x = 0$ (anterior pole) and $x = 1$ (posterior pole), except for a constant flux production of g at $x = 0$ (see Appendix A for details of model development).

The constant flux production of Gal4 at the anterior pole is denoted by the parameter q_g, which equals 1 for four copies of Gal4, and q_{2x} for two copies. Because the copies of *gal4* are at two different sites within the genome, we could not be sure that two copies of *gal4* resulted in one half the production of Gal4 protein as compared to four copies of *gal4*. Therefore, we investigated the behavior of the model for q_{2x} between 0.35 and 0.65. We found that for lower values of q_{2x}, the model was able to adequately

reproduce an attenuated *lacZ* profile in the Gal4x4/Gal80x1 scenario, and an expanded *lacZ* profile in the Gal4x2/Gal80x2 scenario (Fig. 3.3A,B). This supports the plausibility of the hypothesis that the system performs attenuation for one dosage ratio and shuttling for another.

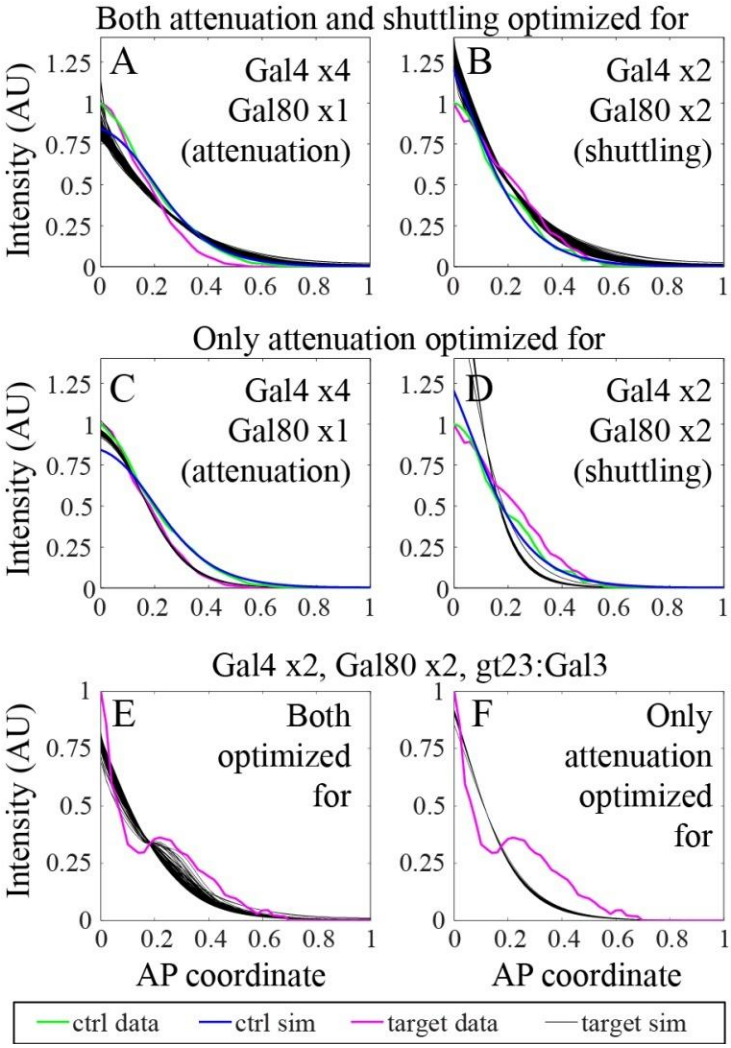

Figure 3.3: A mechanistic model of Gal4/Gal80/Gal3 interactions supports our hypothesis. (A,B) The model, when simultaneously fit to both the attenuation (A) and shuttling (B) data, is able to adequately satisfy both scenarios. The same parameter sets were used in both (A) and (B), with the only difference being that the levels of Gal4 and Gal80 are altered. (C,D) When the model is fit only to the attenuation phenotype, the attenuation fit is better (C), but shuttling does not occur (D). (E) When Gal3 is added to the system, the model exhibits a similar phenotype as experimentally observed. The parameter sets were the same as shown in (A) and (B). (F) However, with parameter sets that resulted from an attenuation-only optimization, as seen in (C) and (D), the presence of Gal3 does not result in a local increase in *lacZ* expression.

3.2.5 - Expression of Gal3 in a stripe results in a peak of lacZ expression

One prediction of the shuttling hypothesis is that forcing the local degradation or capture of the inhibitor (Gal80) should result in a similarly localized peak in signaling activity. Therefore, we introduced the yeast protein Gal3 into the Gal4/Gal80 system. In the presence of Gal3, Gal80 dissociates from Gal4 (Egriboz et al. 2013, Platt, Reece 1998, Smidtas, Schachter & Kepes 2006). This results in freeing of the Gal4 activation domain from Gal4 and activation of *UAS*-linked genes (Egriboz et al. 2013, Platt, Reece 1998, Smidtas, Schachter & Kepes 2006).

Two different enhancer regions were used to create stripes of Gal3 (Fig. 3.4), namely *gt23* (Ochoa-Espinosa et al. 2005) and *evestr2* (Small, Blair & Levine 1992). As predicted by the shuttling hypothesis, the expression of Gal3 in these domains causes a local increase in *lacZ* production (Fig. 3.4).

To control for the possibility that Gal3 expression is causing UAS-*lacZ* transcription without Gal80, we examined embryos that carried two copies of Gal4 and Gal3, but lacked Gal80. Surprisingly, a small increase in *lacZ* expression near the site of Gal3 expression was also observed when no Gal80 was present in the system (Fig. 3.4B-C). While the yeast literature does not suggest that Gal3 interacts with Gal4, our data suggests that Gal3 may have some previously unknown interaction with Gal4. However, Gal3 appears to have a more significant localization effect when Gal80 is in the system (Fig. 3.4B-C), supporting our hypothesis that Gal80 shuttles Gal4.

To confirm that increased *lacZ* expression in the Gal3 domains is explained only by the shuttling phenomenon, we extended our model to include the presence of Gal3 (see Appendix A). When the model is optimized to simultaneously fit both the

attenuation and shuttling regimes (Fig. 3.3A,B), local Gal3 expression can result in a corresponding local increase in *lacZ* expression (Fig. 3.3E). On the other hand, when the model is only optimized to fit the attenuation regime (and thus, shuttling does not occur when the Gal80/Gal4 ratio is high; Fig. 3.3C,D), the presence of Gal3 does not alter the *lacZ* expression profile (Fig. 3.3F). Thus, our Gal3 experimental results unambiguously support the presence of shuttling within our system, as the model without shuttling cannot replicate the experimental results.

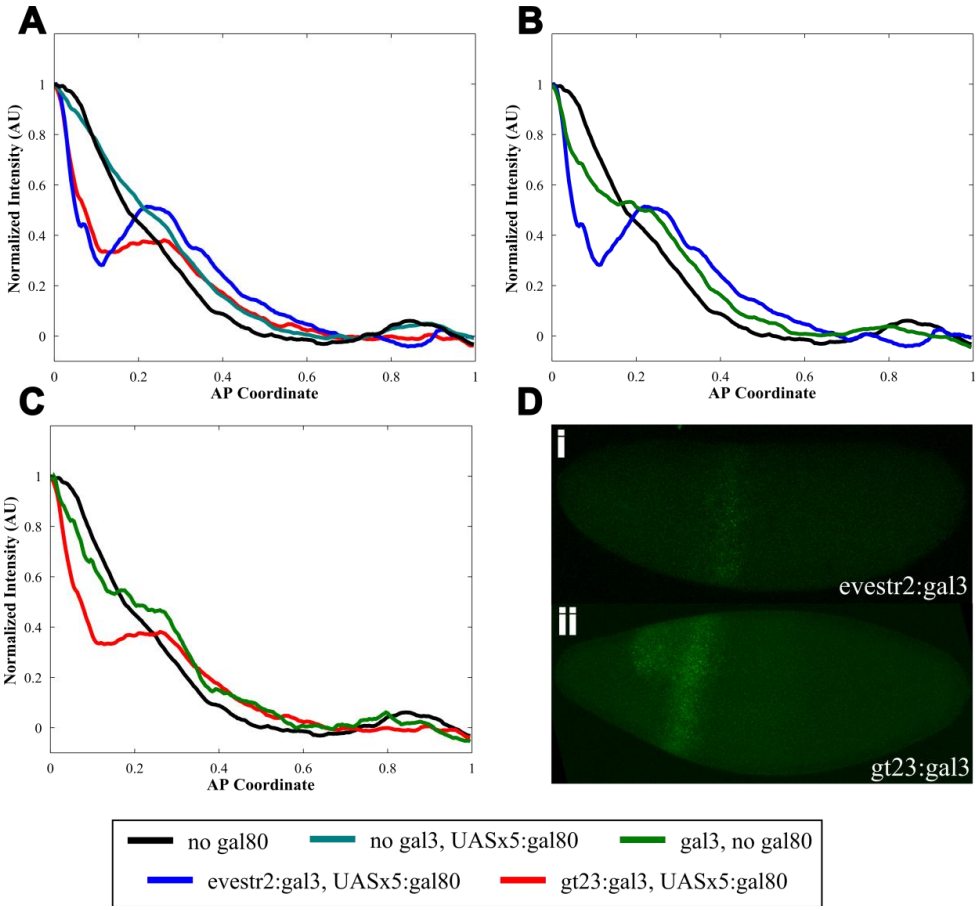

Figure 3.4: Localized Gal3 creates a peak in *lacZ*. (A) Average curves of *lacZ* expression in embryos without *gal80* and without *gal3* (n = 27) and with two copies of *UAS*x5:*gal80* and: no *gal3* (n = 36), *evestr2*:*gal3* (n = 19), or *gt23*:*gal3* (n = 12). (B) Average curves of *lacZ* expression in embryos without *gal80* and without *gal3*, without *gal80* and with *evestr2:gal3* (n = 9), and with both *gal80* and *evestr2:gal3*. (C) Average curves of *lacZ* expression in embryos without *gal80* and no *gal3*, without *gal80* and with *gt23:gal3* (n = 6), and with both *gal80* and *gt23:gal3*. (D) Expression of *gal3* in embryos expressing (i) *evestr2*:*gal3* and (ii) *gt23*:*gal3*.

3.3 - Discussion

A synthetic negative feedback network consisting of *gal4*, *gal80*, and *lacZ* was expressed in the *Drosophila* embryo. This experimental system was able to produce weak negative feedback, marked by attenuation, the shifting of expression of *lacZ* toward the source of activation (anterior pole) and an increase in robustness. When the copy numbers of *gal4* and *gal80* were altered, a shuttling system was created at a high Gal80 to Gal4 ratio. In this shuttling system, the *lacZ* profile expands toward the posterior pole, away from the source of activation. Shuttling has been found in other systems to produce robust gradients from an initial broad morphogen signal (Eldar et al. 2002, Haskel-Ittah et al. 2012). Shuttling requires a broad morphogen gradient which activates expression of the shuttling molecule. In this case the broad morphogen gradient of Gal4 activates expression of the *UAS*-linked *gal80*. This shuttling molecule is able to form a complex with the morphogen. A third molecule can be used which breaks up the shuttling molecule-morphogen complex, releasing the active morphogen. We were able to introduce Gal3, which results in an increase in *lacZ* expression at the source of Gal3. This is caused by the freeing of Gal4 from the Gal4-Gal80 complex. This provides further evidence for our shuttling system.

Previous studies have found evidence for morphogen gradients which enhance their own degradation, this form of negative feedback is known as self-enhanced ligand degradation. In this system morphogens degrade at a higher level where their levels are highest and at a slower rate farther from its source increasing the signaling range for the morphogen. This has been found in wingless and hedgehog patterning in the *Drosophila* wing, BMP signaling and DV axis specification in the zebrafish and *Xenopus* embryos,

Wnt and EGFR signaling systems in mammalian cells, and retinoic acid signaling in zebrafish (Eldar et al. 2003, Yabe et al. 2003, Reversade, De Robertis 2005, Niida et al. 2004, Katoh et al. 1993, White et al. 2007).

Parallels from these systems can be drawn to our synthetic system. From our synthetic system we can understand in isolation the mechanisms at work in these systems better. Namely, how the simple negative feedback system is able to achieve both an attenuation and a shuttling system depending on which system is being observed.

While shutting has only recently been proposed to explain the ability of certain morphogen gradients to be defined and achieve robust patterns, comparisons to other systems suggest that shuttling may exist in other negative feedback systems (Haskel-Ittah et al. 2012). Most importantly this work demonstrated a negative feedback system that is able to produce two very different outputs depending on the spatial domains of expression and relative amounts of these genes. This shows the complexity of gene networks in tissue patterning and other multi-cellular systems. While much previous work has been carried out to understand synthetic gene networks in single-celled systems, much care must be taken to extrapolate these findings into multi-cellular systems.

3.4 - Material and Methods

3.4.1 - Plasmids

All plasmids were constructed from the pUAST parent plasmid (gift from J. Mahaffey). The *UAS*x5:*gal80* was created by inserting gal80 (PCR amplified from genomic DNA of flies containing gal80) into the pUAST digested with *Not*I and *Xba*I. The *UAS*x3 construct was obtained by inserting the *UAS* sequence

(TGCGGAGTACTGTCCTCCGAG) into pBlueScript II SK (+) (from Addgene) flanked by *Sal*I and *Xho*I restriction sites. Subseqent multimerization was performed using restriction digest with *Sal*I/*Not*I and *Xho*I/*Not*I and subsequent ligation. The final *UAS*x3:*gal80* plasmid was created by insertion of *UAS*x3 (PCR amplified from *UAS*x3 in pBlueScript), *hsp70* (PCR amplified from pUAST), and *gal80* into pUAST parent plasmid. The *gal3* constructs were made by inserting the *eve* minimal promoter (PCR amplifed from genomic DNA), *gal3* (PCR amplified from yeast genomic DNA), and either *evestr2* (PCR amplified from genomic *Drosophila* DNA using primers AGATACATaagcttGCCATCAGCGAGATTATTAGTCAA and AGACTCAGctgcagAGGGCTAAGTCGGCGCAAA) or *gt23* (PCR amplified using primers AGATCATaagcttGGGAATTCGGCGACTTGGATCGTGAG and ATGACACActgcagAAAACTGCAGCTGCCCTGCCCTGCTCTG from genomic *Drosophila* DNA) enhancer regions into the *UAS*x3:*gal80* plasmid.

3.4.2 - Fly Stocks

The w-;Gal4-GCN4:Bcd 3'UTR;Gal4-GCN4:Bcd 3'UTR flies were a gift from Dr. Dostatni (Janody et al. 2001). The *UASp*:*lacZ* flies used in this study were obtained from Bloomington Stock Center (BS 3955). The *UAS*x3:*gal80* and *UAS*x5:*gal80* fly lines were created by injection and incorporation of plasmid constructs into the 68A4 landing site (injections performed by Model System Injections into yw;attP2 flies). The *gal3* constructs were incorporated into the 65B2 landing site by fly injection (injections performed by Model System Injections into yw;VK33 flies).

3.4.3 - Embryo Staining and Image Collection and Analysis

Embryos were fixed 2-4 hrs after egg laying per standard protocols. Fluorescent *in situ* hybridization was conducted per published protocols (Kosman et al. 2004) (Proteinase K treatment was omitted) using RNA probes for *lacZ* (biotin conjugated), *gal80* (flourescein conjugated), and *gal3* (digoxigenin conjugated). Primary antibodies to biotin (goat anti-biotin, 1:5000; gift from Immunoreagents), flourescein (rabbit anti-fluorescein, 1:500; ThermoFisher Scientific), and digoxigenin (mouse anti-digoxigenin, 1:500; Roche). Secondary antibodies used were Alexa Flour 488 donkey anti-rabbit (ThermoFisher Scientific), Alexa Flour 546 donkey anti-goat (ThermoFisher Scientific), and Alexa Flour 488 donkey anti-mouse (ThermoFisher Scientific).

Images were taken at the mid-saggittal plane (determined optically) using a Zeiss Confocal microscope.

The images were analyzed using a modified version of the method described in Trisnadi et al. The code was altered to fit the mid-saggittal section of the embryo to an ellipse using elipse_fit.m (written by Tal Henel, available from Matlab Central) using the two foci of the ellipse, the embryo was broken up into two half circles (at the anterior and posterior poles) and a rectangle bridging the two circles. The flourescence intensity around the periphery of the embryo was determined in each section (code available).

References

Basu, S., Mehreja, R., Thiberge, S., Chen, M. & Weiss, R. 2004, "Spatiotemporal control of gene expression with pulse-generating networks", Proceedings of the National Academy of Sciences of the United States of America, vol. 101, no. 17, pp. 6355-6360.

Carrozza, M., John, S., Sil, A., Hopper, J. & Workman, J. 2002, "Gal80 confers specificity on HAT complex interactions with activators", *Journal of Biological Chemistry*, vol. 277, no. 27, pp. 24648-24652.

Crauk, O. & Dostatni, N. 2005, "Bicoid determines sharp and precise target gene expression on the Drosophila embryo", Current Biology, vol. 15, no. 21, pp. 1888-1898.

Driever, W. & Nusslein-Volhard, C. 1988, "The Bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner.", Cell, vol. 54, no. 1, pp. 95-104.

Egriboz, O., Goswami, S., Tao, X., Dotts, K., Schaeffer, C., Pilauri, V. & Hopper, J.E. 2013, "Self-Association of the Gal4 Inhibitor Protein Gal80 Is Impaired by Gal3: Evidence for a New Mechanism in the GAL Gene Switch", Molecular and cellular biology, vol. 33, no. 18, pp. 3667-3674.

Eldar, A., Rosin, D., Shilo, B. & Barkai, N. 2003a, "Self-enhanced ligand degradation underlies robustness of morphogen gradients", Developmental Cell, vol. 5, no. 4, pp. 635-646.

Eldar, A., Rosin, D., Shilo, B. & Barkai, N. 2003b, "Self-Enhanced ligand degradation underlies robustness of morphogen gradients.", Dev Cell, vol. 5, pp. 635-646.

Elliott, D.A. & Brand, A.H. 2008, "The GAL4 system : a versatile system for the expression of genes.", Methods Mol Biol, vol. 420, pp. 79-95.

Elowitz, M.B. & Leibler, S. 2000a, "A synthetic oscillatory network of transcriptional regulators", Nature, vol. 403, no. 6767, pp. 335-338.

Elowitz, M.B. & Leibler, S. 2000b, "A synthetic oscillatory network of transcriptional regulators.", Nature, vol. 403, no. 6767, pp. 335-338.

Friedland, A.E., Lu, T.K., Wang, X., Shi, D., Church, G. & Collins, J.J. 2009, "Synthetic Gene Networks That Count", Science, vol. 324, no. 5931, pp. 1199-1202.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. 1986, "Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network.", Cell, vol. 47, no. 5, pp. 735-746.

Gardner, T., Cantor, C. & Collins, J. 2000, "Construction of a genetic toggle switch in Escherichia coli", Nature, vol. 403, no. 6767, pp. 339-342.

Giniger, E. & Ptashne, M. 1988, "Cooperative Dna-Binding of the Yeast Transcriptional Activator Gal4", Proceedings of the National Academy of Sciences of the United States of America, vol. 85, no. 2, pp. 382-386.

Haskel-Ittah, M., Ben-Zvi, D., Branski-Arieli, M., Schejter, E.D., Shilo, B. & Barkai, N. 2012, "Self-Organized Shuttling: Generating Sharp Dorsoventral Polarity in the Early Drosophila Embryo", Cell, vol. 150, no. 5, pp. 1016-1028.

Hoffmann, A., Levchenko, A., Scott, M. & Baltimore, D. 2002, "The I kappa B-NF-kappa B signaling module: Temporal control and selective gene activation", Science, vol. 298, no. 5596, pp. 1241-1245.

Holley, S., Julich, D., Rauch, G., Geisler, R. & Nusslein-Volhard, C. 2002, "Her1 and the Notch Pathway Function within the Oscillator Mechanism that Regulates Zebrafish Somitogenesis", Development, vol. 129, no. 5, pp. 1175-1183.

Huang, A.M., Rusch, J. & Levine, M. 1997, "An anteroposterior dorsal gradient in the Drosophila embryo", Genes & development, vol. 11, no. 15, pp. 1963-1973.

Janody, F., Sturny, R., Schaeffer, V., Azou, Y. & Dostatni, N. 2001, "Two distinct domains of Bicoid mediate its transcriptional downregulation by the Torso pathway", Development, vol. 128, no. 12, pp. 2281-2290.

Jermusyk, A.A. & Reeves, G.T. 2016, "Transcription Factor Networks", Encyclopedia of Cell Biology, vol. 4, pp. 63-63-71.

Katoh, M., Yazaki, Y., Sugimura, T. & Terada, M. 1993, "C-Erbb3 Gene Encodes Secreted as Well as Transmembrane Receptor Tyrosine Kinase", Biochemical and biophysical research communications, vol. 192, no. 3, pp. 1189-1197.

Kosman, D., Mizutani, C., Lemons, D., Cox, W., McGinnis, W. & Bier, E. 2004, "Multiplex detection of RNA expression in Drosophila embryos", Science, vol. 305, no. 5685, pp. 846-846.

Lander, A.D. 2007, "Morpheus unbound: Reimagining the morphogen gradient", Cell, vol. 128, no. 2, pp. 245-256.

Lander, A.D., Lo, W., Nie, Q. & Wan, F.Y.M. 2009, "The Measure of Success: Constraints, Objectives, and Tradeoffs in Morphogen-mediated Patterning.", Cold Spring Harbor Perspect Biol, vol. 1, no. 1, pp. a002022.

Lohr, D., Venkov, P. & Zlatanova, J. 1995, "Transcriptional regulation in the yeast GAL gene family: a complex genetic network.", FASEB J, vol. 9, no. 9, pp. 777-787.

Macdonald, P.M. & Struhl, G. 1988, "cis-acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos.", Nature, vol. 336, no. 6199, pp. 595-598.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. 2002, "Network motifs: Simple building blocks of complex networks", Science, vol. 298, no. 5594, pp. 824-827.

Mitarai, N., Jensen, M.H. & Semsey, S. 2015, "Coupled Positive and Negative Feedbacks Produce Diverse Gene Expression Patterns in Colonies", Mbio, vol. 6, no. 2, pp. e00059-15.

Niida, A., Hiroko, T., Kasai, M., Furukawa, Y., Nakamura, Y., Suzuki, Y., Sugano, S. & Akiyama, T. 2004, "DKK1, a negative regulator of Wnt signaling, is a target of the beta-catenin/TCF pathway", Oncogene, vol. 23, no. 52, pp. 8520-8526.

Ochoa-Espinosa, A., Yucel, G., Kaplan, L., Pare, A., Pura, N., Oberstein, A., Papatsenko, D. & Small, S. 2005, "The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila", Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 14, pp. 4960-4965.

Oren, M. 1999, "Regulation of the p53 tumor suppressor protein", Journal of Biological Chemistry, vol. 274, no. 51, pp. 36031-36034.

Pan, G., Li, J., Zhou, Y., Zheng, H. & Pei, D. 2006, "A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal", Faseb Journal, vol. 20, no. 10, pp. 1730-+.

Payne, S., Li, B., Cao, Y., Schaeffer, D., Ryser, M.D. & You, L. 2013, "Temporal control of self-organized pattern formation without morphogen gradients in bacteria", Molecular Systems Biology, vol. 9, pp. UNSP 697.

Platt, A. & Reece, R. 1998, "The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex", Embo Journal, vol. 17, no. 14, pp. 4086-4091.

Rawlings, J., Rosler, K. & Harrison, D. 2004, "The JAK/STAT signaling pathway", Journal of cell science, vol. 117, no. 8, pp. 1281-1283.

Reeves, G.T., Kalifa, R., Klein, D.E., Lemmon, M.A. & Shvartsman, S.Y. 2005, "Computational analysis of EGFR inhibition by Argos.", Dev Biol, vol. 284, pp. 523-535.

Reversade, B. & De Robertis, E. 2005, "Regulation of ADMP and BMP2/4/7 at opposite embryonic poles generates a self-regulating morphogenetic field", Cell, vol. 123, no. 6, pp. 1147-1160.

Schaerli, Y., Munteanu, A., Gili, M., Cotterell, J., Sharpe, J. & Isalan, M. 2014, "A unified design space of synthetic stripe-forming networks", Nature Communications, vol. 5, pp. 4905.

Small, S., Blair, A. & Levine, M. 1992, "Regulation of Even-Skipped Stripe-2 in the Drosophila Embryo", Embo Journal, vol. 11, no. 11, pp. 4047-4057.

Smidtas, S., Schachter, V. & Kepes, F. 2006, "The adaptive filter of the yeast galactose pathway", Journal of theoretical biology, vol. 242, no. 2, pp. 372-381.

Spirov, A., Fahmy, K., Schneider, M., Frei, E., Noll, M. & Baumgartner, S. 2009, "Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein gradient.", Development, vol. 136, no. 4, pp. 605-614.

Weil, T.T., Forrest, K.M. & Gavis, E.R. 2006, "Localization of bicoid mRNA in late oocytes is maintained by continual active transport.", Dev Cell, vol. 11, no. 2, pp. 251-262.

White, R.J., Nie, Q., Lander, A.D. & Schilling, T.F. 2007, "Complex regulation of cyp26a1 creates a robust retinoic acid gradient in the zebrafish embryo", Plos Biology, vol. 5, no. 11, pp. 2522-2533.

Yabe, T., Shimizu, T., Muraoka, O., Bae, Y., Hirata, T., Nojima, H., Kawakami, A., Hirano, T. & Hibi, M. 2003, "Ogon/Secreted Frizzled functions as a negative feedback regulator of Bmp signaling", Development, vol. 130, no. 12, pp. 2705-2716.

CHAPTER 4

Determination of Novel Players in the *Drosophila melanogaster* Anterior-Posterior Patterning System Using Natural Variation

Ashley A. Jermusyk, Sarah E. Gharavi, Aslesha S. Tingare, and Gregory T. Reeves

*This chapter will be submitted to *PNAS*

4.1 - Background

Bicoid (Bcd) is expressed in a gradient along the anterior-posterior axis of the early *Drosophila melanogaster* embryo (Lucas et al. 2013, Morrison et al. 2012, Tamari, Barkai 2012). This protein gradient is formed by localized production at the anterior pole from *bicoid* (*bcd*) RNA deposited by the mother (Lucas et al. 2013, Morrison et al. 2012, Tamari, Barkai 2012). This Bcd morphogen gradient activates zygotic expression of the gap genes, *Krüppel* (*Kr*), *knirps* (*kni*), *hunchback* (*hb*), and *giant* (*gt*), in broad stripes along the anterior-posterior axis (Bieler, Pozzorini & Naef 2011, Perkins et al. 2006, Jaeger 2011) Cross-repression between the gap genes serve to refine their borders (Perkins et al. 2006, Jaeger 2011, Sokolowski, Erdmann & ten Wolde Pieter Rein 2012). The gap genes activate the downstream pair-rule genes which form the parasegments of the embryo (Small, Blair & Levine 1992). The pair rule genes control the expression of the segment polarity genes which form the segments of the embryo (McGinnis, Krumlauf 1992, von Dassow et al. 2000).

While there is much known about the connections within this network there are still unidentified components. Previous work has used a collection of wild-caught in-bred lines to identify novel genes responsible for phenotypic changes in *Drosophila* (Mackay et al. 2012, Huang et al. 2014). A few of these lines have been used to quantify variation in gene expression in AP patterning genes (Jiang et al. 2015). We sought to find new connections within this network by quantifying subtle natural variation among these lines in gene expression of *Kr* and Even-skipped (Eve). Variations in gene expression were then linked to genomic differences between the lines. This study found how small genomic changes, even single nucleotide changes, outside of previously

characterized enhancer regions can measurably impact gene expression patterns. By studying the effects of these small genomic changes we were also able to determine novel transcription factors (*medea*, *ultraspiracle*, *glial cells missing*, *pangolin*, and *orthopedia*) that act in the anterior-posterior patterning system. These genes were mainly found to impact the expression of *Kr* and Eve in subtle ways. However, one gene, *pangolin*, was found to have significant effects on the system. This evidence points to a large number of unexplored genes that act within the early embryo to control anterior-posterior patterning.

4.2 - Results and Discussion

4.2.1 - Differences exist in Kr and Eve expression among wild-caught lines

The expression patterns of *Kr* mRNA and Eve protein were determined in 13 of the DGRP fly lines and in a laboratory control strain (*yw*; see Materials and Methods). *Kr* is expressed in a broad stripe 43 – 53% embryo length (Fig. 4.1A,C). Eve is expressed in seven narrow stripes (at 32, 40, 47, 54, 61, 68, and 77% embryo length) (Fig. 4.1B,D). We measured the expression patterns for both *Kr* and Eve (see Materials and Methods) in the mid-saggittal plane of the embryo. Variability among the lines was found in the positioning of the two *Kr* borders and Eve stripes 1 – 6 (ANOVA, $p < 0.012$) (Table 4.1). Due to the comparatively high variability within each line in positioning of Eve stripe 7, no statistically significant difference was found among the lines for that stripe (ANOVA, $p = 0.08$). Once we established variability between the lines, we determined the associations responsible for this variation by conducting a pair-wise comparison for the positions of the anterior and posterior border of Kr and of the peaks of the seven Eve stripes among the DGRP fly lines and the *yw* control. The results of this

pair-wise analysis for *Kr* anterior border are seen in Fig. 4.1E (for analysis of all positions see Fig. B.1).

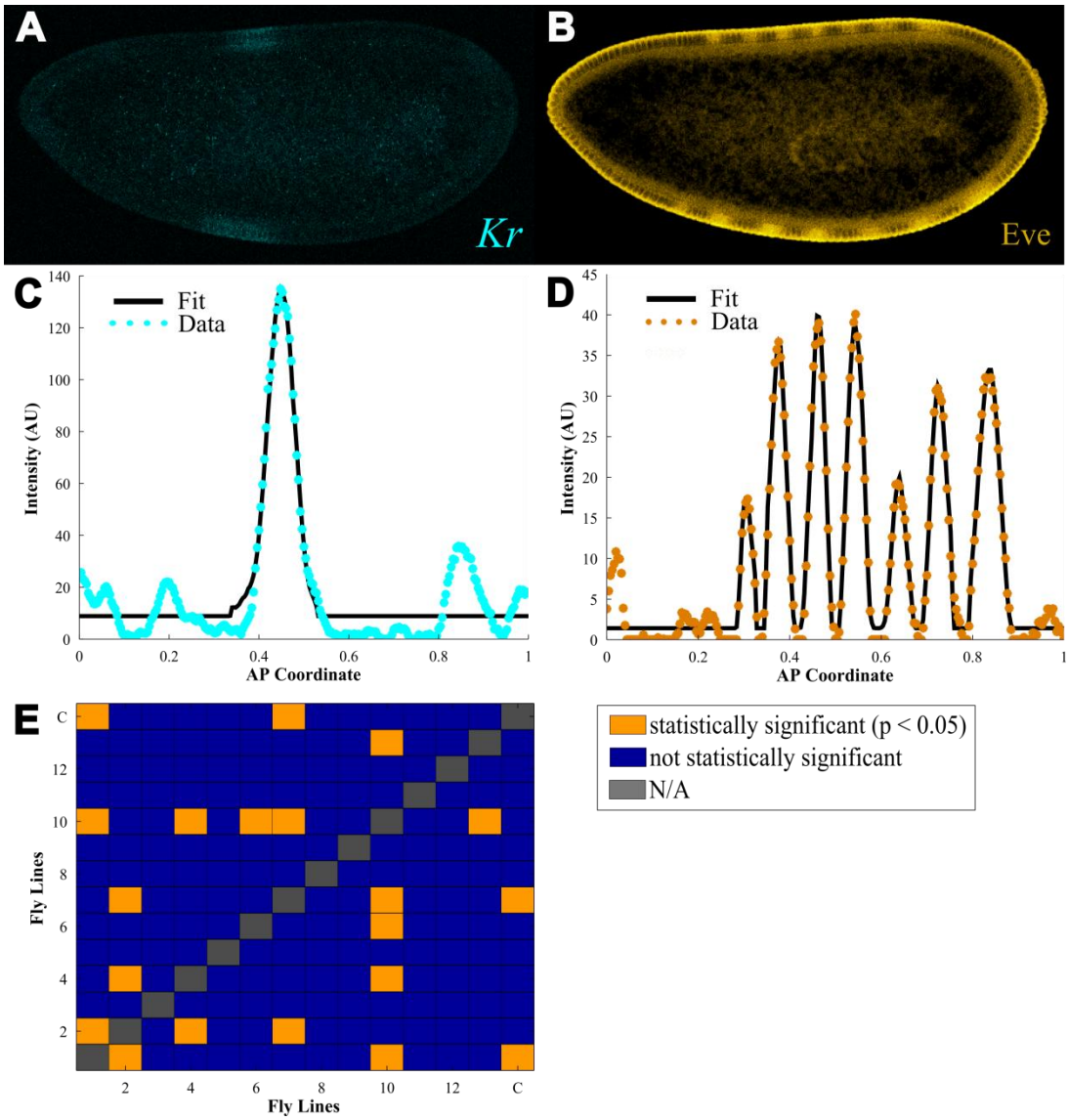

Figure 4.1: *Kr* and Eve expression quantified in the embryo reveals changes in gene expression among the fly lines. Normal expression of (A) *Kr* and (B) Eve as measured via *in situ* hybridization at the mid-saggittal section in an embryo (for these images and all images, anterior is to the left). Quantification of this expression along the dorsal half of this embryo where (*) is the normalized expression at each point and the solid black curve is the fit for (C) *Kr* and (B) Eve. (D) Heat map showing comparison of position of *Kr* anterior border in each of the fly lines, where orange denotes statistically significant (per post hoc Tukey-Kramer test, $p < 0.05$) differences between the lines and blue denotes no statistically significant difference between the lines. The fly lines are in the order of: RAL150, RAL306, RAL307, RAL315, RAL317, RAL360, RAL41, RAL57, RAL705, RAL761, RAL765, RAL799, RAL801, and laboratory control.

Table 4.1: ANOVA analysis of positioning of *Kr* and Eve across fly lines

Border or Stripe	P-value (no control)	P-value (w/control)
Kr (Anterior)	**2.36E-05**	**1.50E-06**
Kr (Posterior)	**4.28E-06**	**6.77E-07**
Eve Stripe 1	**1.56E-06**	**6.29E-06**
Eve Stripe 2	**1.98E-11**	**2.74E-11**
Eve Stripe 3	**4.69E-11**	**1.09E-12**
Eve Stripe 4	**1.54E-06**	**3.48E-12**
Eve Stripe 5	**0.012**	**7.91E-13**
Eve Stripe 6	**0.001**	**2.86E-05**
Eve Stripe 7	0.078	**0.039**

4.2.2 - Association Mapping to Locate Significant SNPs

Association mapping was used to determine if these differences in gene expression were correlated to specific genomic differences among the lines. All single nucleotide polymorphisms (SNPs) and short insertions and deletions (indels) 20 kb upstream and downstream of *Kr* and *eve* were evaluated, this region includes 661 mutations near *Kr* and 646 mutations near *eve*. For each SNP or indel within this region, a t-test was performed for each *Kr* border or Eve stripe with the lines grouped based on presence or absence of the mutation (Fig 4.2A-B). A SNP or indel is taken as significant where there is a difference ($p < 0.05$) between the position in the mutant lines compared to non-mutant lines (see Fig. B.2). We found 5 statistically significant mutations near the *Kr* locus, and 47 near *eve* (of those near the *eve* locus 13 were in known enhancers or within *eve*). For this many mutations tested, with this high p-value cutoff, we expect some false positives. To screen for false positives, the SNPs and indels with p-values less than 0.05 were tested for their ability to produce expression *in vivo*. This was accomplished by grouping one or more significant SNPs into putative enhancer regions (Fig. 4.2). In this study, four putative enhancers for *Kr* and twelve for *eve* were tested.

The genomic positions of these regions and the primers used to create constructs with these regions are listed in Table B.1.

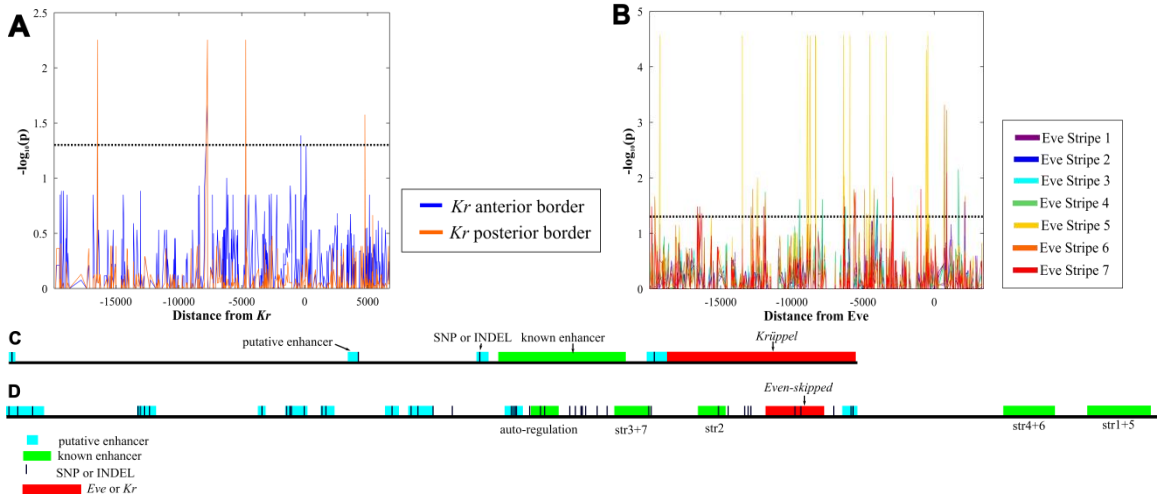

Figure 4.2: Results of association mapping analysis. Probability a given SNP or indel is correlated with changes in gene positioning for (A) *Kr* and (B) Eve. (C) Region surrounding the *Kr* gene with the SNPs and indels (dark blue bars) found to be associated with changes in *Kr* expression and the putative enhancers they were tested in. Known enhancers are shown in green and putative enhancers tested are shown in cyan. (D) Locations of SNPs and indels found to be associated with changes in *eve* expression. Putative enhancer regions tested (cyan) and known enhancer regions (green, stripe regulated shown below, (Fujioka et al. 1999, Small, Blair & Levine 1996, Small, Blair & Levine 1992)) are shown. Both (C) and (D) are drawn to scale.

4.2.3 - Testing of Putative Enhancers

Of the tested putative enhancers, three for *eve* and one for *Kr* were able to drive distinct expression *in vivo* (Fig. 4.3), representative embryo images are shown for all enhancers in Fig. B.4. The minimal promoter used with these putative enhancers drives expression of a stripe at roughly 20-30% embryo length (Fig. 4.3I); only putative enhancers that generate expression outside of this region were explored further (Liberman, Stathopoulos 2009). These expression patterns are relatively weak (Fig 4.3-A, C, E, and G), which is to be expected since they most likely contain only one binding site for a transcription factor. In addition, our methodology focuses on discovering

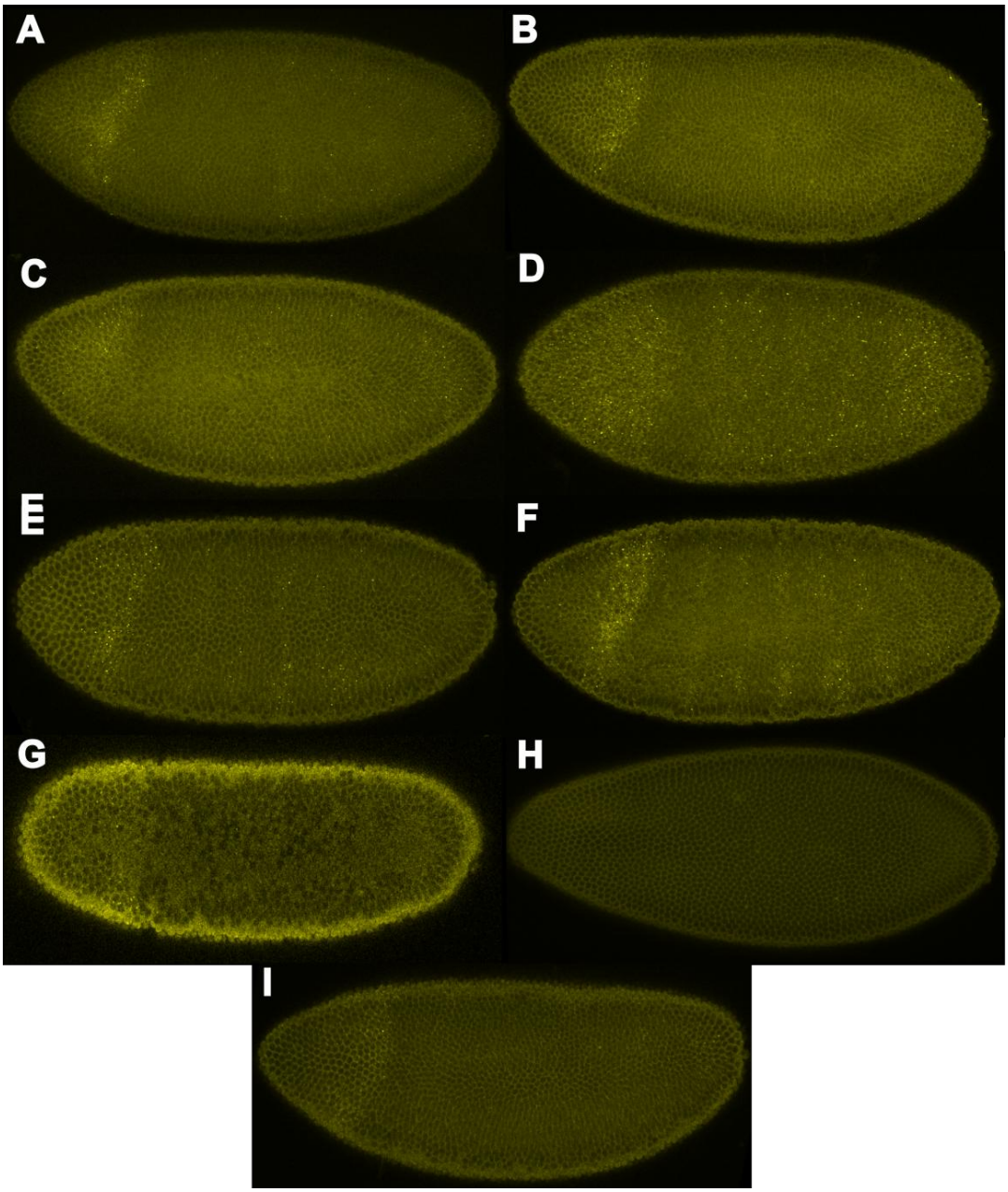

Figure 4.3: Results of Putative Enhancer Testing. (A) Expression of *lacZ* due to EveA enhancer is localized to the posterior region at a stripe at approximately 50% embryo length. (B) In the mutated EveA enhancer expression is lost in the posterior of the embryo. (C) EveB causes expression along the anterior and posterior poles of the embryo. (D) This anterior expression is not lost in the mutated EveB enhancer, in fact expression increases throughout the embryo, in rough stripes. (E) The anterior pole and weak stripes of *lacZ* expression are driven by the EveC enhancer. (F) The expression in the stripes is increased in the mutated EveC enhancer, however expression in the anterior cap is lost. (G) The KrA enhancer drives expression at the anterior pole. (H) Expression is lost in the mutated KrA enhancer. (I) The putative enhancer trap plasmid without any enhancer region (just the minimal promoter) drives a broad stripe of expression between 20-30% embryo length.

sources for subtle variation in gene expression; the binding sites responsible for these subtle effects should only be able to produce minimal levels of expression. The putative enhancers that were able to produce expression *in vivo* were mutated at the site of the SNP to the mutant version of this SNP. These mutated enhancers produced altered expression patterns (Fig. 4.3B,D,F,H). For the EveA enhancer, an AACA deletion was made corresponding to the indel within this region. This deletion results in loss of posterior expression found with the non-mutated enhancer. The alteration of the SNP within the EveB enhancer (from A to T) results in an increase in expression throughout the embryo in broad stripes. A loss in expression at the anterior pole of the embryo and simultaneously an increase in expression of stripes results from altering the EveC enhancer at the SNP (C to T mutation). Mutating the indel (C insertion) in the KrA enhancer, results in a loss of expression in the anterior cap of the embryo.

4.2.4 - Determining novel transcription factors

This ability of these enhancers to drive expression points to the presence of transcription factor binding sites within these putative enhancers. These transcription factor binding sites are most likely at the site of the SNP within the enhancer. Therefore these enhancer regions were then analyzed to determine which transcription factors were most likely binding and generating this expression. Position Weight Matrices were used to compile a list of transcription factor binding sites that may be present at the SNPs and indels of interest (Fig. B.3, Table 4.2). Where available, ChIP data was also used to rule out or suggest transcription factors (MacArthur et al. 2009).We ruled out for further

investigation any transcription factors already known to interact with the AP patterning system. The remaining transcription factors -- *gcm* (*glial cells missing*), *med* (*medea*), *otp* (*orthopedia*), *usp* (*ultraspiracle*), and *pan* (*pangolin*) -- represent possible novel components of the AP network. We then tested these transcription factors for their ability to affect *Krüppel* and Even-skipped expression using mutant fly lines (see Methods), in comparison to *yw* control expression patterns.

Table 4.2: Transcription factors identified as being likely candidates for binding to SNPs and indels of interest. Genes identified in bold were tested in this study.

Transcription Factor	**Enhancer Trap**	**Suggested By**	**Notes**
bcd	Eve3	PWM	AP patterning gene
cad	Eve2, Eve3	ChIP	already know interact
D	Eve2, Eve3	ChIP	already know interact
dl	Eve2, Eve8	ChIP	already know intereact
en	Eve3	PWM	AP patterning gene
gcm	**Eve2**	**PWM**	**tested here**
gt	Eve3	PWM	AP patterning gene
hkb	Kr4	PWM	eliminated by ChIP data
kni	Kr4	PWM	AP patterning gene
Kr	Kr4	PWM	AP patterning gene
Med	**Eve8, Kr4**	**ChIP**	**tested here**
otp	**Kr4**	**PWM**	**tested here**
pan	**Eve3**	**PWM**	**tested here**
slp1	Eve2, Eve3	ChIP	already tested
sna	Eve2	PWM	eliminated by ChIP data
tin	Kr4, Eve2	PWM	effects eve
tll	Eve8	PWM	eliminated by ChIP data
twi	Eve2, Eve3	ChIP	already tested
usp	**Eve8**	**PWM**	**tested here**

4.2.5 - gcm *effects positioning of* Kr *and Eve stripes 6 and 7*

The positioning of *Kr* and Eve stripes 6 and 7 was found to be effected by *gcm* (*glial cells missing*). We tested three different *gcm* RNAi lines, and found posterior shifts in the *Kr* domain and in the Eve 6 and 7 domains (see Fig. 4.4, Table B.3, Table B.4, and Materials and Methods). This shift in the positioning of Eve stripe 6 is consistent with the shift in Eve stripe 6 correlated with the indel suspected of being at a transcription factor binding site for *gcm* (per PWM data). This indel was explored through the EveA putative enhancer which produces expression in the posterior region of the embryo. *Kr* and Eve stripes 6 and 7 are expressed in this region. However, *gcm* is known to be expressed in the anterior half of the embryo (15-35% embryo length on the ventral side). This would suggest an indirect activing on *Kr* and *eve* by Gcm through some other intermediary signal. However, if Gcm does in fact have a direct interaction with *eve* and *Kr*, as our data suggests, this may be due to diffusion of Gcm towards the posterior of the embryo or some low level of expression in this region.

4.2.6 - Shifts in Eve stripes and Kr *borders are observed in* usp *mutants*

The maternal gene *ultraspiracle* (*usp*) was found to effect expression of *Kr* and Eve in the early embryo. *usp* is expressed throughout the early embryo (Hammonds et al. 2013, Tomancak et al. 2007, Tomancak et al. 2002), however only weak expression remains by mid-NC14. Two fly lines mutated for *usp*, one amorphic and one hypomorphic, were found to produce shifts in the borders of *Kr* and Eve stripes 1, 2 and 7 (Fig. 4.4, Table B.3, Table B.4, and Materials and Methods). *usp* was tested because PWM analysis points to a binding site at the SNP in EveC. This SNP was correlated with

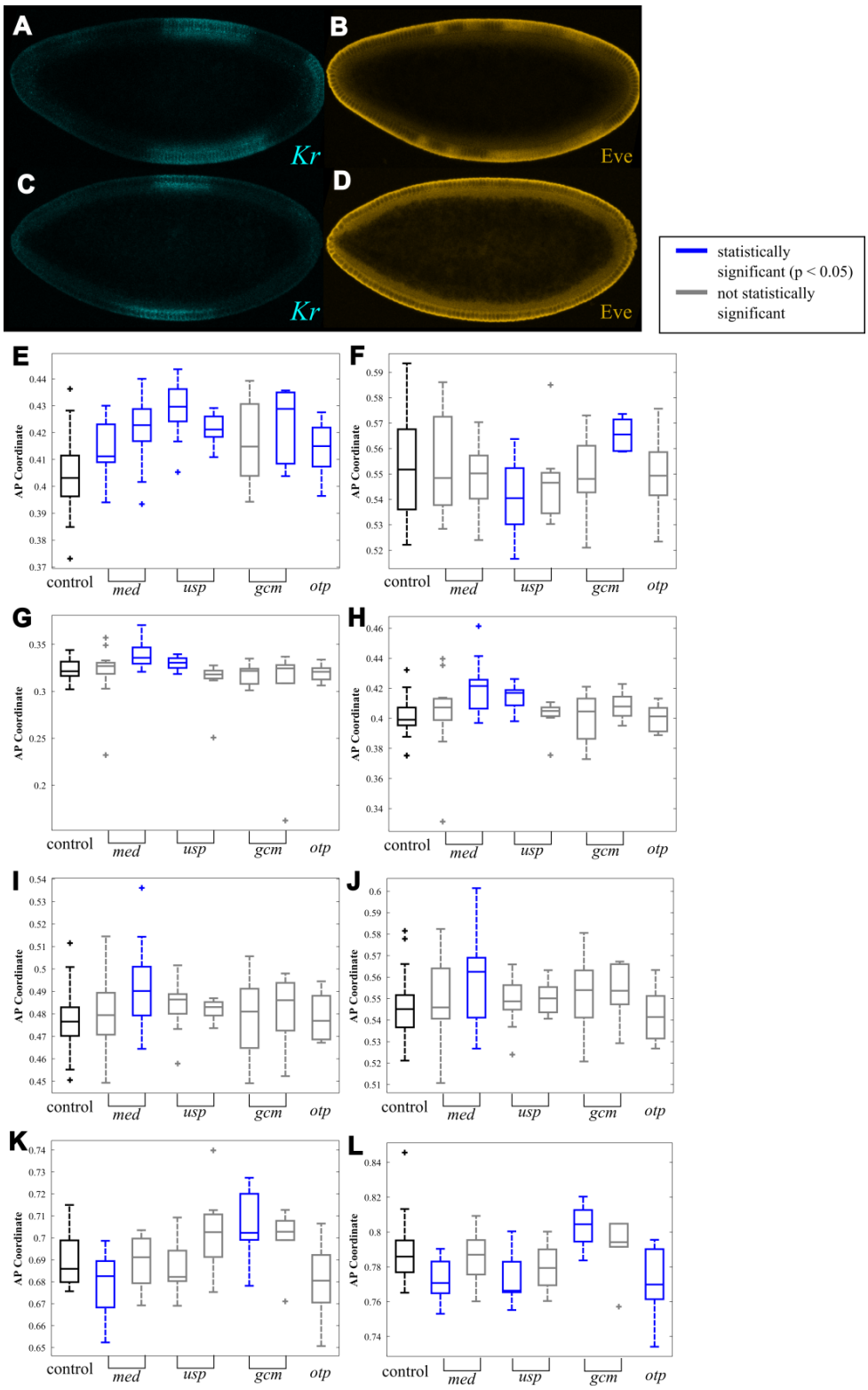

Figure 4.4: Shifts in *Kr* and Eve seen in mutants. In *pan* mutant BS26743 (A) *Kr* and (B) Eve expression. In mutant BS22312, (A) *Kr* and (B) Eve expression. Variation in positioning of (E) *Kr* anterior border, (F) *Kr* posterior border, (G) Eve stripe 1, (H) Eve stripe 2, (I) Eve stripe 3, (J) Eve stripe 4, (K) Eve stripe 6, and (L) Eve stripe 7.

a shift in Eve stripe 5; this enhancer trap produces expression in the anterior pole and in stripes throughout the embryo. These stripes of expression due to the putative EveC enhancer (Fig. 4.3) are consistent with the changes seen in the positioning of *Kr* and Eve stripes 1, 2, and 7 observed in the lines with mutated levels of *usp* expression (Fig. 4.4).

4.2.7 - medea *results in subtle shifts in* Kr *and Eve throughout the AP axis*

ChIP data (supported by PWM analysis) suggests a binding site for *medea* (*med*) in either EveA or EveB, at the SNP contained within these enhancer traps. The SNPs within the EveA and EveB putative enhancers were correlated with a shift in Eve stripe 6 and stripe 5 respectively. Testing of *medea* mutants showed shifts in both the *Kr* anterior border and positioning of all Eve stripes (Fig. 4.4, Table B.3, Table B.4, and Materials and Methods). These effects spread throughout the entire embryo are consistent with expression observed due to the mutant EveB putative enhancer. Since *med* is maternally deposited and found ubiquitously throughout the early embryo these effects are consistent with the region where Medea is known to be present. Medea is an effector molecule for the Decapentaplegic (Dpp) pathway, which patterns the dorsal axis of the embryo (Raftery, Sutherland 2003). It is therefore unclear how Medea may act to affect expression along the AP axis, particularly in DV positions where Dpp signaling is not active.

4.2.8 - Positioning of Kr *and Eve stripe 7 are shifted in* otp *mutants*

orthopedia (*otp*) was suggested from PWM analysis of possible binding sites at the SNP in the KrA putative enhancer. This enhancer drives expression at the anterior

pole. The SNP in KrA was found to be correlated with a shift in the anterior border of *Kr*. A fly line deficient for *otp* (via RNAi knockdown) results in a shift in the expression of the anterior border of *Kr* and in Eve stripe 7 (Fig. 4.4, Table B.3, Table B.4, and Materials and Methods). Otp is a Hox gene which is known to be active following gastrulation (Hammonds et al. 2013, Tomancak et al. 2007, Tomancak et al. 2002, Simeone et al. 1994). However, RNAseq data has found transcripts in this time period (2-4 hour old embryos) (Graveley et al. 2011). This suggests some low level of *otp* expression that affects *Kr* and possibly also Eve expression.

4.2.9 - pangolin *mutations result in large shifts in* Kr *and Eve expression*

The most significant effects were observed for fly lines mutated for *pangolin* (*pan*). *pan* expression is expressed ubiquitously throughout the early embryo and was tested based on PWM analysis of EveB. The EveB putative enhancer generates expression at the anterior pole; however the mutated enhancer trap generates expression throughout the embryo. Two different mutant fly lines, one expressing an RNAi knockdown (BS26743) and one with an insertion in the gene (BS22312) were tested. The *Kr* and Eve patterns produced by these fly lines were variable within these mutant populations. Significant differences compared to wild-type patterns were observed, examples for each fly line are seen in Fig. 4.4. For the insertion fly line (Fig. 4.4C,D), at least half of the flies tested exhibited this pattern. Over a quarter of the flies tested in the RNAi knockdown line (Fig 4.4A,B) showed this large expansion of the *Kr* domain and subsequent disruption in the Eve pattern. These fly lines were viable and these variations do not appear to effect survival.

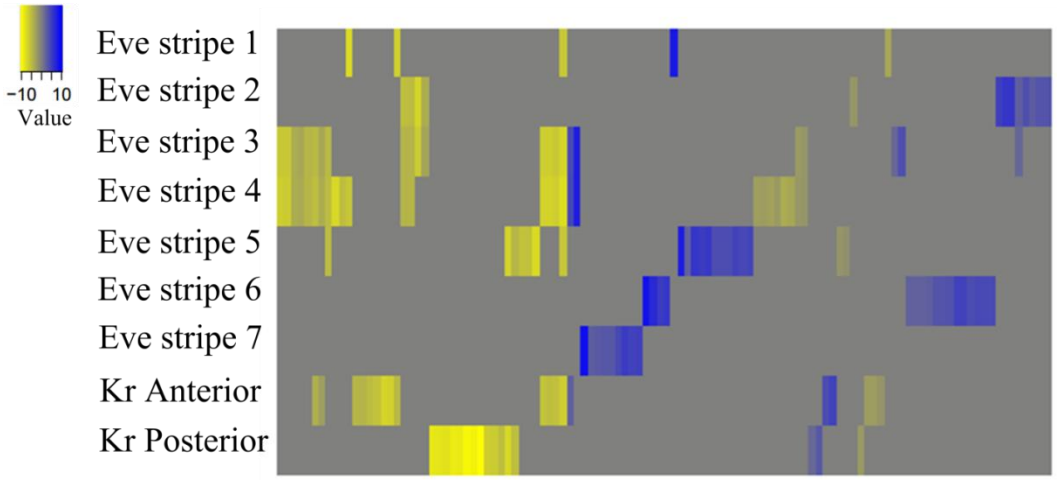

Figure 4.5: Differentially expressed genes linked to changes in Eve and *Kr* positioning. Results of RNAseq analysis, where genes are correlated with changes in positioning of *Kr* and Eve.

Table 4.3: Genes whose expression levels are correlated with positioning of *Kr* and Eve

Gene	**Expression Pattern**
GNBP3	broad DV stripe
e(r)	faint broad DV stripe
CG10672	faint expression at center
CG10591	faint stripe 30-90% EL
CG4449	stripes
CG6180	stripes
Pvf2	two DV stripes
eIF-5A	ubiquitous
RpS25	ubiquitous
CG1354	ubiquitous
CG7091	ubiquitous
Rpb5	ubiquitous - missing AP cap
CG8089	ventral side
SelR	ventral side

4.2.10 - Variations in the transcriptome suggest novel AP patterning genes

The transcriptomes of the same DGRP fly lines (and *yw* control) were determined using RNAseq. This data was analyzed using the edgeR algorithm with a false discovery

rate cut-off of 0.1%. For this analysis the lines were grouped based on positioning of its borders of *Kr* and Eve stripes. This analysis was carried out separately for each position, where three groups were created for each position (anterior shift, no shift, or posterior shift compared to yw control). From this analysis we can create heat maps to highlight differentially expressed transcripts correlated with variations in the positioning of *Kr* and Eve (Fig.4.5). These differentially expressed genes which are known to be expressed in the early embryo are listed in Table 4.3 (a complete list of all differentially expressed genes is listed in Table B.5).

4.3 - Conclusions

We found that SNPs and indels present between natural fly lines correlate with subtle variations in *Kr* and Eve expression patterns. Testing of possible regulatory regions containing these genomic variations identified certain SNPs which are able to produce activity and are therefore likely transcription factor binding sites. Through this analysis we were able to identify novel transcription factors (*usp*, *med*, *gcm*, and *otp*) that when mutated produce subtle variations in the position of *Kr* and Eve stripes. These subtle variations are consistent with the variations seen when the SNPs suspected of being their binding sites are mutated. This mutant analysis also identified *pangolin*, which is able to produce large variations in *Kr* and Eve. The results of these analyses points to a greater network of genes involved in the anterior-posterior patterning system. In addition, this study demonstrates the ability of a SNP to produce subtle, yet identifiable variations in gene expression. The methodology used in this study can be applied to further studies using the DGRP fly lines. A larger sample of lines (or all

DGRP lines) can be tested, which would allow for SNPs to be explored throughout the genome (at significant distance from the gene of interest). This can identify trans-acting factors which were previously difficult to identify.

4.4 – Materials and Methods

4.4.1 - Embryo Staining and Image Collection

Embryos 2-4 hrs after egg lay, were fixed using formaldehyde per standard protocols. Subsequently, fluorescent *in situ* hybridization was used to image RNA and protein expression per published protocols ((Kosman et al. 2004) with proteinase K treatment omitted). RNA probes for *lacZ* (biotin conjugated) and *Kr* (flourescein conjugated) were used. Primary antibodies to biotin (goat, anti-biotin, 1:5000, gift from Immunoreagents), Eve (mouse anti-Eve, 1:10; Developmental Studies Hybridoma Bank), and flourescein (rabbit, anti-flourescein, 1:500; ThermoFisher Scientific). Secondary antibodies used were Alexa Flour 488 donkey anti-rabbit (ThermoFisher Scientific), Alexa Flour 546 donkey anti-goat (ThermoFisher Scientific), and AlexaFlour 546 donkey anti-mouse (ThermoFisher Scientific). Images were taken using a Zeiss Confocal microscope. Quantification of *Kr* and Eve was performed on images taken at the mid-saggittal plane and analyzed using Matlab (see Ch. 3).

4.4.2 - Plasmid Construction

The putative enhancer plasmids were cloned into vector (Liberman, Stathopoulos 2009) using EcoRI, BglII, or AscI. The Evep:lacZ plasmid is the vector with no enhancer trap inserted. Enhancer regions were amplified from genomic DNA from yw flies using

primers listed in Table B.1. Mutations were introduced into the enhancer traps using mutagenesis PCR (primers listed in Table B.2). All PCR was carried out using Q5 Polymerase (NEB).

4.4.3 - Fly lines

yw was used as a laboratory control strain. Natural variation fly lines were provided by Trudy MacKay [reference]. The lines denoted in this paper as 1-13, are RAL41, RAL57, RAL105, RAL306, RAL307, RAL315, RAL317, RAL360, RAL705, RAL761, RAL765, RAL799, and RAL801. Mutants for suspect transcription factors were from Blooming Stock Center: *Medea* (BS9033 [Med^{1}] and BS9006 [Med^{5}] from ethyl methanesulfonate mutagenesis), *pangolin* (BS26743 and BS22312- both RNAi knockdown lines), *ultraspiracle* (hypomorphic line BS4660 and amorphic line BS31414), *glial cells missing* (RNAi knockdown lines BS28913, BS31518, and BS31519), and *orthopedia* (RNAi knockdown line BS57582). The enhancer trap and mutant enhancer trap fly lines were created by injection and incorporation of plasmid constructs into the 68A4 landing site (yw; attP2 flies) for the KrA:lacZ, EveL:lacZ, EveI:lacZ, and EveK:lacZ injections were performed by Model System injections. KrB:lacZ, KrC:lacZ, KrD:lacZ,, EveA:lacZ, EveB:lacZ, EveC:lacZ, EveD:lacZ:, EveE:lacZ, EveF:lacZ, EveG:lacZ, EveH:lacZ, EveJ:lacZ injections were performed by Genetic Services, Inc. KrBmut:lacZ, EveBmut:lacZ, EveCmut:lacZ, Evep:lacZ and EveGmut:lacZ injections were performed by GenetiVision Inc.

4.4.4 - Identification of Novel Transcription Factors

Identification of novel transcription factors using position weight matrices were carried out for the region surrounding the SNP within the enhancers that generated expression *in vivo*. The position weight matrices used were obtained from (MacArthur et al. 2009, Down et al. 2007, Kulakovskiy, Makeev 2009, Noyes et al. 2008a, Noyes et al. 2008b, Papatsenko, Levine 2007, Bergman, Carlson & Celniker 2005, Pollard et al. 2006). Probability of a transcription factor binding site being present within a given series was calculated by multiplying the probability of the given nucleotide at each position in the sequence and dividing this by the probability of a random sequence (calculated from a 10,000 bp *Drosophila melanogaster* exon region). Transcription factors where $p < 0.0005$ were explored. Chromatin immunoprecipitation data was taken from MacArthur et al, 2009 (MacArthur et al. 2009).

4.4.5 - RNAseq

RNA was obtained from 0-1 hr old embryos. cDNA libraries were constructed using 350 ng of RNA (concentration determined via BioAnalyzer). Libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (E740S) and NEBNext Poly(A) mRNA Magnetic Isolation Module (1E7490S) and associated NEBNext Multiplex Oligos for Illumina (Primers Set 1 and Primers Set 2) (E7335S and E7500S). Sequencing was carried out by North Carolina State University's Genome Sciences Laboratory using Illumina HiSeq 2500. Transcript reads were compared to the *Drosphila*

melanogaster Genome version 5 using BWAaln. Statistical analysis was carried out using edgeR in R.

References

Bergman, C., Carlson, J. & Celniker, S. 2005, "Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila melanogaster", *Bioinformatics,* vol. 21, no. 8, pp. 1747-1749.

Bieler, J., Pozzorini, C. & Naef, F. 2011, "Whole-embryo modeling of early segmentation in Drosophila identifies robust and fragile expression domains.", *Biophys J,* vol. 101, no. 2, pp. 287-296.

Down, T.A., Bergman, C.M., Su, J. & Hubbard, T.J.P. 2007, "Large-scale discovery of promoter motifs in Drosophila melanogaster", *Plos Computational Biology*, vol. 3, no. 1, pp. 95-109.

Fujioka, M., Emi-Sarker, Y., Yusibova, G.L., Goto, T. & Jaynes, J.B. 1999, "Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients.", *Development,* vol. 126, no. 11, pp. 2527-2538.

Graveley, B.R., Brooks, A.N., Carlson, J., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., Brown, J.B., Cherbas, L., Davis, C.A., Dobin, A., Li, R., Lin, W., Malone, J.H., Mattiuzzo, N.R., Miller, D., Sturgill, D., Tuch, B.B., Zaleski, C., Zhang, D., Blanchette, M., Dudoit, S., Eads, B., Green, R.E., Hammonds, A., Jiang, L., Kapranov, P., Langton, L., Perrimon, N., Sandler, J.E., Wan, K.H., Willingham, A., Zhang, Y., Zou, Y., Andrews, J., Bickel, P.J., Brenner, S.E., Brent, M.R., Cherbas, P., Gingeras, T.R., Hoskins, R.A., Kaufman, T.C., Oliver, B. & Celniker, S.E. 2011, "The developmental

transcriptome of Drosophila melanogaster", *Nature*, vol. 471, no. 7339, pp. 473-479.

Hammonds, A.S., Bristow, C.A., Fisher, W.W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E. & Celniker, S.E. 2013, "Spatial expression of transcription factors in Drosophila embryonic organ development", *Genome biology*, vol. 14, no. 12, pp. R140.

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ramia, M., Tarone, A.M., Turlapati, L., Zichner, T., Zhu, D., Lyman, R.F., Magwire, M.M., Blankenburg, K., Carbone, M.A., Chang, K., Ellis, L.L., Fernandez, S., Han, Y., Highnam, G., Hjelmen, C.E., Jack, J.R., Javaid, M., Jayaseelan, J., Kalra, D., Lee, S., Lewis, L., Munidasa, M., Ongeri, F., Patel, S., Perales, L., Perez, A., Pu, L., Rollmann, S.M., Ruth, R., Saada, N., Warner, C., Williams, A., Wu, Y., Yamamoto, A., Zhang, Y., Zhu, Y., Anholt, R.R.H., Korbel, J.O., Mittelman, D., Muzny, D.M., Gibbs, R.A., Barbadilla, A., Johnston, J.S., Stone, E.A., Richards, S., Deplancke, B. & Mackay, T.F.C. 2014, "Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines", *Genome research*, vol. 24, no. 7, pp. 1193-1208.

Jaeger, J. 2011, "The gap gene network", *Cellular and Molecular Life Sciences*, vol. 68, no. 2, pp. 243-274.

Jiang, P., Ludwig, M.Z., Kreitman, M. & Reinitz, J. 2015, "Natural variation of the expression pattern of the segmentation gene even-skipped in melanogaster", *Developmental biology*, vol. 405, no. 1, pp. 173-181.

Kosman, D., Mizutani, C., Lemons, D., Cox, W., McGinnis, W. & Bier, E. 2004, "Multiplex detection of RNA expression in Drosophila embryos", *Science,* vol. 305, no. 5685, pp. 846-846.

Kulakovskiy, I.V. & Makeev, V.J. 2009, "Discovery of DNA motifs recognized by transcription factors through integration of different experimental sources", *Molecular Biophysics,* vol. 54, no. 6, pp. 667-667-674.

Liberman, L.M. & Stathopoulos, A. 2009, "Design flexibility in \textitcis-regulatory control of gene expression: synthetic and comparative evidence.", *Dev Biol,* vol. 327, no. 2, pp. 578-589.

Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes Jose, Walczak, A.M. & Coppey, M.a.D.,Nathalie 2013, "Live imaging of bicoid-dependent transcription in Drosophila embryos.", *Curr Biol,* vol. 23, no. 21, pp. 2135-2139.

MacArthur, S., Li, X., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keraenen, S.V.E., Knowles, D.W., Stapleton, M., Bickel, P., Biggin, M.D. & Eisen, M.B. 2009, "Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions", *Genome biology,* vol. 10, no. 7, pp. R80.

Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S.a.H.,Yi, Magwire, M.M., Cridland, J.M., Richardson, M.F., Anholt, R.R.H., Barrón, M., Bess, Crystal and Blankenburg,Kerstin Petra, Carbone, M.A., Castellano, D., Chaboub, L., Duncan, L., Harris, Z., Javaid, M., Jayaseelan, J.C., Jhangiani, S.N., Jordan, K.W., Lara, F., Lawrence, Faye and Lee,Sandra L.,

Librado, P., Linheiro, R.S., Lyman, R.F., Mackey, A.J., Munidasa, M., Muzny, D.M., Nazareth, L., Newsham, I., Perales, L.a.P.,Ling-Ling, Qu, C., R\`amia, M., Reid, J.G., Rollmann, S.M., Rozas, J., Saada, N., Turlapati, L., Worley, K.C., Wu, Y., Yamamoto, A., Zhu, Y., Bergman, C.M., Thornton, K.R., Mittelman, D. & Gibbs, R.A. 2012, "The Drosophila melanogaster Genetic Reference Panel.", *Nature,* vol. 482, no. 7384, pp. 173-178.

McGinnis, W. & Krumlauf, R. 1992, "Homeobox genes and axial patterning", *Cell,* vol. 68, no. 2, pp. 283.

Morrison, A.H., Scheeler, M., Dubuis, J. & Gregor, T. 2012, "Quantifying the Bicoid morphogen gradient in living fly embryos.", *Cold Spring Harb Protoc,* vol. 2012, no. 4, pp. 398-406.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H. & Wolfe, S.A. 2008a, "Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites", *Cell,* vol. 133, no. 7, pp. 1277-1289.

Noyes, M.B., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M.H. & Wolfe, S.A. 2008b, "A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system", *Nucleic acids research,* vol. 36, no. 8, pp. 2547-2560.

Papatsenko, D. & Levine, M. 2007, "A rationale for the enhanceosome and other evolutionarily constrained enhancers", *Current Biology,* vol. 17, no. 22, pp. R955-R957.

Perkins, T.J., Jaeger, J., Reinitz, J. & Glass, L. 2006, "Reverse engineering the gap gene network of Drosophila melanogaster", *Plos Computational Biology*, vol. 2, no. 5, pp. 417-428.

Pollard, D.A., Moses, A.M., Iyer, V.N. & Eisen, M.B. 2006, "Detecting the limits of regulatory element conservation and divergence estimation using pairwise and multiple alignments", *Bmc Bioinformatics*, vol. 7, pp. 376.

Raftery, L. & Sutherland, D. 2003, "Gradients and thresholds: BMP response gradients unveiled in Drosophila embryos", *Trends in Genetics*, vol. 19, no. 12, pp. 701-708.

Simeone, A., D'Aice, M., Nigro, V., Casanova, J., Graziani, F., Acampora, D. & Avantaggiato, V. 1994, "Orthopedia, a Novel Homeobox-Containing Gene Expressed in the Developing Cns of both Mouse and Drosophila", *Neuron*, vol. 13, no. 1, pp. 83-101.

Small, S., Blair, A. & Levine, M. 1992, "Regulation of even-skipped stripe 2 in the Drosophila embryo.", *EMBO J*, vol. 11, no. 11, pp. 4047-4057.

Small, S., Blair, A. & Levine, M. 1996, "Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo.", *Dev Biol*, vol. 175, no. 2, pp. 314-324.

Sokolowski, T.R., Erdmann, T. & ten Wolde Pieter Rein 2012, "Mutual repression enhances the steepness and precision of gene expression boundaries.", *PLoS Comput Biol*, vol. 8, no. 8, pp. e1002654.

Tamari, Z. & Barkai, N. 2012, "Improved readout precision of the Bicoid morphogen gradient by early decoding.", *J Biol Phys*, vol. 38, no. 2, pp. 317-329.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E. & Rubin, G.M. 2007, "Global analysis of patterns of gene expression during Drosophila embryogenesis", *Genome biology,* vol. 8, no. 7, pp. R145.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E. & Rubin, G.M. 2002, "Systematic determination of patterns of gene expression during Drosophila embryogenesis", *Genome Biol.,* vol. 3, no. 12, pp. 1-1-88.

von Dassow, G., Meir, E., Munro, E. & Odell, G. 2000, "The segment polarity network is a robust development module", *Nature,* vol. 406, no. 6792, pp. 188-192.

CHAPTER 5

Creation and Modeling of Synthetic Sharpness and Scaling Gene Networks in Drosophila Embryos

Ashley A. Jermusyk, Nicholas P. Murphy, Ethan C. Simmons, and Gregory T. Reeves

5.1 – Background

Previous studies of endogenous signaling networks have sought to understand and identify specific interactions within complex patterning systems. These complex signaling networks are made up of commonly found simple gene circuits. One of these gene circuits, cross-repression, can be set-up in such a manner to exhibit multiple system outputs. These outputs, which are not a part of the individual parts, are known as emergent properties. In the case of cross-repression these emergent properties are are sharpness and scaling (Jagla et al. 2002, Jaeger 2011, Xu et al. 2014, Reversade, De Robertis 2005, Houchmandzadeh, Wieschaus & Leibler 2005). To better understand the specific ability of cross-repression to exhibit these two emergent properties, we examine this circuit in isolation using a synthetic gene network. By looking at this circuit in this manner we will be able to understand the biological limits and factors most important for the proper function of this network and how different emergent properties can be exhibited by the system. We suggest our findings may generalize to other multi-cellular systems.

5.1.1 - Patterning

The initial patterning in multi-cellular organisms is set-up by morphogens. Morphogen gradients are formed when a protein is produced and then diffuses away from a localized source (Wolpert 1969). This creates an exponentially decaying concentration gradient which then gives rise to expression of downstream genes at different positions based on concentration thresholds of the morphogen (Wolpert 1969). This gives rise to

broad domains of expression; however these broad domains are further refined by interactions between these downstream genes (Wolpert 1969).

5.1.2 - Sharpness

When a cross-repression gene network is set-up in a certain manner it is able to produce sharp borders. In this system the border of expression of overlapping genes is refined by this mutual repression. This emergent property is observed in the anterior-posterior patterning system within the *Drosophila melanogaster* embryo (Perkins et al. 2006, Jaeger 2011, Sokolowski, Erdmann & ten Wolde Pieter Rein 2012). In this system, the morphogen gradient, Bicoid, activates expression of the gap genes (Perkins et al. 2006, Jaeger 2011). The domains of the gap genes are refined by cross-repression between neighboring genes (Perkins et al. 2006, Jaeger 2011, Sokolowski, Erdmann & ten Wolde Pieter Rein 2012). This produces distinct regions of expression for each of the gap genes which then go onto regulate the expression of downstream patterning genes. A synthetic cross-repression network was created where two genes are expressed in overlapping domains (*lacI* and *tetR*) due to a morphogen gradient (Gal4). This repression between these genes leads to sharp borders between these domains.

5.1.3 - Scaling

Scaling is an emergent property that can also be achieved using a cross-repression network. This phenotype is created when cross-repression occurs between two different genes which are activated by opposing morphogen gradients. This allows for tissues to input total tissue size and thereby reproducibly position genes at a given percent tissue

size. This mechanism has been proposed to achieve scaling in anterior-posterior patterning within the early *Drosophila* embryo. Specifically, it has been proposed that Hunchback, which is expressed at 50% embryo length, is positioned robustly despite noise in the Bicoid morphogen gradient (Houchmandzadeh, Wieschaus & Leibler 2005). A second opposing morphogen gradient, such as Caudal, may be responsible for providing tissue size input into this system (Morishita, Iwasa 2009, McHale, Rappel & Levine 2006). Scaling may also occur in zebrafish embryos where two opposing morphogen gradients, Bone Morphogenic Protein and Nodal, are suffiecient for patterning (Xu et al. 2014, Reversade, De Robertis 2005). The ability of scaling to reproducibly position domains regardless of embryo size was tested using a synthetic gene network. In this synthetic network, two opposing morphogen gradients activate expression of genes expressed on opposite ends of the tissue. These two proteins (TetR and cI) repress one another. The ability of TetR and cI to be expressed reproducibly at 50% embryo length due to this scaling mechanism can be tested.

5.2 - Results

5.2.1 - Preliminary Testing of the Sharpness Network

A synthetic sharpness network was created and expressed along the anterior-posterior axis. A synthetic Gal4 gradient was formed with its peak at the anterior pole. Gal4 is able to activate genes downstream of *UAS* sites (upstream activating sequences) (Lohr, Venkov & Zlatanova 1995, Giniger, Ptashne 1988, Elliott, Brand 2008). In this case, Gal4 activates *UAS*-linked *lacI* and *tetR*. By linking *lacI* and *tetR* to a different number of *UAS* sites (three for *lacI* and seven for *tetR*), they are expressed in different,

overlapping domains (see Chapter 3 and (Lohr, Venkov & Zlatanova 1995, Giniger, Ptashne 1988, Elliott, Brand 2008)). In addition, *lacI* and *tetR* cross-repress each other, as they are linked to *tetO* and *laO* sites, respectively (Brown et al. 1987, Mortlock, Low & Crisanti 2003). Because both *lacI* and *tetR* are fusion genes with the *Krüppel* repressor domain (KrRD), when *lacI* and *tetR* bind to their respective *lacO* and *tetO* sites, the downstream gene is repressed (Licht et al. 1994).

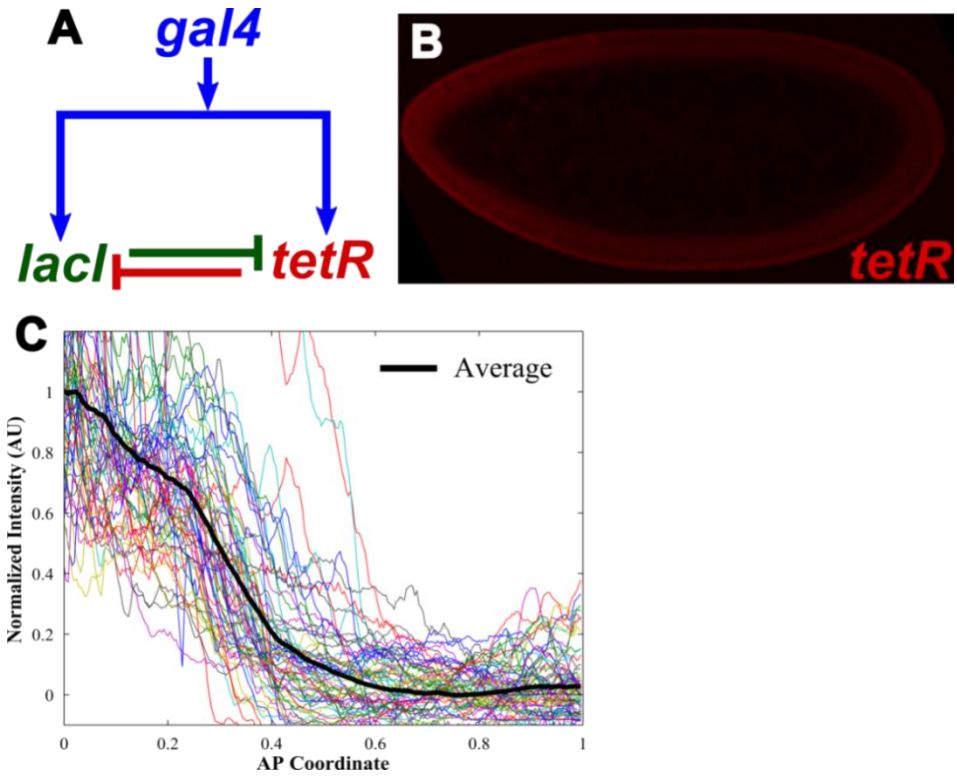

Figure 5.1: Sharpness network. (A) network diagram, Gal4 activates LacI and TetR, LacI and TetR repress each other. (B) *tetR* mRNA expression at the mid-saggittal section of an embryo with the *UAS*x7:*lacO*x8:*tetR-krRD* construct and no LacI. (C) Quantification of *tetR* expression due to Gal4 (no LacI), where each colored curve is the dorsal or ventral side of one embryo and the average is shown in black.

These *lacI* and *tetR* constructs, shown in were incorporated into flies (see Methods). The expression pattern of *tetR* due to Gal4 (without LacI in the embryo) was determined in this network (Fig. 5.1) via fluorescent *in situ* hybridization and subsequent quantification using image analysis protocols (see Materials and Methods). As expected

from previous studies (see Chapter 3) the *tetR* expression is in a broad, graded domain. Future work will be needed to quantify the domain for the other component (*lacI*) of this network. In addition the changes in the domains of *lacI* and *tetR* in the final network will need to be fully characterized.

5.2.2 - Preliminary Testing of the Scaling Network

In the *Drosophila melanogaster* embryo two opposing morphogen gradients exist along the anterior-posterior axis, Bicoid and Nanos (Gavis, Lehmann 1992, Macdonald, Struhl 1988, Weil, Forrest & Gavis 2006, Driever, Nusslein-Volhard 1988, Janody et al. 2001, Gavis, Lehmann 1994). We used the 3'UTR of these two genes to create opposing gradients of exogenous activators (Gal4 and LexA-VP16 respectively) (Gavis, Lehmann 1992, Macdonald, Struhl 1988, Weil, Forrest & Gavis 2006, Driever, Nusslein-Volhard 1988, Janody et al. 2001, Gavis, Lehmann 1994, Spirov et al. 2009, Huang, Rusch & Levine 1997). These opposing morphogen gradients are able to drive expression of downstream genes linked to binding sequences for these activators (Lohr, Venkov & Zlatanova 1995, Giniger, Ptashne 1988, Elliott, Brand 2008, Yagi, Mayer & Basler 2010, Lai, Lee 2006). The Gal4:bcd 3'UTR gradient is used to drive expression of *UAS*-linked *tetR-KrRD* (Lohr, Venkov & Zlatanova 1995, Giniger, Ptashne 1988, Elliott, Brand 2008). The opposing LexA-VP16 gradient activates expression of *lexO* linked *cI-AntPHD* (Yagi, Mayer & Basler 2010, Lai, Lee 2006). The *tetR-KrRD* construct is designed with *OR* sites, which are bound by cI-AntPHD, repressing *tetR* expression (Meyer, Maurer & Ptashne 1980). Similarly, the *cI-AntPHD* construct contains *tetO* sites

for the binding of, and repression by, TetR-KrRD (Mortlock, Low & Crisanti 2003). This complete network (Fig. 5.2) was incorporated into flies.

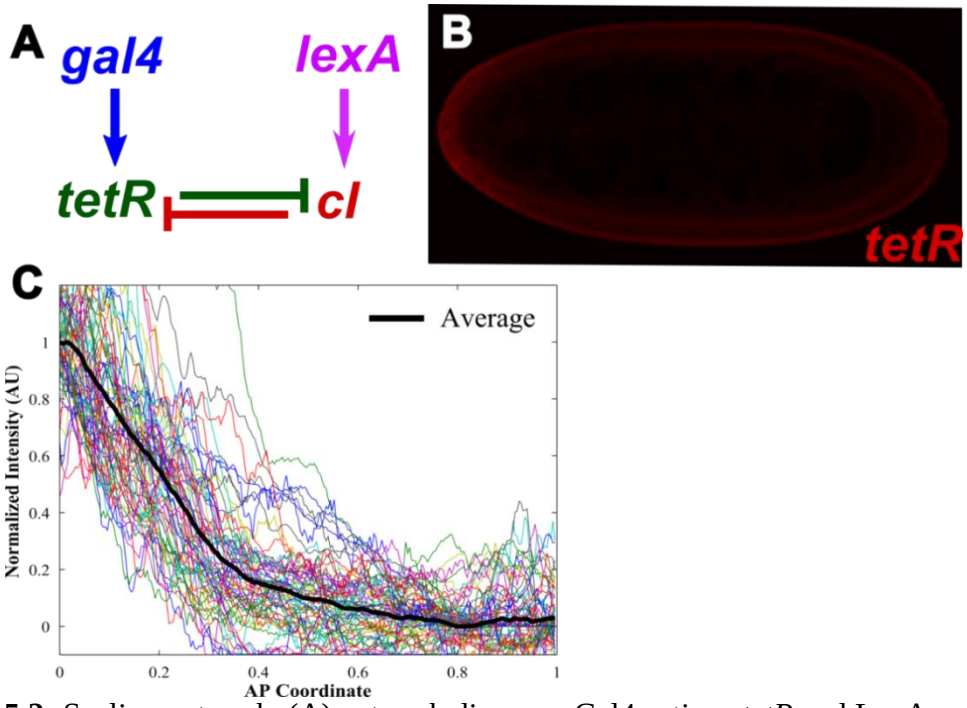

Figure 5.2: Scaling network. (A) network diagram, Gal4 actives *tetR* and LexA activates *cI*. cI and TetR undergo cross-repression. (B) *tetR* mRNA expression at the mid-saggittal section of an embryo with the *UAS*x7:*OR1-3*:*tetR-krRD* construct and no cI. (C) Quantification of *tetR* expression due to Gal4 (no cI), where each colored curve is the dorsal or ventral side of one embryo and the average is shown in black.

Using fluorescent *in situ* hybridization and subsequent analysis (see Materials and Methods), the expression profile for *tetR* (due to Gal4 without cI and LexA) in the early embryo was determined (Fig. 5.2). *tetR* is expressed in the anterior third of the embryo in a graded domain. The positioning of the border of this domain can be described by a single statistic by determining where the profile is at 50% maximum. This occurs at 0.217 ± 0.077 embryo length. The precision with which this domain is patterned should increase with the input of cI into the embryo. This is because variability due to changes in embryo size should be corrected by the input of cI which is patterned by the opposing

morphogen gradient of LexA:nos 3'UTR. This change in the *tetR* domain, as well as *cI* expression will need to be determined in the embryo.

5.3 - Future Work

While some preliminary data was obtained on these networks, there is still much work needed to characterize these systems. The profiles for *lacI* (sharpness network) and *cI* (scaling network) need to be determined, both with and without *tetR* present. Combined with our current data for *tetR*, this missing data will describe the each of the network components in isolation. The expression in the complete network (with *tetR* and either *cI* or *lacI* as applicable) must also be measured.

In the sharpness network a sharp border should be formed between the domains of *lacI* and *tetR*. Additionally, the expression of *tetR* will decrease where it overlaps with the domain of *lacI* (due to repression by LacI binding at *lacO* sites). In the scaling network, the positioning of both *cI* and *tetR* should be more reproducible since error due to differentially sized embryos will be eliminated.

If these networks do not show the expected patterns, changes should be made to the experimental system. One reason for this would be these genes are not expressed at the expected domains; if this is a problem, it should be apparent from our data of each gene in isolation. Due to our preliminary testing of *tetR* and previous work with the *UAS* system (see Ch. 3), this is most likely not an issue with our sharpness network or with *tetR* in the scaling network since the ability of *UAS* sites to drive expression over a given range is well established for our system. Without the data for our final network we cannot determine if the concentration of proteins within the system are appropriate. This

can be adjusted by changing the copy number of transgenes within the system or altering the promoter. The repression by the proteins in the system must also be sufficient to achieve the desired cross-repression motif. This can be adjusted by adding or decreasing the number of binding sites for the protein (*lacO*, *tetO*, or *OR* sites as appropriate).

5.4 - Conclusions

Synthetic gene networks are a useful tool in understanding small gene circuits in isolation. Cross-repression can be explored using two different spatially expressed systems, exhibiting sharpness and scaling. The constructs within these networks were created and incorporated into flies. Only preliminary data was obtained for these systems. Future work will be needed to characterize these experimental networks. Using these experimental networks we will be able to gain insights in how scaling and sharpness can be achieved using cross-repression in the tissue patterning system.

5.5 - Materials and Methods:

5.5.1 - Plasmid Construction

*tetO*x4 and *tetO*x7 were obtained by inserting the *tetO* sequence (TCCCTATCAGTGATAGAGATCT) into pBlueScript II SK(+) (from Addgene) using *Xho*I and *Sal*I. Following the addition of one *tetO* site, the plasmid was digested with *Sal*I and *Not*I as well as *Xho*I and *Not*I, these two purified digestion products were ligated to obtain two *tetO* sites in tandem. This process was repeated as needed to obtain *tetO*x4 and tetOx7. The *UAS*x3:*gal80* plasmid (from gal80 paper) was digested with XhoI and ligated with *tetO*x4 (digested with *Xho*I and *Sal*I) to obtain *UAS*x3:*tetO*x4. Subsequently,

*UAS*x3:*tetO*x4 was digested with *Not*I and *Kpn*I and *lacI* (PCR amplified from pBlueScript) was inserted to obtain *UAS*x3:*tetO*x4:*lacI*. Finally, *krRD* (PCR amplified from genomic DNA (Licht et al. 1994)) was inserted into *UAS*x3:*tetO*x4:*lacI* digested with *Kpn*I and *Xba*I to obtain the final *UAS*x3:*tetO*x4:*lacI:KrRD* plasmid. The *lacO* sequence (TCGAATTGTGAGCGCTCACAAT) was inserted into pBlueScript and multimerized in the same manner to obtained *lacO*x8. *UAS*x7was created by inserting the *UAS* sequence (TGCGGAGTACTGTCCTCCGAG) into pBlueScript and subsequent multimerization. *UAS*x7:*lacO*x8:*tetR-KrRD* was obtained by inserting *UAS*x7 (digested out of *UAS*x7 in pBlueScript), l*acO*x8 (digested out of *lacO*x8 in pBlueScript), *tetR* (PCR amplified from pLenti CMV TetR Blast (716-1) [Addgene Plasmid # 17492] (Campeau et al. 2009), gift from K. Grant), and *KrRD* (PCR amplified from *Drosophila* genomic DNA). *hsp83*:*lexA-VP16-mCherry*:*nos* 3'UTR was obtained by inserting *hsp83* (PCR amplified from genomic DNA), *lexA-VP16* (PCR amplified from genomic DNA, Bloomington Stock 29958), *mCherry* (PCR amplified, (Shaner et al. 2004)) and *nos* 3'UTR (PCR amplified from genomic DNA) into *UAS*x3:*gal80* (see Ch. 3). The *lexO*x4 sequence (TCGACTGCTGTATATAAAACCAGTGGTTATATGTACAGTACTGCTGT ATATAAAACCAGTGGTTATATGTACAGTACGTCGA) was inserted into pBlueScript and multimerized as above to obtain *lexO*x8. *lexO*x8:*tetO*x7:*cI-antPHD* was obtained by inserting *lexO*x8 (digested out of *lexO*x8 in pBlueScript), *tetO*x7 (digested out of *tetO*x7 in pBlueScript), *cI* (PCR amplified from pTSMA plasmid, gift from L.C. You (Kobayashi et al. 2004)), and *antPHD* (PCR amplified from genomic DNA (Regulski et al. 1985)) into *UAS*x3:gal80 (see Ch. 3). *OR1-3* was inserted into pBlueScript by Gibson cloning following PCR with primers

AATATCTAACACCGTGCGTGTTGACTATTTTACCTCTGGCGGTGATAgtcgacggta tcgataagcttgatatcgaattcctgcagcc and AAATAGTCAACACGCACGGTGTTAGATATTT ATCCCTTGCGGTGATActcgaggggggcccggtacccaatt. *UAS*x7:*OR1-3*:*tetR-KrRD* was obtained by inserting *OR1-3* (digested out of *OR1-3* in pBlueScript), *tetR*, and *KrRD* (PCR amplified from genomic DNA (Licht et al. 1994)) into *UAS*x7:*gal80* (see Ch. 3).

5.5.2 - Fly Stocks

The w-;Gal4-GCN4:Bcd 3'UTR;Gal4-GCN4:Bcd 3'UTR flies used were a gift from Dr. Dostatni (Janody et al. 2001). The *UAS*x7:*OR13*:*tetR-KrRD, hsp83*:*lexA-VP16*:*nos* 3'UTR, and *UAS*x7:*lacO*x8:*tetR-KrRD* flies were created by injection and incorporation of the plasmid construct into the 68A4 landing site (by Molecular System Injections into yw; attP2 flies). The *lexO*x8:*tetO*x7:*cI-antPHD* and *UAS*x3:*tetO*x4:*lacI-KrRD* flies were created by injection and incorporation of the plasmid construct into the 25C7 landing site (by Molecular System Injection into yw;attP40 flies).

5.5.3 - Embryo Staining and Image Collection and Analysis

Embryos were fixed 2-4 hrs following egg lay per standard protocols. Subsequently, flouurescent *in situ* hybridization was conducted per published protocols (Kosman et al. 2004) with Proteinase K treatment omitted). An RNA probe for *tetR* (biotin conjugated) was used. Primary antibodies to biotin (goat anti-biotin 1:5000; gift from Immunoreagents). Secondary antibodies were Alexa Flour 546 donkey anti-goat (ThermoFisher Scientific). All images were taken at the mid-saggittal plane using a Zeiss Confocal microscope.

References

Brown, M., Figge, J., Hansen, U., Wright, C., Jeang, K., Khoury, G., Livingston, D. & Roberts, T. 1987, "Iac Repressor can Regulate Expression from a Hybrid-Sv40 Early Promoter Containing a Lac Operator in Animal-Cells", *Cell*, vol. 49, no. 5, pp. 603-612.

Campeau, E., Ruhl, V.E., Rodier, F., Smith, C.L., Rahmberg, B.L., Fuss, J.O., Campisi, J., Yaswen, P., Cooper, P.K. & Kaufmann, P.D. 2009, "A Versatile Viral System for Expression and Depletion of Proteins in Mammalian Cells", *Plos One*, vol. 4, no. 8, pp. e6529.

Driever, W. & Nusslein-Volhard, C. 1988, "The Bicoid protein determines position in theDrosophila embryo in a concentration-dependent manner.", *Cell*, vol. 54, no. 1, pp. 95-104.

Elliott, D.A. & Brand, A.H. 2008, "The GAL4 system : a versatile system for the expression of genes.", *Methods Mol Biol*, vol. 420, pp. 79-95.

Gavis, E.R. & Lehmann, R. 1994, "Translational regulation of nanos by RNA localization.", *Nature*, vol. 369, no. 6478, pp. 315-318.

Gavis, E.R. & Lehmann, R. 1992, "Localization of Nanos Rna Controls Embryonic Polarity", *Cell*, vol. 71, no. 2, pp. 301-313.

Giniger, E. & Ptashne, M. 1988, "Cooperative Dna-Binding of the Yeast Transcriptional Activator Gal4", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 2, pp. 382-386.

Houchmandzadeh, B., Wieschaus, E. & Leibler, S. 2005, "Precise domain specification in the developing Drosophila embryo.", *Phys Rev E Stat Nonlin Soft Matter Phys,* vol. 72, no. 6 Pt 1, pp. 061920.

Huang, A.M., Rusch, J. & Levine, M. 1997, "An anteroposterior dorsal gradient in the Drosophila embryo", *Genes & development,* vol. 11, no. 15, pp. 1963-1973.

Jaeger, J. 2011, "The gap gene network", *Cellular and Molecular Life Sciences,* vol. 68, no. 2, pp. 243-274.

Jagla, T., Bidet, Y., Da Ponte, J., Dastugue, B. & Jagla, K. 2002, "Cross-repressive interactions of identity genes are essential for proper specification of cardiac and muscular fates in Drosophila", *Development,* vol. 129, no. 4, pp. 1037-1047.

Janody, F., Sturny, R., Schaeffer, V., Azou, Y. & Dostatni, N. 2001, "Two distinct domains of Bicoid mediate its transcriptional downregulation by the Torso pathway", *Development,* vol. 128, no. 12, pp. 2281-2290.

Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T., Cantor, C. & Collins, J. 2004, "Programmable cells: Interfacing natural and engineered gene networks", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 101, no. 22, pp. 8414-8419.

Kosman, D., Mizutani, C., Lemons, D., Cox, W., McGinnis, W. & Bier, E. 2004, "Multiplex detection of RNA expression in Drosophila embryos", *Science,* vol. 305, no. 5685, pp. 846-846.

Lai, S. & Lee, T. 2006, "Genetic mosaic with dual binary transcriptional systems in Drosophila.", *Nat Neurosci,* vol. 9, no. 5, pp. 703-709.

Licht, J., Hannarose, W., Reddy, J., English, M., Ro, M., Grossel, M., Shaknovich, R. & Hansen, U. 1994, "Mapping and Mutagenesis of the Amino-Terminal Transcriptional Repression Domain of the Drosophila Kruppel Protein", *Molecular and cellular biology,* vol. 14, no. 6, pp. 4057-4066.

Lohr, D., Venkov, P. & Zlatanova, J. 1995, "Transcriptional Regulation in the Yeast Gal Gene Family - a Complex Genetic Network Rid B-3273-2009", *Faseb Journal,* vol. 9, no. 9, pp. 777-787.

Macdonald, P.M. & Struhl, G. 1988, "cis-acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos.", *Nature,* vol. 336, no. 6199, pp. 595-598.

McHale, P., Rappel, W. & Levine, H. 2006, "Embryonic pattern scaling achieved by oppositely directed morphogen gradients", *Physical Biology,* vol. 3, no. 2, pp. 107-120.

Meyer, B., Maurer, R. & Ptashne, M. 1980, "Gene-Regulation at the Right Operator (Or) of Bacteriophage-Lambda .2. Or1, Or2, and Or3 - their Roles in Mediating the Effects of Repressor and Cro", *Journal of Molecular Biology,* vol. 139, no. 2, pp. 163-194.

Morishita, Y. & Iwasa, Y. 2009, "Accuracy of positional information provided by multiple morphogen gradients with correlated noise", *Physical Review E,* vol. 79, no. 6, pp. 061905.

Mortlock, A., Low, W. & Crisanti, A. 2003, "Suppression of gene expression by a cell-permeable Tet repressor.", *Nucleic Acids Res,* vol. 31, no. 23, pp. e152.

Perkins, T.J., Jaeger, J., Reinitz, J. & Glass, L. 2006, "Reverse engineering the gap gene network of Drosophila melanogaster", *Plos Computational Biology*, vol. 2, no. 5, pp. 417-428.

Regulski, M., Harding, K., Kostriken, R., Karch, F., Levine, M. & McGinnis, W. 1985, "Homeo Box Genes of the Antennapedia and Bithorax Complexes of Drosophila", *Cell*, vol. 43, no. 1, pp. 71-80.

Reversade, B. & De Robertis, E. 2005, "Regulation of ADMP and BMP2/4/7 at opposite embryonic poles generates a self-regulating morphogenetic field", *Cell*, vol. 123, no. 6, pp. 1147-1160.

Shaner, N., Campbell, R., Steinbach, P., Giepmans, B., Palmer, A. & Tsien, R. 2004, "Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein", *Nature biotechnology*, vol. 22, no. 12, pp. 1567-1572.

Sokolowski, T.R., Erdmann, T. & ten Wolde Pieter Rein 2012, "Mutual repression enhances the steepness and precision of gene expression boundaries.", *PLoS Comput Biol*, vol. 8, no. 8, pp. e1002654.

Spirov, A., Fahmy, K., Schneider, M., Frei, E., Noll, M. & Baumgartner, S. 2009, "Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein gradient.", *Development*, vol. 136, no. 4, pp. 605-614.

Weil, T.T., Forrest, K.M. & Gavis, E.R. 2006, "Localization of bicoid mRNA in late oocytes is maintained by continual active transport", *Developmental Cell*, vol. 11, no. 2, pp. 251-262.

Wolpert, L. 1969, "Positional information and the spatial pattern of cellular differentiation.", *J Theor Biol,* vol. 25, no. 1, pp. 1-47.

Xu, P., Houssin, N., Ferri-Lagneau, K.F., Thisse, B. & Thisse, C. 2014, "Construction of a Vertebrate Embryo from Two Opposing Morphogen Gradients", *Science,* vol. 344, no. 6179, pp. 87-89.

Yagi, R., Mayer, F. & Basler, K. 2010, "Refined LexA transactivators and their use in combination with the Drosophila Gal4 system", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 107, no. 37, pp. 16166-16171.

CHAPTER 6

New Tools for Controlling Gene Expression in *Drosophila melanogaster* embryos using Hammerhead Ribozymes and Short Upstream Open Reading Frames

Ashley A. Jermusyk, Thomas Jacobsen, Paetra N. Muller, and Gregory T. Reeves

6.1 – Background

It is desirable to alter gene expression of either endogenous genes or genes within synthetic network systems. Several different methods have been used or are currently being explored for tuning gene expression. Among these, the oldest method for tuning endogenous gene expression is to insert mutations within the extended gene regions (Reviewed in Bokel 2008, Venken, Simpson & Bellen 2011). This can be done using mutagenic agents (such as UV radiation, x-rays, or ethyl methanesulfonate) or by inserting transposable elements (Reviewed in Bokel 2008, Venken, Simpson & Bellen 2011). However, these mutations are random and therefore it is difficult to produce a specific change in expression levels.

Newer methods have been developed for tuning gene expression at the RNA or protein level in endogenous or synthetic systems. At the RNA level, RNAi knock-down, self-cleaving hammerhead ribozymes, and synthetic promoter systems to control transcription can be used (Deans, Cantor & Collins 2007, Horn et al. 2011, Guzman et al. 1995, Auslander, Ketzer & Hartig 2010, Yen et al. 2004). RNAi has been validated for both endogenous and synthetic systems; however, it requires delivery of RNA molecules into the cell or secondary expression systems (Deans, Cantor & Collins 2007, Horn et al. 2011). Furthermore, it is difficult to use RNAi with the degree of control necessary to tune gene expression to intermediate levels (Deans, Cantor & Collins 2007, Horn et al. 2011). Where RNAi methods lack control, the special promoter systems are able to provide this level of tunable expression. However, these systems rely on secondary molecules which are difficult to deliver to multi-cellular organisms, such as *Drosophila* (Guzman et al. 1995). This study explores self-cleaving hammerhead ribozymes for

control of RNA levels because unlike RNAi, they can be tuned by adjusting surrounding sequences to effect folding, and unlike synthetic promoter systems, no secondary molecules are needed (Auslander, Ketzer & Hartig 2010, Yen et al. 2004). When using hammerhead ribozymes to control gene expression, the hammerhead ribozyme sequence is placed on the mRNA transcript with the desired gene, Fig. 6.1A (Auslander, Ketzer & Hartig 2010, Yen et al. 2004). A decrease in mRNA levels occurs via mRNA degradation through self-cleavage of the properly folded hammerhead ribozyme (Fig. 6.1A) (Auslander, Ketzer & Hartig 2010, Yen et al. 2004). This results in decreased amount of both mRNA and protein. The amount of mRNA degradation can be tuned by adjusting the ability of the ribozyme to be properly folded, which is done by changing the sequence immediately upstream of the ribozyme (Fig. 6.1C) (Auslander, Ketzer & Hartig 2010, Yen et al. 2004). In addition, the ribozyme can be inserted either 5' or 3' to the open reading frame for the gene of interest, thus cleaving the 5- cap or 3'-poly(A) tail (Auslander, Ketzer & Hartig 2010, Yen et al. 2004). While both locations should lead to mRNA degradation, previous studies have found more pronounced effects when inserted 5' (Auslander, Ketzer & Hartig 2010, Yen et al. 2004). This could be a result of not only a loss of mRNA stability, but also a loss in the ability for the ribosome to bind and translate the mRNA. This system has been previously validated in mammalian cells to tune gene expression levels (Auslaender et al. 2014, Wieland, Auslaender & Fussenegger 2012).

Various methods also exist for tuning gene expression at the protein level, including: protein tags for targeted degradation, short upstream open reading frames to prevent translation, and synthetic ribosome binding sites (McGinness, Baker & Sauer

2006, Ferreira, Overton & Wang 2013, Salis, Mirsky & Voigt 2009). Targeted protein degradation via tags has been tested in bacterial systems, however this system would require significant modification to function in a new organism, such as *Drosophila* (McGinness, Baker & Sauer 2006). Both synthetic ribosome binding sites and short upstream open reading frames should be easily adaptable for use in *Drosophila* (McGinness, Baker & Sauer 2006, Ferreira, Overton & Wang 2013, Salis, Mirsky & Voigt 2009). We chose to explore short upstream open reading frames for tuning protein levels because more significant changes in expression levels are possible, however ribosome binding sites for these upstream open reading frames and the open reading frame (ORF) for gene of interest can be altered to provide more intermediate levels of control. Upstream ORFs are also known to be present and effect translation in of genes within humans and mice, the parasite *Plasmodium falciparum*, plants (*Arabidopsis*), and *Drosophila* (Calvo, Pagliarini & Mootha 2009, Kumar, Srinivas & Patankar 2015, Matsui et al. 2007, von Arnim, Jia & Vaughn 2014, Medenbach, Seiler & Hentze 2011). Since upstream ORFs are already known to be present in a wide variety of organisms including *Drosophila*, this presents an easily adaptable method for our application.

In our experimental system for evaluating short upstream open reading frames (uORFs), we placed three uORFs encoding two amino acids upstream of the ORF for the gene of interest (Fig. 6.1D). The protein of interest is only produced when ribosomes "leak" past the uORFs (Fig. 6.1E) (Ferreira, Overton & Wang 2013). The amount of protein being produced can be adjusted by altering the ability of the ribosomes to "leak" past the uORFs to the ORF for the gene of interest (Ferreira, Overton & Wang 2013). This can be done by adjusting the number of uORFs: the more uORFs, the fewer

ribosomes that are able to leak past to the ORF of interest (Ferreira, Overton & Wang 2013). The ability of the ribosome to recognize the ORFs can also be adjusted by altering the sequence surrounding the start codon, known as the Kozak sequence (Ferreira, Overton & Wang 2013).

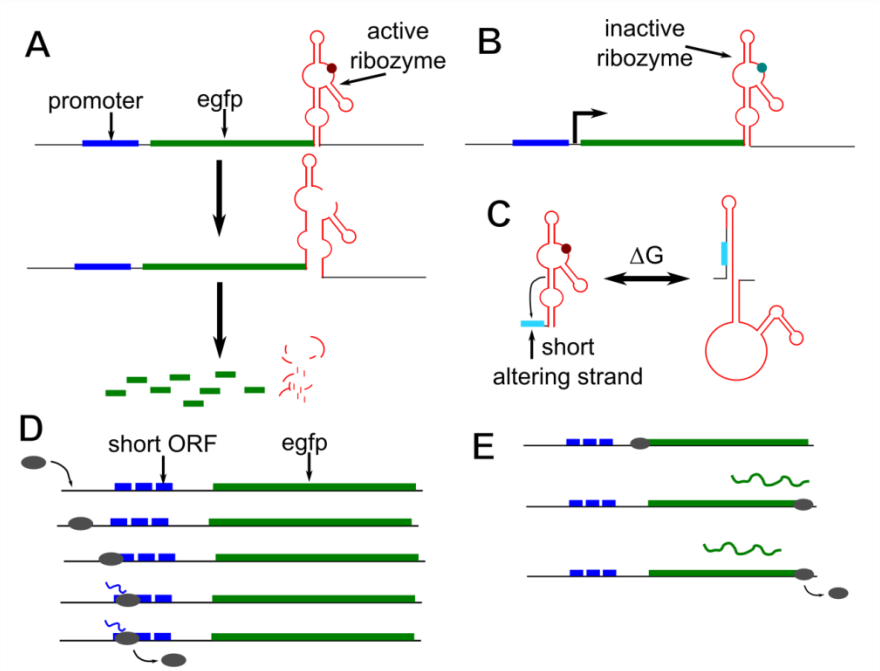

Figure 6.1: Set-up of ribozyme and upstream open reading frames for controlling gene expression. (A) The hammerhead ribozyme is part of the mRNA containing the gene of interest. The hammerhead ribozyme undergoes self-cleavage and the mRNA is degraded. (B) A single base pair mutation (denoted by the now teal circle), inactivates the ribozyme so that this self-cleavage does not occur, allowing for translation of the gene of interest. (C) A sequence can be inserted immediately upstream of the ribozyme which is complementary to second stem of the ribozyme, disrupting the formation of the hammerhead structure. (D) Short open reading frames inserted upstream of the open reading frame for the gene of interest are recognized and translated by the ribosome. Some ribosomes will "leak" past the initial short upstream open reading frame (uORF) to translate a second or third short uORF. (E) A small number of ribosomes will "leak" past these uORFs and translate the gene of interest.

The utility of ribozymes and uORFs to control gene expression in *Drosophila* embryos was evaluated in this study. We first tested the ability of the self-cleaving hammerhead ribozyme and uORFs to decrease gene expression of *egfp* (*enhanced green*

flourescent protein) in *Drosophila* embryos using a ubiquitous promoter. These results suggested the ribozyme is functional within the embryo, however due to low expression of *egfp,* further testing was performed with different promoter systems to create a better system for validation of these methods. It is desirable to not only knockdown expression using the ribozyme, but to tune gene expression to intermediate levels, which can be done by inserting short competing sequences immediately upstream of the ribozyme. The ability of these sequences to tune gene expression was first explored in mammalian cells. Based on these results, a selection of these sequences can then be evaluated in *Drosophila* embryos using the appropriate promoter system to demonstrate their functionality and range of control in tuning gene expression within the embryo. Testing of the uORFs to control protein levels was conducted in *Drosophila* embryos using a ubiquitous promoter; however embryos expressing the uORF construct showed higher expression than control uORF constructs. This may be a result of the low level of expression seen in these embryos, or a problem with our control construct. More work will need to be done to determine optimize this system for proper functionality in *Drosophila* embryos.

6.2 - Results

6.2.1 - Testing of Ribozyme Using Ubiquitous Promoter

The ability of the self-cleaving hammerhead ribozyme to alter gene expression was evaluated by comparing GFP expression when the active or inactive ribozyme is present. This analysis was performed using the promoter *hsp83,* which generates ubiquitous expression throughout the early embryo. We measured expression of

Enhanced Green Flourescent Protein (eGFP) in *Drosophila melanogaster* embryos, via antibody staining and subsequent imaging of the mid-saggittal plane using confocal microscopy (Fig. 6. 2A-B). Antibody staining for eGFP was chosen because of the high level of autofluorescence and the low overall level of eGFP expression. Antibody staining also allows for an increase throughput in our testing and is necessary when using zygotic promoters since sufficient time is not available in that case for the proper folding of eGFP following production (such as those promoters used in section 6.2.2) (Casso, Ramirez-Weber & Kornberg 2000, Dunipace, Ozdemir & Stathopoulos 2011). The same image processing conditions (maximum and minimum corrections) were used for both embryos. We then quantified expression around the periphery of the embryo and took the average (see Materials and Methods). Our analysis found that there was a low level of expression in the inactive ribozyme flies (without normalization average fluorescence is 110.0 ± 22.1 AU in inactive flies compared to 91.6 ± 23.9 AU in background control) (Fig. 6.2C). The expression in flies carrying the active ribozyme was similar to that of the background control (74.3 ± 15.2 AU). We found the flies carrying the inactive ribozyme had higher expression than that with the active ribozyme ($p = 0.008$). However, this difference was very small (1.5 fold reduction by the active ribozyme to 68% of the activity with the inactive ribozyme). This appears higher than is observed in the embryo images (Fig. 6.2A-B) due to the image correction applied (specifically the choice for minimum and maximum intensity thresholds applied). This is most likely due to the high background, therefore we desired to change the promoter used so that a greater dynamic range was available to study the differences in gene expression.

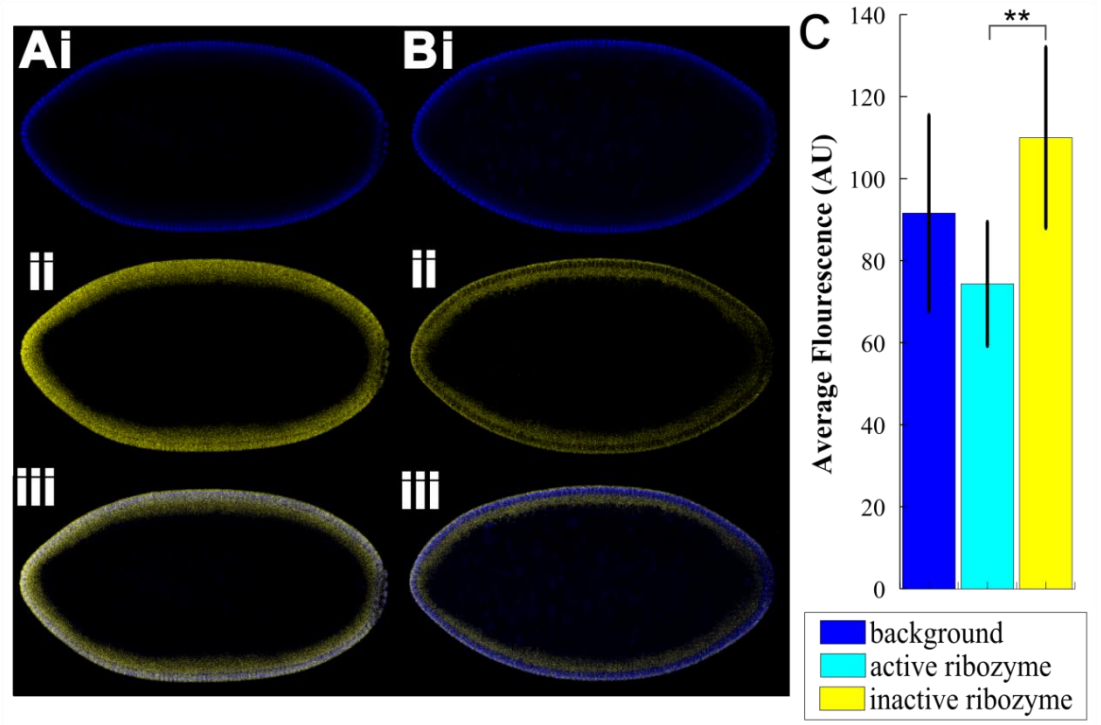

Figure 6.2: Evaluation of Ribozymes Using a Ubiquitous Promoter. (A) Mid-saggittal section of an embryo expressing EGFP using the constitutive *hsp83* promoter with the inactive ribozyme; (i) DAPI staining, showing nuclei around the periphery of the embryo, (ii) antibody staining for GFP, (iii) merged image showing GFP (yellow) and DAPI (blue). (B) Embryo expressing *egfp* linked to the active ribozyme, (i) DAPI, (ii) anti-GFP, (iii) merged image of (i) and (ii) (the same image processing was applied to both images). (C) Comparison of EGFP expression (as measured by antibody staining) in embryos with no construct (background control – blue), construct containing the active ribozyme (cyan), and inactive ribozyme (yellow). Expression in embryos expressing EGFP under control of the constitutive hsp83 promoter is higher ($p = 0.008$) in embryos with the inactive ribozyme ($n = 9$) compare to the active ribozyme ($n = 4$). Both the inactive and active ribozyme are not statistically significantly different from background ($n = 6$).

6.2.2 - Analysis of Different Enhancers for Ribozyme Testing

To generate a higher level of eGFP expression we tested three different enhancers upstream of the eve minimal promoter. In this case, we used enhancers that would generate a spatially-dependent pattern of gene expression; this provides for an easier measurement of eGFP expression since a clear pattern should be visible and proof of eGFP expression. This should eliminate the need to compare expression versus a

separate background or autoflourescence control. The three enhancers used were *gt23*, *KrCD2*, and *kni* proximal (Fig. 6.3A) (Perry, Boettiger & Levine 2011, Ochoa-Espinosa et al. 2005). Our testing failed to show localized protein expression at expected positions. This is possibly due to the high diffusivity of eGFP within the early embryo. Increased diffusion of an intracellular protein in the embryo is possible because the *Drosophila* embryo develops as a syncytium, meaning nuclear divisions occur without corresponding cell divisions for the first fourteen nuclear cycles.

Because no localized expression was observed, we again took the average intensity along the periphery of the embryo following antibody staining. The level of eGFP expression (with average background control subtraction) in flies carrying the inactive ribozyme with all of these enhancers was still very similar to background levels (for *KrCD2* = 3.4 ± 8.2 AU, for *gt23* = 10.9 ± 5.5 AU, for *kni* prox. = 23.3 ± 13.6 AU), or similar to zero. The active ribozyme was found to have lower expression in the *KrCD2* flies ($p < 0.009$), but in the flies containing the other enhancers, this change was not statistically significant (Fig. 6.3B-D). An analysis (Fig. 6.3E) of expression in embryos containing these enhancer constructs and in embryos using the *hsp83* ubiquitous promoter, shows fluorescence levels are even lower in the embryos with the enhancer constructs. This is partially due to the method used to calculate fluorescence in these embryos. Our analysis used an average around the entire embryo, where these enhancer constructs should only generate expression in a localized segment of the embryo, thereby lowering the calculated average value. Of these enhancers, the *kni* prox. enhancer shows the greatest level of expression and is the most promising of those tested.

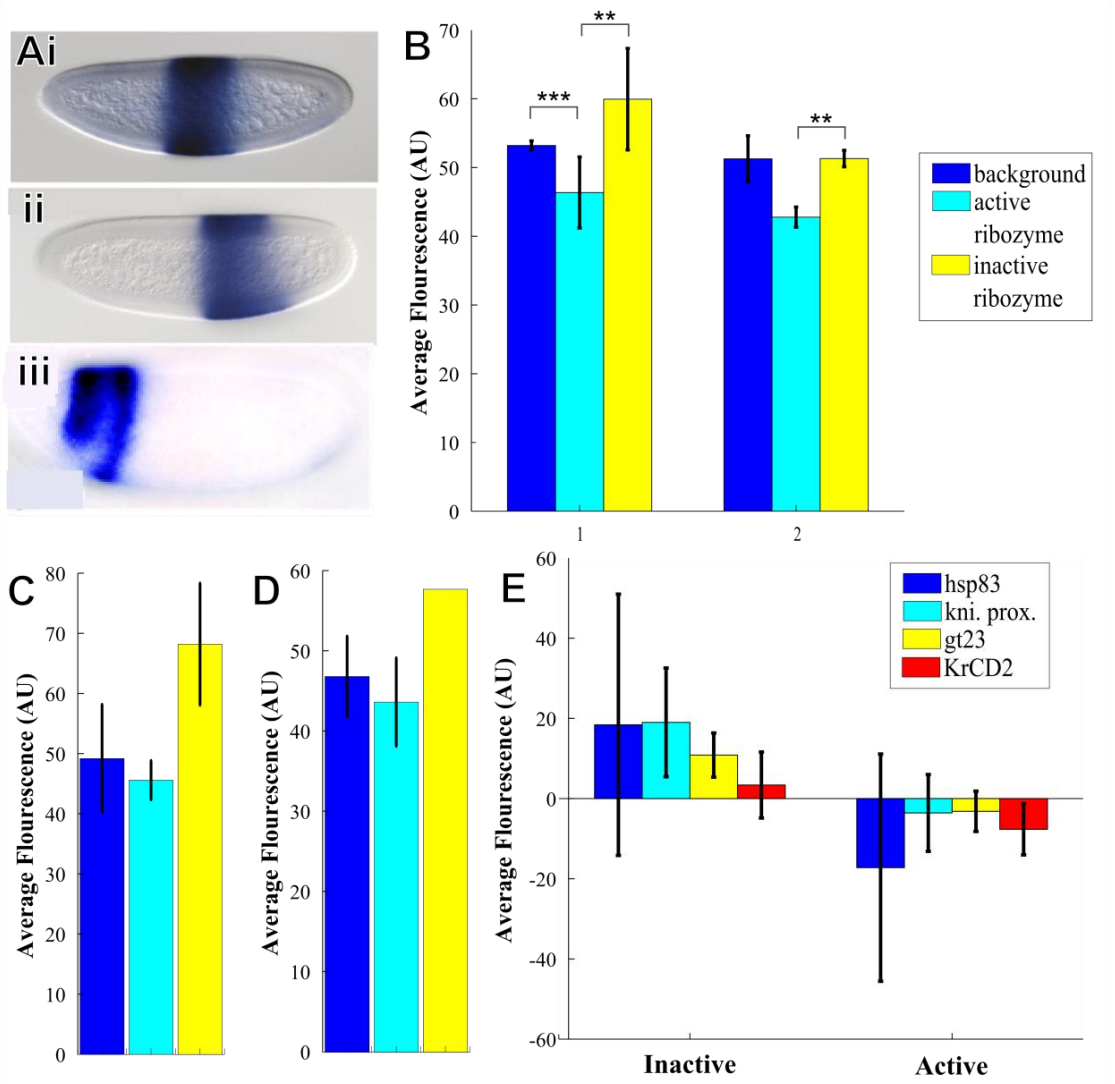

Figure 6.3: Evaluation of Ribozymes Using Enhancer Constructs. (A) *In situ* hybridization for enhancer regions (i) *KrCD2* (Perry, Boettiger & Levine 2011), (ii) *kni* prox. (Perry, Boettiger & Levine 2011), and (iii) *gt23* (Ochoa-Espinosa et al. 2005). (B) Embryos expressing the Kr CD2 construct were tested on two different days for EGFP expression. On day 1, expression in embryos containing the active ribozyme (n = 15) was statistically significantly lower (p = 0.0002) than background (n = 2); embryos expressing the inactive ribozyme (n = 6) have statistically significantly higher EGFP levels than those expressing the active ribozyme (p = 0.004), but not background. On day 2, EGFP levels in embryos containing the inactive ribozyme (n = 2) was statistically significantly higher than those containing the active ribozyme (n = 3) (p = 0.008), however no statistically significant difference was observed compared to background (n = 4). (C) In embryos containing the kni. proximal enhancer construct, eGFP expression is not statistically significantly different between embryos with the inactive ribozyme (n = 2), active ribozyme (n = 8), and background (n = 3). (D) eGFP expression in embryos containing the gt23 enhancer construct (background, n = 6; active, n = 28; inactive, n = 1). (E) (G) Comparison of background subtracted fluorescence in embryos expressing the active and inactive ribozyme constructs with each of the promoter/enhancers tested.

6.2.3 - Effect of Upstream Competing Sequences

As mentioned above, our desire is to tune gene expression to specific levels. Therefore we looked at tuning the activity of the ribozyme by inserting different short sequences directly upstream of the ribozyme. These sequences were designed to be complimentary to, and therefore to hybridize with, the second stem of the ribozyme (Fig. 6.1C). Ten different sequences were designed that have different degrees of complementarity to this stem of the ribozyme. With a given sequence inserted, the RNA can adopt various conformations; conformations without the properly folded hammerhead would not be cleaved (and therefore minimal mRNA degradation would occur). The Gibbs free energy associated with these various active and inactive conformations can be used to predict the fraction of mRNA molecules possessing the active hammerhead conformation (Table 6.1). The ability of these sequences to alter the activity of the ribozyme in tuning gene expression was tested in mammalian cells. Mammalian cells were transfected with constructs containing each of these sequences, sorted by flow cytometry, and the fluorescence was measured. The average of total fluorescence above the gate (set by untransfected cells, see Fig. 6.4A), with fluorescence due to background autoflourescence subtracted (fluorescence in untransfected cells), was used. The intensity (normalized to the inactive ribozyme with no competing sequence), shown in Table 6.1, is greater for the inactive than the active ribozyme for each pair except with the R3 upstream competing sequence inserted (Fig. 6.4D). While it is expected that all cells with the inactive ribozyme to have similar fluorescence, variations were observed; this may be because the competing sequence, being upstream of the *egfp* open reading frame, affects the ability of the ribosome to translate the protein.One way to

compare the ability of these upstream competing sequences to affect eGFP expression is to examine the fold change between the inactive and active ribozyme in each case, Table 6.1 and Fig. 6.4E. These upstream competing sequences were able to tune gene expression between a 28.4 and -1.7 fold decrease in expression levels (calculated as the background subtracted and normalized fluorescence of the active ribozyme over the inactive ribozyme).

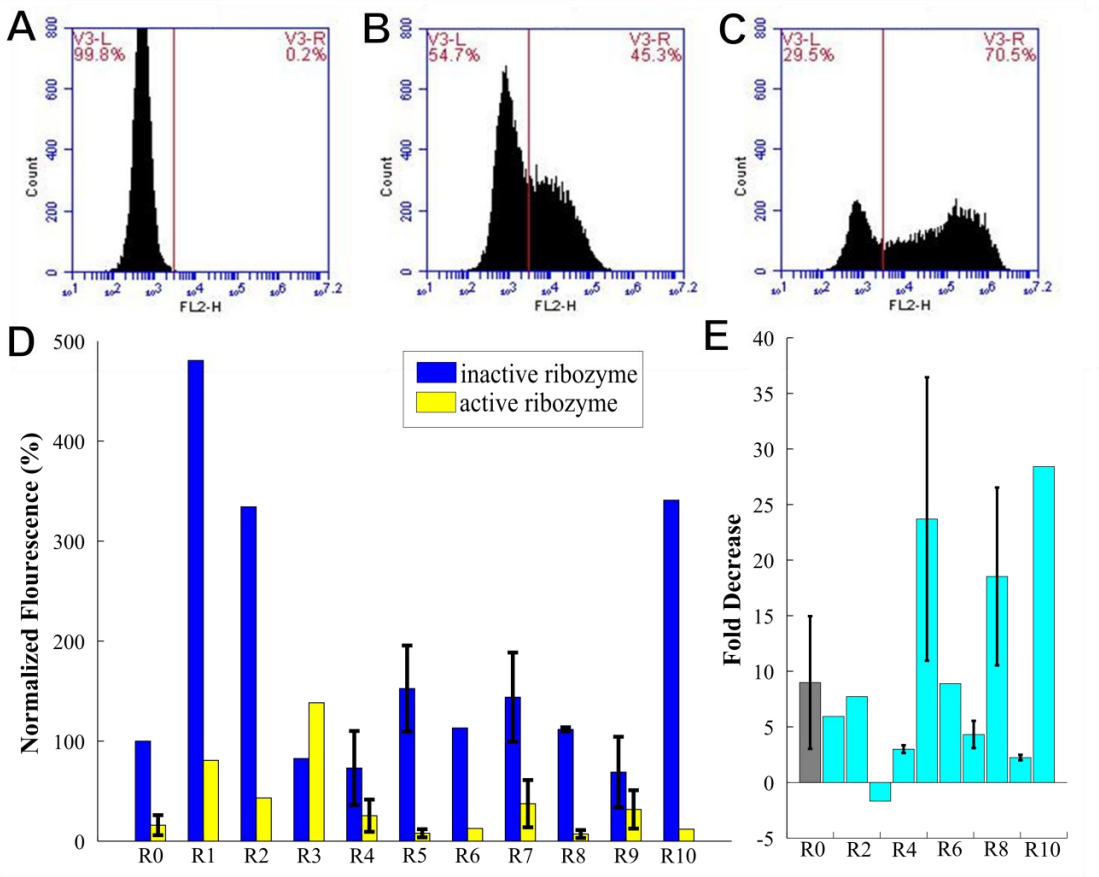

Figure 6.4: Ability of Upstream Competing Sequences to Effect Ribozyme Activity in Mammalian Cells. (A) Histograms showing fluorescence in a population of untransfected mammalian cells. Red line is the gate, set at with 2% of cell population above the gate. Fluorescence in mammalian cells transfected with the active ribozyme construct (B) and inactive ribozyme construct (C). (D) Normalized fluorescence in cells transfected with the active ribozyme (blue) or inactive ribozyme (yellow) constructs containing upstream competing sequences (all values normalized to average fluorescence in cells with the inactive ribozyme with no short upstream competing sequence [R0]). (E) Fold change in fluorescence between cells containing inactive and active ribozyme constructs (calculated as inactive/active fluorescence).

Table 6.1: Changes in eGFP expression in mammalian cells due to insertion of upstream competing sequences

Upstream Sequence	Active Hammerhead, ΔG (kcal/mol)	No Hammerhead, ΔG (kcal/mol)	Fraction Active	Fraction Inactive	Normalized Intensity		Fold Decrease
					Inactive Ribozyme	Active Ribozyme	
None (R0)	-32.0, -31.6	-30.7, -30.6	0.872	0.128	100.0 ± 0.0	16.0 ± 10.0	9.0 ± 6.0
R1	-33.7, -33.7, -33.4	-32.4, -32.3	0.921	0.079	480.6	80.8	5.9
R2	-35.3	-34.0, -33.6	0.844	0.156	334.2	43.3	7.7
R3	-35	-33.7, -33.5, -33.4	0.779	0.221	82.6	138.4	-1.7
R4	-35.5	-34.2, -34.2, -33.9	0.759	0.241	73.1 ± 37.1	25.5 ± 16.1	3.0 ± 0.4
R5	-35.8, -35.7, -35.2	-35.8	0.690	0.310	152.6 ± 43.0	7.9 ± 4.1	23.7 ± 12.7
R6	-33.9, -32.7	-33.4, -33.1, -32.6, -32.5, -32.4	0.526	0.474	113.2	12.7	8.9
R7	-33.6, -32.6	-33.4, -33, -32.5, -32.4, -32.3	0.439	0.561	144.0 ± 44.7	37.5 ± 23.6	4.3 ± 1.2
R8	-33.5	-34.3, -33.4, -33.3	0.160	0.840	111.7 ± 2.1	7.2 ± 4.0	18.5 ± 8.0
R9	-35.0	-36.0, -35.1, -35.0	0.121	0.879	69.0 ± 35.4	31.8 ± 19.2	2.2 ± 0.2
R10	-	-37.6, -37.5, -37.3, -36.3	0.000	1.000	340.9	12.0	28.4

6.2.4 - Testing of Upstream Open Reading Frames Using Ubiquitous Promoter

The ability of short upstream open reading frames to alter protein expression was evaluated using the ubiquitous *hsp83* promoter. For comparison, a control plasmid was used where the ATGs were replaced with TTG. The average fluorescence intensity was found around the embryo following antibody staining (as above) to measure the relative amounts of eGFP in embryos expressing the uORF and control uORF constructs. The background subtracted average fluorescence level in flies expressing the control uORF construct was very low (0.9 ± 11.5 AU), Fig. 6.5C. However, expression was higher in the flies expressing the uORF construct (with background subtraction, 10.6 ± 11.5 AU), this would mean there is more eGFP when the uORFs are inserted, which is the opposite of what is expected. This increase in observed eGFP levels in the uORF flies compared to the control uORF flies may be due to the high variability compared to the overall level of expression.

This initial testing shows there are insufficient levels of expression of eGFP using the *hsp83* promoter to detect differences between the uORF and uORF control expressing flies. This is consistent with the small changes observed using this promoter seen while testing the ribozyme system and suggests this system will need to validated using a promoter system that generates sufficient levels of eGFP expression. If this observed increase in expression with the uORF construct as compared to the control uORF construct is not simply due to noise within the system at our low levels of expression, then this means the control uORF construct is causing a decrease in expression, which could be due to a number of factors. One possibility is the control uORFs used are somehow affecting the ability of the ribosome to scan the RNA for the start codon;

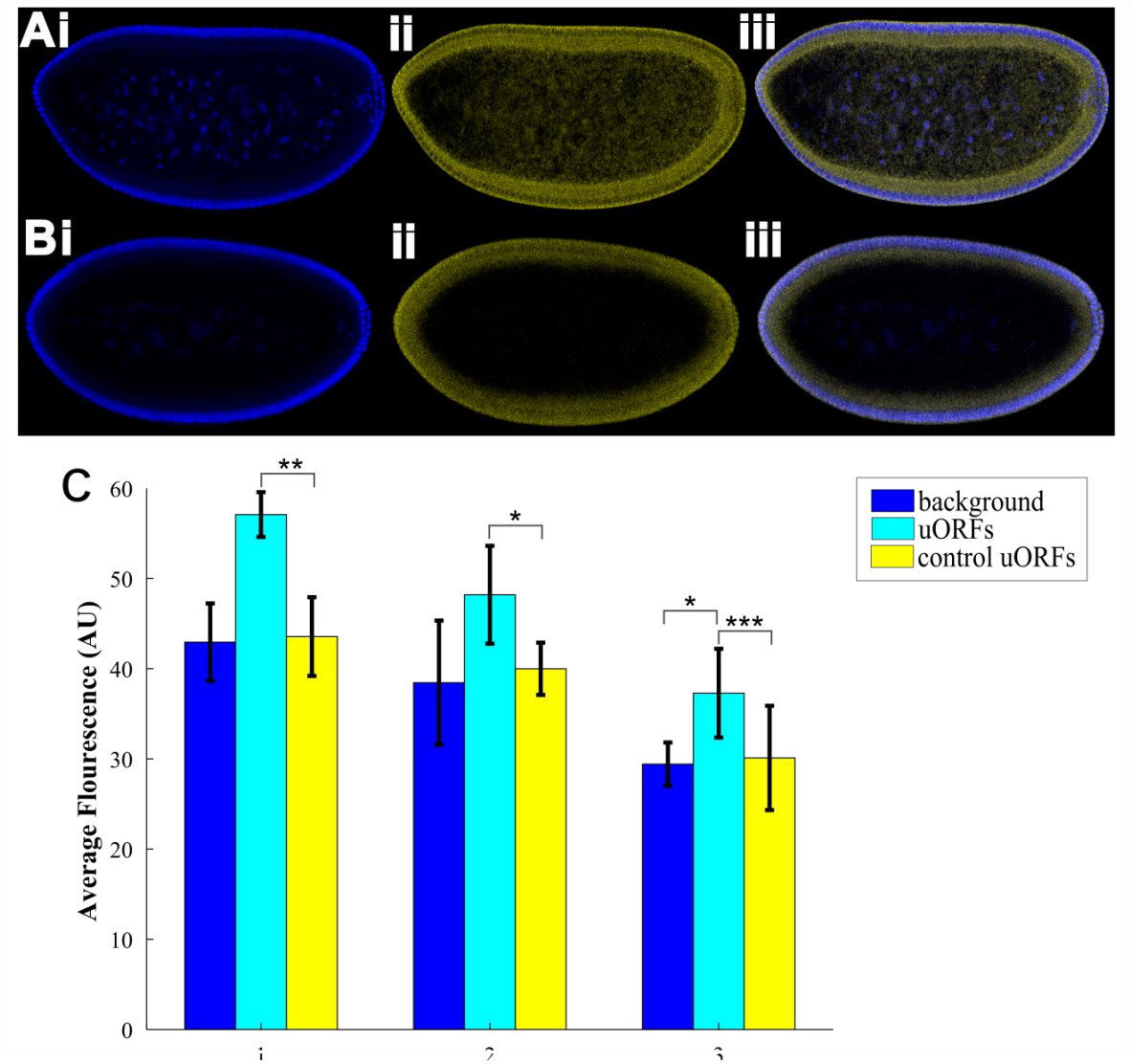

Figure 6.5: Evaluation of short upstream open reading frames for controlling gene expression. (A) Embryo expressing EGFP using the constitutive hsp83 promoter with the control uORF construct; (i) DAPI nuclear staining, (ii) antibody staining for GFP, (iii) merged image showing GFP (yellow) and DAPI (blue). (B) EGFP expression in embryo with the uORF construct; (i) DAPI, (ii) anti-GFP, (iii) merged anti-GFP and DAPI. (C) Comparison of EGFP expression in embryos containing uORF construct (cyan), control uORF construct (yellow), and background control (blue) (three replicates on separate days). On day 1, expression in embryos expressing the uORF construct (n = 3) is statistically significantly higher (p = 0.001) than those expressing the control uORF construct (n = 5), however no difference is found between either construct and background levels (n = 7). On day 2, uORF construct expressing embryo (n = 5) was again found to be statistically significantly higher (p = 0.026) than control uORF construct expressing embryo (n = 4), no difference was found compared to background (n = 6). On day 3, embryos expressing the uORF construct (n = 13) had statistically significantly higher expression than control uORF construct (n =23, p = 0.0005) and background control (n = 15, p = 0.04).

however, this is unlikely since the same changes are being made as were used in previous testing of this system (Ferreira, Overton & Wang 2013). Much work will need to be done to explore the reason for these unexpected results.

6.3 - Future Work

6.3.1 - Future Hammerhead Ribozyme Work

We were able to demonstrate the ability of a self-cleaving hammerhead ribozyme to alter gene expression of eGFP in the early *Drosophila* embryo. However, the levels of eGFP without the active ribozyme were relatively low. This resulted in a small overall difference between the active and inactive cases. A different promoter and enhancer combination could be used to produce greater levels of expression. In addition, changing the gene being expressed could allow us to measure higher levels of expression. Since the *Drosophila* embryo develops as a syncytium (at our time window, the nuclei are present around the periphery of the embryo, which is still one cell), species are more freely able to diffusive through the embryo than would occur in a normal multi-cellular system. A decrease in diffusivity could be achieved by using a protein that is anchored to the cell membrane. One such protein is mCD8-GFP which has been previously shown in *Drosophila* to label the cell surface (due to the mCD8), but still retain GFP expression (Lee, Luo 1999). This would eliminate problems due to the diffusive nature of eGFP following translation. In addition it would be desirable to measure the RNA levels of *egfp*, which would allow us to measure the amount of RNA that is present (and not degraded following self-cleavage). This RNA level should correspond to the protein levels of eGFP.

The ability of the various upstream competing sequences to tune the hammerhead ribozyme-mediated knock-down of gene expression was determined in mammalian cells. We would like to use these results to predict the tuning of the functional activity of the ribozyme in *Drosophila* embryos. This tuning will generate expression at intermediate levels between the maximum expression (inactive ribozyme) and minimum (active ribozyme with no altering strand). To properly tune expression along this dynamic range, we should place the original (R0), R7, R8, R9, and R10 in flies for testing. This should present fold changes from the initial inactive ribozyme (9.0 fold change), to 4.3, 18.5, 2.2, and 28.4 respectively.

6.3.2 - Future Short Upstream ORFs Work

Upstream ORFs are known to be present and affect translation rates in *Drosophila*, mammals, plants, and parasites (Calvo, Pagliarini & Mootha 2009, Kumar, Srinivas & Patankar 2015, Matsui et al. 2007, von Arnim, Jia & Vaughn 2014, Medenbach, Seiler & Hentze 2011). Some of this work suggests transcription factor binding between these upstream ORFs and the major ORF may mediate the ability of these upstream ORFs to control expression (Medenbach, Seiler & Hentze 2011). However, we were unable to verify the insertion of short upstream ORFs directly upstream of the major ORF on their own are able to control gene expression in the *Drosophila* embryo. If we are able to decrease gene expression with the addition of these short uORFs, it is then desirable to adjust the gene expression to intermediate levels. This can be done in a few different manners. The translation initiation site (TIS) can be altered such that the ribosome differentially recognizes the start codon. The optimal TIS

is GCCACCAUGGG (Kozak 1987). Small changes in this sequence effect the translation efficiency. For example, the -3A has an enhanced efficiency by 58% relative to -3U (Kozak 1987). Alterations can be made in the +4 and +5 position to effect translation. Since these changes would be in the tripeptides that are be translated by the short uORFs, it is not problematic to alter this sequence (Noderer et al. 2014). Changes in the number of short uORFs and in the TIS for each uORF should alter the amount of eGFP (or other gene of interest) that is translated. These alterations to the system can be made once a good promoter/enhancer system is found to work for testing; this can be accomplished with the hammerhead ribozyme and then verified with multiple short uORFs, as attempted in the preliminary testing above.

6.4 - Conclusions

In order to design a system for tuning gene expression in the *Drosophila* embryo, we explored self-cleaving hammerhead ribozymes and short upstream open reading frames. These techniques for tuning gene expression can be applied to both synthetic gene networks and endogenous studies. Specifically, we seek to construct synthetic gene networks within the *Drosophila melanogaster* embryo. In order to properly express these networks, it is necessary for genes within the network to be expressed in similar levels and therefore it is necessary to finely tune expression of individual genes within the system. This is difficult to do with such coarse methods currently available, such as altering copy number of the genes expressed or changing promoters. Therefore, the ability of ribozymes or short uORFs to predictively and reproducibly tune gene expression will greatly aid in these studies. Due to low levels of expression of our *egfp*

constructs we were unable to validate the use of these systems in *Drosophila* embryos. However, in parallel the ability to tune ribozyme activity was tested in mammalian cells using upstream competing sequences. It was found these sequences are able to increase the range of utility for ribozymes and tune their activity to intermediate levels as desired. Work is still needed on these systems to demonstrate their functionality in *Drosophila* embryos.

6.5 - Materials and Methods

6.5.1 - Plasmids

Plasmids for expression in *Drosophila* were constructed from the pUAST parent plasmid (gift from J. Mahaffey). The *hsp83*:*active*:*egfp, hsp83*:*inactive*:*egfp, hsp83*:*ORF*:*egfp,* and *hsp83*:*ORFC*:*egfp* plasmids were created by first inserting *egfp* and *hsp83*. Followed by insertion of the applicable ribozyme or *ORF* (PCR amplified using the corresponding gBlock, created by Integrated DNA Technologies, Inc.). The sequences of these gBlocks are as follows: for the *active ribozyme* tagctagctgactacgtacgagcagctacgactaagcttATTCGC CCGCACAGGTTGCGCACTTTTCGACCGTATCACAACACTGATCTACCCTAGTA TTCACAGGAAGTTGCATCCCTGCATCCAGAAGCCTCTAGAAGTTTCTAGAGA CTTCCAGTTCGGGTCGGGTTTTTCTATAAAAGCAGACGCGCGGCGTTTGCCGG TTcgagtcTTGAAAAAAATTTCGTACGGTGTGCGTCGTAACAACAAGCAgaattcGC ActcgagCTGAGGTGCAGGTACATCCAGCTGACGAGTCCCAAATAGGACGAAAC GCGCTTCGGTGCGTCCTGGATTCCACTGCTATCCACgcggccgctactatgcatgcgatcggc tacgatc, for the *inactive ribozyme* tagctagctgactacgtacgagcagctacgactaagcttATTCGCCCG CACAGGTTGCGCACTTTTCGACCGTATCACAACACTGATCTACCCTAGTATTC

ACAGGAAGTTGCATCCCTGCATCCAGAAGCCTCTAGAAGTTTCTAGAGACTT
CCAGTTCGGGTCGGGTTTTTCTATAAAAGCAGACGCGCGGCGTTTGCCGGTTc
gagtcTTGAAAAAAATTTCGTACGGTGTGCGTCGTAACAACAAGCAgaattcGCActc
gagCTGAGGTGCAGGTACATCCAGCTGACGAGTCCCAAATAGGACGAGACGC
GCTTCGGTGCGTCCTGGATTCCACTGCTATCCACgcggccgctactatgcatgcgatcggctac
gatc, for the *ORF* tagctagctgactacgtacgagcagctacgactaagcttATTCGCCCGCACAGGTTG
CGCACTTTTCGACCGTATCACAACACTGATCTACCCTAGTATTCACAGGAAGT
TGCATCCCTGCATCCAGAAGCCTCTAGAAGTTTCTAGAGACTTCCAGTTCGGG
TCGGGTTTTTCTATAAAAGCAGACGCGCGGCGTTTGCCGGTTcgagtcTTGAAAA
AAATTTCGTACGGTGTGCGTCGTAACAACAAGCAgaattcGCActcgagACCATGGG
TTGATAAACCATGGGTTGATAAACCATGGGTTGATTACCgcggccgctactatgcatgcga
tcggctacgatc, and for the *ORFC* tagctagctgactacgtacgagcagctacgactaagcttATTCGCCC
GCACAGGTTGCGCACTTTTCGACCGTATCACAACACTGATCTACCCTAGTATT
CACAGGAAGTTGCATCCCTGCATCCAGAAGCCTCTAGAAGTTTCTAGAGACT
TCCAGTTCGGGTCGGGTTTTTCTATAAAAGCAGACGCGCGGCGTTTGCCGGTT
cgagtcTTGAAAAAAATTTCGTACGGTGTGCGTCGTAACAACAAGCAgaattcGCAct
cgagACCTTGGGTAGATAAACCTTGGGTAGATAAACCTTGGGTAGATTACCgcgg
ccgctactatgcatgcgatcggctacgatc. The *KrCD2*:*egfp*:*active*, *KrCD2*:*egfp*:*inactive*, *kni.prox*:*egfp*:*active*, *kni.prox*:*egfp*:*inactive*, *gt23*:*egfp*:*active*, and *gt23*:*egfp*:*inactive* plasmids were created by first digesting the *hsp83*:*active*:*egfp* plasmid with *Eco*RI and *Not*I to insert the *eve basal* promoter (PCR amplified from genomic DNA using AGATACATgaattcGAGCGCAGCGGTATAAAAGGGC and ATTCGAgcggccgcGGTCCACGGGACTGGCGTCGTGA). This *evep*:*egfp* plasmid was

then digested with *Hin*dIII and *Eco*RI and the desired enhancer region (PCR amplified from genomic DNA) was added. The *KrCD2* enhancer was amplified using primers AGATACATaagcttGTAAGTTCCCATATTTCGGACCTTATC and AGATACATgaattcTGGGTACTTCGCTGAGTTGAGTGAGTTG (Perry, Boettiger & Levine 2011). The *kni* extended proximal was amplified using AGATACATaagcttGGTGGTGCGGTTCTTCTTGTCGATGATG and AGATACATgaattcGAGTGAGTGAGAAATCCAGCCGCCCTTAG (Perry, Boettiger & Levine 2011). The *gt23* enhancer was PCR amplified using AGATCATaagcttGGGAATTCGGCGACTTGGATCGTGAG and ATGACACAgaattcAAAACTGCAGCTGCCCTGCCCTGCTCTG (Ochoa-Espinosa et al. 2005).

The plasmids used for expression in HEK293T cells were constructed from the pcDNA3.1+ vector (Invitrogen, Cat#: V790-20). The *active*:*egfp* and *inactive*:*egfp* plasmids were constructed by PCR amplification of the previously mentioned *hsp83*:*active*:*egfp* and *hsp83*:*inactive*:*egfp*, respectively (using ggctagcgtttaaacttaaagcttACGGTGTGCGTCGTAACA and ggtttaaacgggcccttctagaTTACTTGTACAGCTCGTCCATG), followed by restriction cloning (*Hin*dIII and *Xba*I). The upstream altering strands were constructed by annealing pairs of short oligos (see Table C.1) and subsequent phosphorylation. These sequences were cloned into the ribozyme plasmids using *Eco*RI and *Xho*I.

6.5.2 - Embryo Staining, Imaging, and Analysis

Embryos 2-4 hrs after egg lay were formaldehyde fixed per standard protocols. Fluorescent *in situ* hybridization was performed per published protocols ((Kosman et al. 2004) with Proteinase K treatment omitted) using an RNA probe for *egfp* (fluorescein conjugated). Primary antibodies to fluorescein (goat anti-fluorescein, 1:500, Rockland), green fluorescent protein (chicken anti-GFP, 1:500; Novus Biologicals), and red fluorescent protein (rabbit anti-RFP, 1:200; Abcam). Secondary antibodies used were Alexa Flour 488 anti-rabbit (ThermoFisher Scientific), Alexa Flour 546 anti-goat (ThermoFisher Scientific), Alexa Flour 647 anti-chicken (Millipore). Embryo images were taken at the mid-saggittal plane using a Zeiss Confocal microscope. The images were then analyzed using the method describe in Jermusyk gal80 paper.

6.5.3 - Fly Stocks

The *hsp83*:*active*:*egfp*, *hsp83*:*inactive*:*egfp*, *hsp83*:*ORF*:*egfp*, and *hsp83*:*ORFC*:*egfp* plasmids were injected into yw;attP2 (68A4) flies by Genetic Services, Inc. The *KrCD2*:*egfp*:*active*, *KrCD2*:*egfp*:*inactive*, *kni*.prox:*egfp*:*active*, *kni*.prox:*egfp*:*inactive*, *gt23*:*egfp*:*active*, and *gt23*:*egfp*:*inactive* plasmids were injected into yw;attP2 (68A4) flies by GenetiVision.

6.54 - Mammalian Cell Culture Testing

The ribozyme plasmids were transfected in HEK293T cells (gift from B. Rao) using FuGene HD (Promega, Cat#: E2311) with a 4:1 reagent:DNA ratio. After a 48-hour

incubation period, the cells were trypsinized and analyzed using flow cytometry (BD Accuri C6 Flow Cytometer).

References

Auslaender, S., Stuecheli, P., Rehm, C., Auslaender, D., Hartig, J.S. & Fussenegger, M. 2014, "A general design strategy for protein-responsive riboswitches in mammalian cells", *Nature Methods,* vol. 11, no. 11, pp. 1154-1160.

Auslander, S., Ketzer, P. & Hartig, J.S. 2010, "A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression.", *Mol Biosyst,* vol. 6, no. 5, pp. 807-814.

Bokel, C. 2008, "EMS Screens: from mutagenesis to screening and mapping", *Methods in Molecular Biology,* vol. 420, pp. 119-119-138.

Casso, D., Ramirez-Weber, F. & Kornberg, T.B. 2000, "GFP-tagged balancer chromosomes for Drosophila melanogaster", *Mechanisms of development,* vol. 91, no. 1-2, pp. 451-454.

Calvo, S.E., Pagliarini, D.J. & Mootha, V.K. 2009, "Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 18, pp. 7507-7512.Deans, T.L., Cantor, C.R. & Collins, J.J. 2007, "A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells", *Cell,* vol. 130, no. 2, pp. 363-372.

Dunipace, L., Ozdemir, A. & Stathopoulos, A. 2011, "Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression.", *Development,* vol. 138, no. 18, pp. 4075-4084.

Ferreira, J.P., Overton, K.W. & Wang, C.L. 2013, "Tuning gene expression with synthetic upstream open reading frames", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 110, no. 28, pp. 11284-11289.

Guzman, L., Belin, D., Carson, M. & Beckwith, J. 1995, "Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P-Bad Promoter", *Journal of Bacteriology,* vol. 177, no. 14, pp. 4121-4130.

Horn, T., Sandmann, T., Fischer, B., Axelsson, E., Huber, W. & Boutros, M. 2011, "Mapping of signaling networks through synthetic genetic interaction analysis by RNAi", *Nature Methods,* vol. 8, no. 4, pp. 341-U91.

Kosman, D., Mizutani, C., Lemons, D., Cox, W., McGinnis, W. & Bier, E. 2004, "Multiplex detection of RNA expression in Drosophila embryos", *Science,* vol. 305, no. 5685, pp. 846-846.

Kozak, M. 1987, "At Least 6 Nucleotides Preceding the Aug Initiator Codon Enhance Translation in Mammalian-Cells", *Journal of Molecular Biology,* vol. 196, no. 4, pp. 947-950.

Kumar, M., Srinivas, V. & Patankar, S. 2015, "Upstream AUGs and upstream ORFs can regulate the downstream ORF in Plasmodium falciparum", *Malaria Journal,* vol. 14, pp. 512.

Lee, T. & Luo, L. 1999, "Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis", *Neuron,* vol. 22, no. 3, pp. 451-461.

Matsui, M., Yachie, N., Okada, Y., Saito, R. & Tomita, M. 2007, "Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse", *FEBS letters,* vol. 581, no. 22, pp. 4184-4188.

McGinness, K., Baker, T. & Sauer, R. 2006, "Engineering controllable protein degradation", *Molecular cell,* vol. 22, no. 5, pp. 701-707.

Medenbach, J., Seiler, M. & Hentze, M.W. 2011, "Translational Control via Protein-Regulated Upstream Open Reading Frames", *Cell,* vol. 145, no. 6, pp. 902-913.

Noderer, W.L., Flockhart, R.J., Bhaduri, A., de Arce, A.J.D., Zhang, J., Khavari, P.A. & Wang, C.L. 2014, "Quantitative analysis of mammalian translation initiation sites by FACS-seq", *Molecular Systems Biology,* vol. 10, no. 8.

Ochoa-Espinosa, A., Yucel, G., Kaplan, L., Pare, A., Pura, N., Oberstein, A., Papatsenko, D. & Small, S. 2005, "The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 14, pp. 4960-4965.

Perry, M.W., Boettiger, A.N. & Levine, M. 2011, "Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo.", *Proc Natl Acad Sci U S A,* vol. 108, no. 33, pp. 13570-13575.

Salis, H.M., Mirsky, E.A. & Voigt, C.A. 2009, "Automated design of synthetic ribosome binding sites to control protein expression", *Nature biotechnology,* vol. 27, no. 10, pp. 946-U112.

Venken, K.J.T., Simpson, J.H. & Bellen, H.J. 2011, "Genetic Manipulation of Genes and Cells in the Nervous System of the Fruit Fly", *Neuron,* vol. 72, no. 2, pp. 202-230.

von Arnim, A.G., Jia, Q. & Vaughn, J.N. 2014, "Regulation of plant translation by upstream open reading frames", *Plant Science,* vol. 214, pp. 1-12.Wieland, M., Auslaender, D. & Fussenegger, M. 2012, "Engineering of ribozyme-based riboswitches for mammalian cells", *Methods,* vol. 56, no. 3, pp. 351-357.

Yen, L., Svendsen, J., Lee, J., Gray, J., Magnier, M., Baba, T., D'Amato, R. & Mulligan, R. 2004, "Exogenous control of mammalian gene expression through modulation of RNA self-cleavage", *Nature,* vol. 431, no. 7007, pp. 471-476.

APPENDICES

APPENDIX A

Supplementary Information for Chapter 3

A.1 - Supplemental Experimental Figures

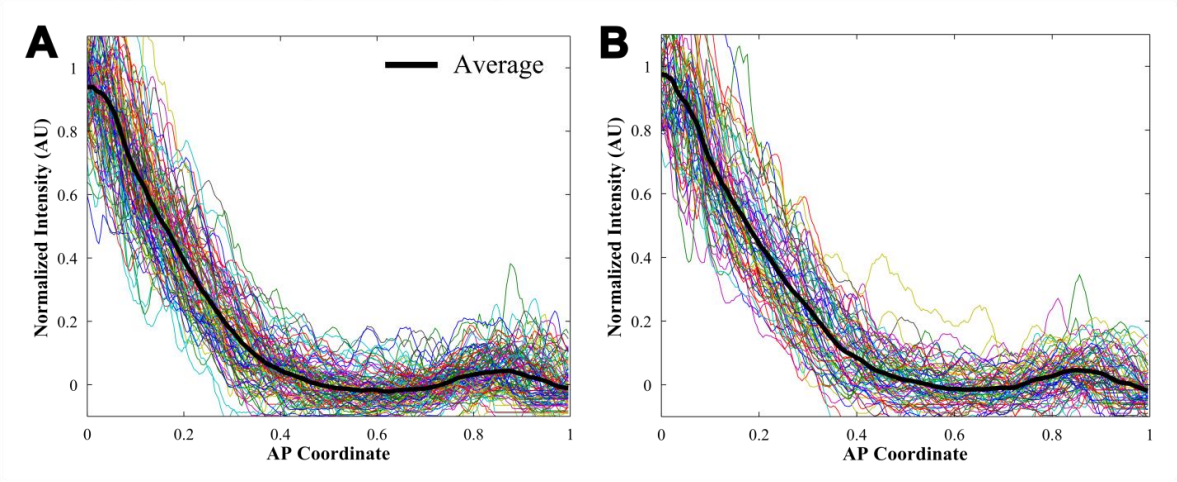

Figure A.1: *lacZ* expression in embryos with four copies of *gal4* and one copy of *gal80*. Each colored curve is the ventral or dorsal side of an individual embryo, (A) *UAS*x5:*gal80* and (B) *UAS*x3:*gal80*.

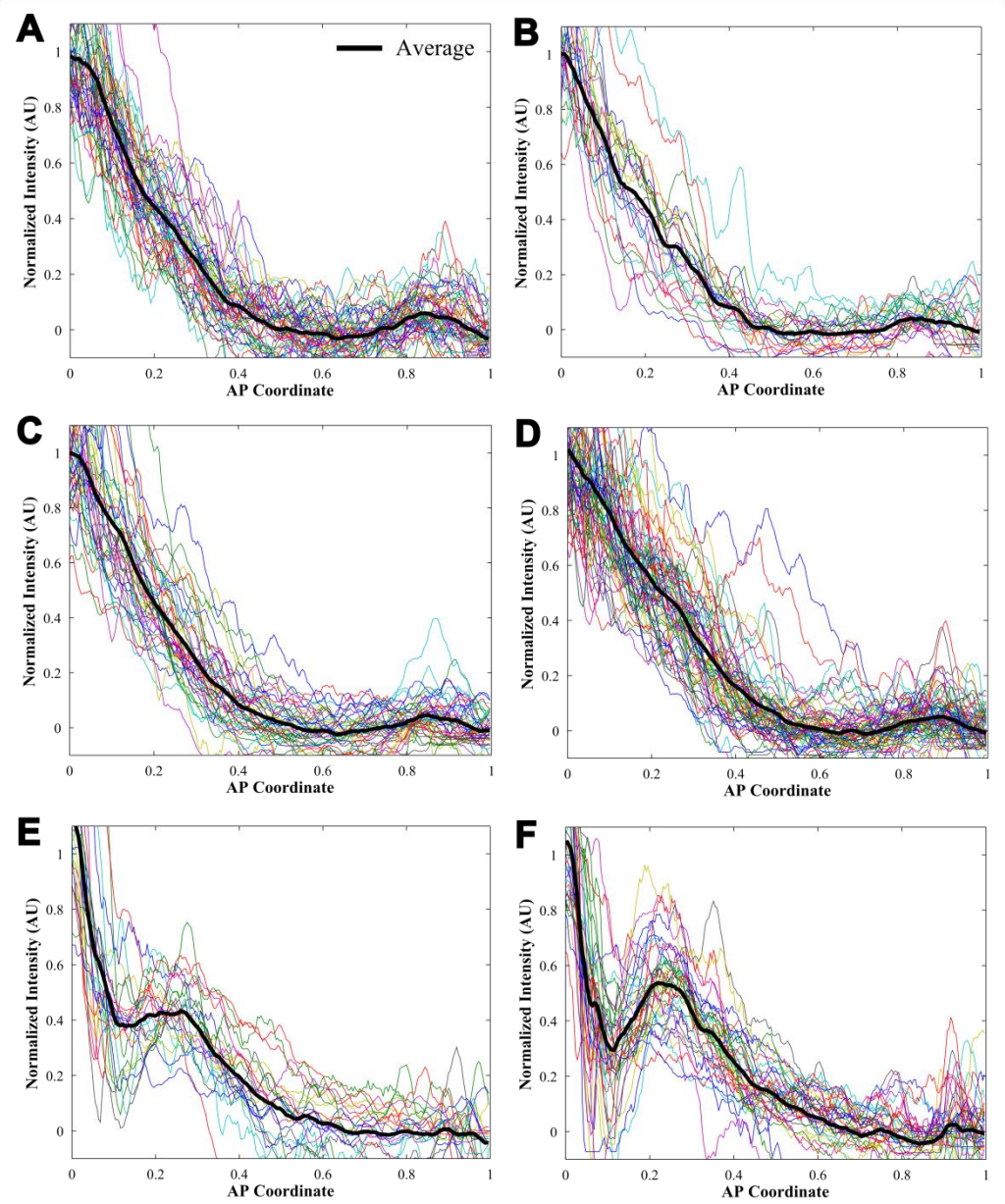

Figure A.2: *lacZ* expression in embryos with two copies of *gal4*. Embryos with: (A) no *gal80*, (B) two copies of *UAS*x3:*gal80*, (C) one copy of *UAS*x5:*gal80*, (D) two copies of *UAS*x5:*gal80* (no *gal3*), (E) two copies of *UAS*x5:*gal80* and *gt23*:*gal3* and (F) two copies of *UAS*x5:*gal80* and *evestr2*:*gal3*.

A.2 - Gal4/Gal80 Model formulation

The model was formulated with the following equations corresponding to Gal4, Gal80, and Gal4/Gal80 complex, respectively:

$$\frac{\partial G}{\partial t} = D_G \frac{\partial^2 G}{\partial X^2} - k_G G - k_{bind}(GR - K_D C)$$

$$\frac{\partial R}{\partial t} = D_R \frac{\partial^2 R}{\partial X^2} - k_R R - k_{bind}(GR - K_D C) + Q_R F_R(G)$$

$$\frac{\partial C}{\partial t} = D_C \frac{\partial^2 C}{\partial X^2} - k_C C + k_{bind}(GR - K_D C)$$

These equations are subject to no flux boundary conditions at $X = 0$ (anterior pole) and $X = L$ (posterior pole), except for the flux of Gal4 at the anterior pole is:

$$D_G \frac{\partial G}{\partial X}\Big|_{X=0} = -Q_G$$

The production of Gal80, as a function of Gal4, is as follows (Papatsenko, Levine 2011):

$$F_R(G) = \frac{(1 + G/K_{DNA})^n - 1}{(1 + G/K_{DNA})^n}$$

Here $n = 5$ for UASx5 Gal80.

A.3 - Scaling

Assuming steady state, we next scale these equations by the following transformations:

$$g = G/\bar{G}\,, r = R/\bar{R}\,, c = C/\bar{C}\,, x = X/L$$

The concentration scales are defined as follows:

$$\bar{G} = \frac{Q_{G,0}}{k_G L}, \bar{R} = \frac{Q_{R,0}}{k_R}, \bar{C} = \bar{G}$$

Here $Q_{G,0}$ is the flux of Gal4 that results from translation at the anterior pole in Gal4 x4 embryos, and $Q_{R,0}$ is the maximal production rate of Gal80 in embryos carrying on copy of *gal80*. This scaling results in the equations found in the main text. The constant flux boundary condition becomes

$$\lambda_g^2 \frac{dg}{dx}\Big|_{x=0} = -q_g$$

And the production function for Gal80 becomes:

$$f_R(g) = \frac{(1+g/K)^n - 1}{(1+g/K)^n}$$

with the following parameter definitions:

$$\lambda_g^2 = \frac{D_G}{k_G L^2}, \lambda_r^2 = \frac{D_R}{k_R L^2}, \lambda_c^2 = \frac{D_C}{k_G L^2}$$

$$\mu = \frac{k_{bind}\bar{R}}{k_G}, \nu = \frac{K_D}{\bar{R}}, \beta = \frac{k_G \bar{G}}{k_R \bar{R}}$$

$$\rho_c = \frac{k_C}{k_G}, q_r = \frac{Q_R}{Q_{R,0}}, q_g = \frac{Q_G}{Q_{G,0}}, K = \frac{K_{DNA}}{\bar{G}}$$

A.4 - Gal3

When considering Gal3, the following equations (for Gal3 and Gal3/Gal80 complex, respectively) were added to the model:

$$\frac{\partial B}{\partial t} = D_B \frac{\partial^2 B}{\partial X^2} - k_B B - k_3(BR - K_3 P) + Q_B F_B(X)$$

$$\frac{\partial P}{\partial t} = D_P \frac{\partial^2 P}{\partial X^2} - k_P P + k_3(BR - K_3 P)$$

Additionally, the term $-k_3(BR - K_3 P)$ also is included in the equation for Gal80. These equations are both subject to no-flux boundary conditions at $X = 0$ and $X = L$. The

production term for Gal3, $F_B(X)$, reflects the localization of Gal3 from the *gt23* enhancer (see Fig. A.3).

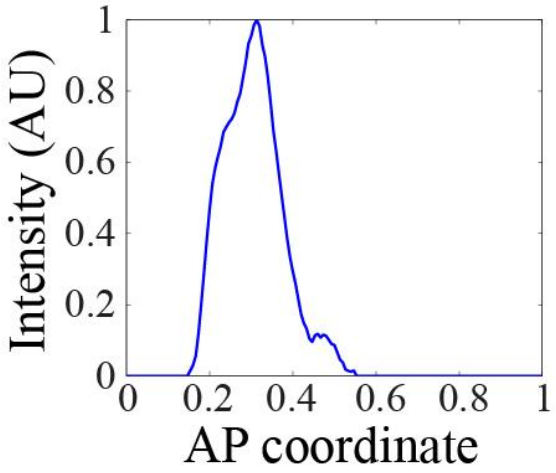

Figure A.3: Localization of *gt23:gal3* expression.

To scale these equations, we made the following transformations (at steady state):

$$b = B/\bar{B}\,, \qquad p = P/\bar{P}$$

The concentration scales are defined as follows:

$$\bar{B} = \frac{Q_{B,0}}{k_B L}, \bar{P} = \bar{R}$$

Here $Q_{B,0}$ is the maximal production rate of Gal3 in embryos carrying on copy of *gt23:gal3*.

This scaling results in the following equations:

$$0 = \lambda_b^2 \frac{d^2 b}{dx^2} - b - \eta\gamma(br - \sigma p) + q_b f_b(x)$$

$$0 = \lambda_p^2 \frac{d^2 p}{dx^2} - \rho_p p + \eta(br - \sigma p)$$

with the following parameter definitions:

$$\lambda_b^2 = \frac{D_B}{k_B L^2}, \lambda_p^2 = \frac{D_P}{k_R L^2}$$

$$\eta = \frac{k_3 \bar{B}}{k_R}, \sigma = \frac{K_3}{\bar{B}}, \gamma = \frac{k_R \bar{R}}{k_B \bar{B}}$$

$$\rho_p = \frac{k_P}{k_R}, q_b = \frac{Q_B}{Q_{B,0}}$$

A.5 - Optimization of the model to the observed lacZ profiles

Our optimization methodology proceeded as follows. First, a value of q_{2x} is chosen. Next, the best-fit values for λ_g and K are found by fitting the model to the *lacZ* data in embryos that lack *gal80*. Next, we use evolutionary optimization (Runarsson, Yao 2000, Runarsson, Yao 2005) to fit the model to the lacZ data for embryos with *gal80* (but without *gal3*). This gives us the further parameters λ_r, λ_c, μ, ν, β, and ρ_c. Finally, we use the *lacZ* data from embryos with *gal80* and *gt23:gal3* to constrain the parameters associated with Gal3. After these fits are found, the value of q_{2x} is varied (as 0.35, 0.40, 0.45, … 0.65) and we run the procedure again. Details follow.

A.5.1 - Optimization of embryos with no gal80

The first set of optimizations were performed on the *lacZ* data in embryos with no *gal80*. The model without Gal80 reduces to a single equation with only three unknown parameters: λ_g, K, and q_{2x}. q_{2x} was chosen as described above as an input to the optimization procedure. For a given choice of q_{2x}, λ_g was varied on an equally-spaced grid with 31 points from 0.05 to 0.2, and the Gal4 equation was solved for Gal4 x4 ($q_r = 1$) and

Gal4 x2 ($q_r = q_{2x}$). Next, K was varied on a log-space with 200 points from 0.1 to 1000. The Gal4 gradients were then used to compute the lacZ profiles in both cases (where the lacZ profile was equal to $f_r(g)$). The cost function was defined as follows:

$$F = \sum_{i=a}^{i=b} f_{4x,i}^2 + f_{2x,i}^2$$

where a is the index for $x_i = 0.1$, and b is the index for $x_i = 0.9$ (we limit ourselves to the central 80% of the embryo to avoid curvature effects). The cost coefficient for Gal4 x4 embryos is

$$f_{4x,i} = \frac{Y_{4x,i} - \alpha_0 L_{4x,i}}{S_{4x,i}}$$

and where $Y_{4x,i}$ is the value of the observed mean lacZ profile at x_i in Gal4 x4 embryos, $L_{4x,i}$ is the value of the simulated lacZ profile at x_i in Gal4 x4 embryos, and $S_{4x,i}$ is the standard deviation of the observed lacZ profiles at x_i in Gal4 x4 embryos. α_0 is a scaling constant that minimizes the difference between Y_{4x} and L_{4x} for x between x_a and x_b.

Similarly, the cost coefficient for Gal4 x2 embryos is

$$f_{2x,i} = \frac{Y_{2x,i} - \alpha_0 L_{2x,i}/\alpha_{2x}}{S_{2x,i}}$$

where $Y_{2x,i}$ is the value of the observed mean lacZ profile at x_i in Gal4 x2 embryos, $L_{2x,i}$ is the value of the simulated lacZ profile at x_i in Gal4 x2 embryos, and $S_{2x,i}$ is the standard deviation of the observed lacZ profiles at x_i in Gal4 x2 embryos. α_{2x} is a scaling constant that minimizes the difference between Y_{2x} and $\alpha_0 L_{2x}$ for x between x_a and x_b.

Using this procedure, we found the best-fit λ_g and K for every value of q_{2x} that served as input.

A.5.2 - Optimization of embryos with gal80 but no gal3

The next set of optimizations were performed on the *lacZ* data in embryos with *gal80* but no *gal3*. The model with Gal80 has the additional parameters λ_r, λ_c, μ, ν, β, and ρ_c. (λ_g and K were held fixed at their best-fit values from the above optimization.)

We used an improved stochastic evolutionary optimization with penalty constraints (Runarsson, Yao 2000, Runarsson, Yao 2005) to find a cloud of parameter sets that best- fit the *lacZ* data from embryos with *gal80* but no *gal3*. For each input value of q_{2x}, we found 100 evolutionary-optimized parameter sets.

The cost function was defined as follows:

$$F = \sum_{i=a}^{i=b} (1-\chi) f_{1x80,i}^2 + \chi f_{2x80,i}^2$$

where a and b are described above. The cost coefficient for Gal4 x4/Gal80 x1 embryos is

$$f_{1x80,i} = \frac{Y_{1x80,i} - \alpha_0 L_{1x80,i}/\alpha_{1x80}}{S_{1x80,i}}$$

and where$Y_{1x80,i}$ is the value of the observed mean lacZ profile at x_i in Gal4 x4/Gal80 x1 embryos, $L_{1x80,i}$ is the value of the simulated *lacZ* profile at x_i in Gal4 x4/Gal80 x1 embryos, and $S_{1x80,i}$ is the standard deviation of the observed *lacZ* profiles at x_i in Gal4 x4/Gal80 x1 embryos. α_{1x80} is a scaling constant that minimizes the difference between Y_{1x80} and $\alpha_0 L_{1x80}$ for x between x_a and x_b.

The cost coefficient for Gal4 x2/Gal80 x2 embryos is analogous. The parameter χ is a scaling parameter that describes the weighting between the two sets of embryos. We chose $\chi = 0.5$ (for optimization of both scenarios) and $\chi = 0$ (for optimization of attenuation only).

Using this procedure, we found clouds of optimized parameter sets for the Gal80 parameters.

A.5.3 - Optimization of embryos with gal80 and gal3

The final optimization was performed on the *lacZ* data in embryos with both *gal80* and *gal3*. The model with Gal80 and Gal3 has the additional parameters λ_b, λ_p, η, σ, γ, and ρ_p. To simplify the optimization, we assumed both diffusion length scales were zero, that non-specific degradation of the Gal3/Gal80 complex was equal to that of Gal80 alone (so that $\rho_p = 1$), and that relatively tight binding occurs ($\sigma = 0.01$). These assumptions left us with two parameters to vary for the optimization: the forward binding rate for Gal3 and Gal80, η, and the ratio of maximal production rates of Gal80 to Gal3, γ.

For each parameter set found by evolutionary optimization above, we used brute-force optimization to determine the best-fit (η, γ) pair. γ was varied on an log-spaced grid with 31 points from 10^{-1} to 10^{2}, and η was varied on a log-space with 51 points from 10^{0} to 10^{6}.

The cost function was defined as follows:

$$F = \sum_{i=a}^{i=b} f_{3,i}^2$$

where a and b are described above. The cost coefficient is

$$f_{3,i} = \frac{Y_{3,i} - \alpha_0 L_{3,i}/\alpha_3}{S_{3,i}}$$

and where$Y_{3,i}$ is the value of the observed mean lacZ profile at x_i in Gal4 x2/Gal80 x2/Gal3 x1 embryos, $L_{3,i}$ is the value of the simulated *lacZ* profile at x_i, and $S_{3,i}$ is the standard deviation of the observed *lacZ* profiles at x_i. α_3 is a scaling constant that minimizes the difference between Y_3 and $\alpha_0 L_3$ for x between x_a and x_b.

References

Papatsenko, D. & Levine, M. 2011, "The Drosophila Gap Gene Network Is Composed of Two Parallel Toggle Switches", *Plos One,* vol. 6, no. 7, pp. e21145.

Runarsson, T. & Yao, X. 2005, "Search biases in constrained evolutionary optimization", *Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews,* vol. 35, no. 2, pp. 233-243.

Runarsson, T. & Yao, X. 2000, "Stochastic ranking for constrained evolutionary optimization", *Ieee Transactions on Evolutionary Computation,* vol. 4, no. 3, pp. 284-294.

APPENDIX B

Supplementary Information for Chapter 4

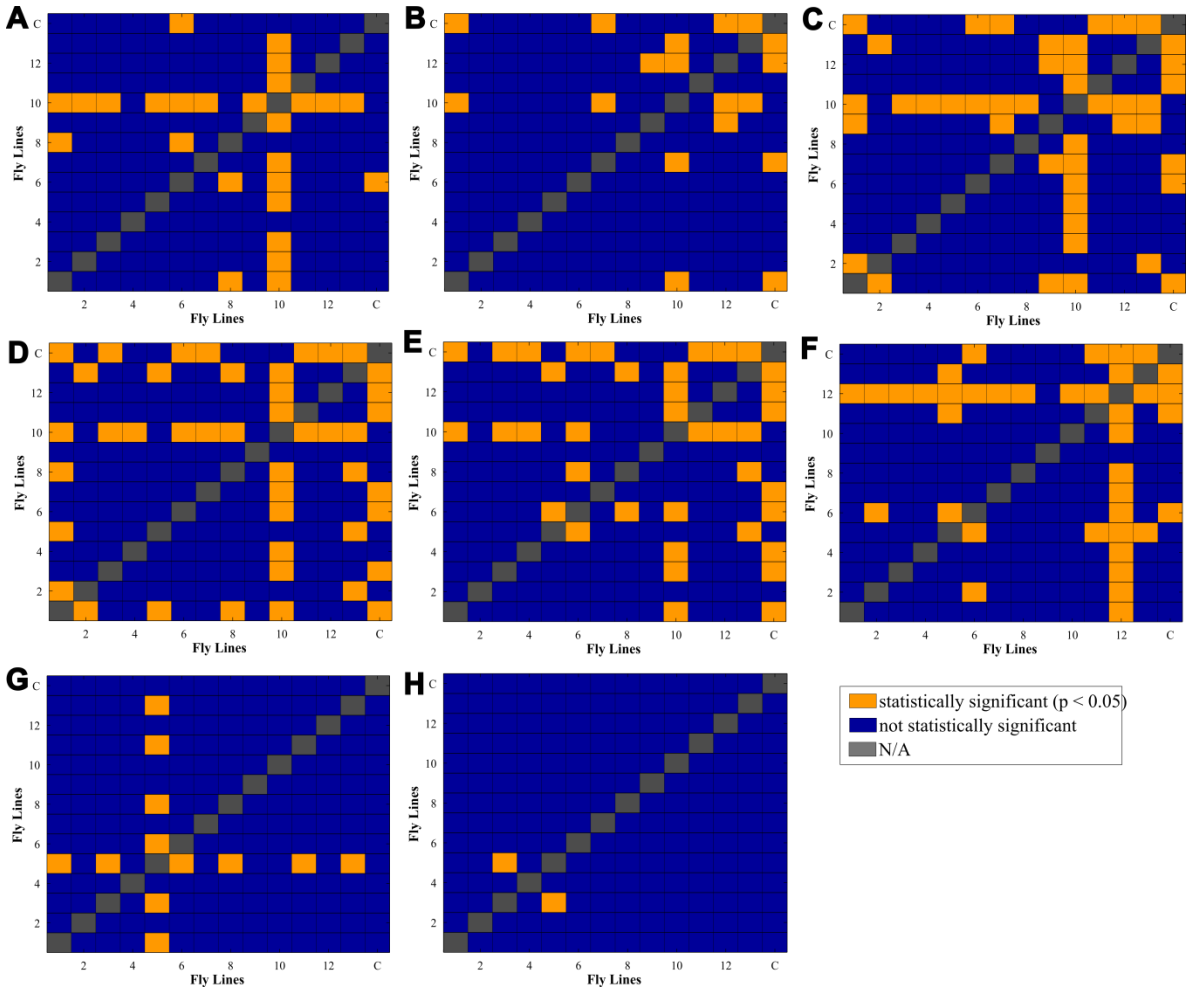

Figure B.1: Pair-wise comparison of position of *Kr* and Eve in each of the fly lines. Results of post-hoc Tukey-Kramer test, where orange denotes statistically significant ($p < 0.05$) differences between the lines and blue denotes no statistically significant difference between the lines. The fly lines are in the order of: RAL150, RAL306, RAL307, RAL315, RAL317, RAL360, RAL41, RAL57, RAL705, RAL761, RAL765, RAL799, RAL801, and laboratory control; for (A) Kr Posterior, (B) Eve stripe 1, (C) Eve stripe 2, (D) Eve stripe 3, (E) Eve stripe 4, (F) Eve stripe 5, (G) Eve stripe 6, and (H) Eve stripe 7.

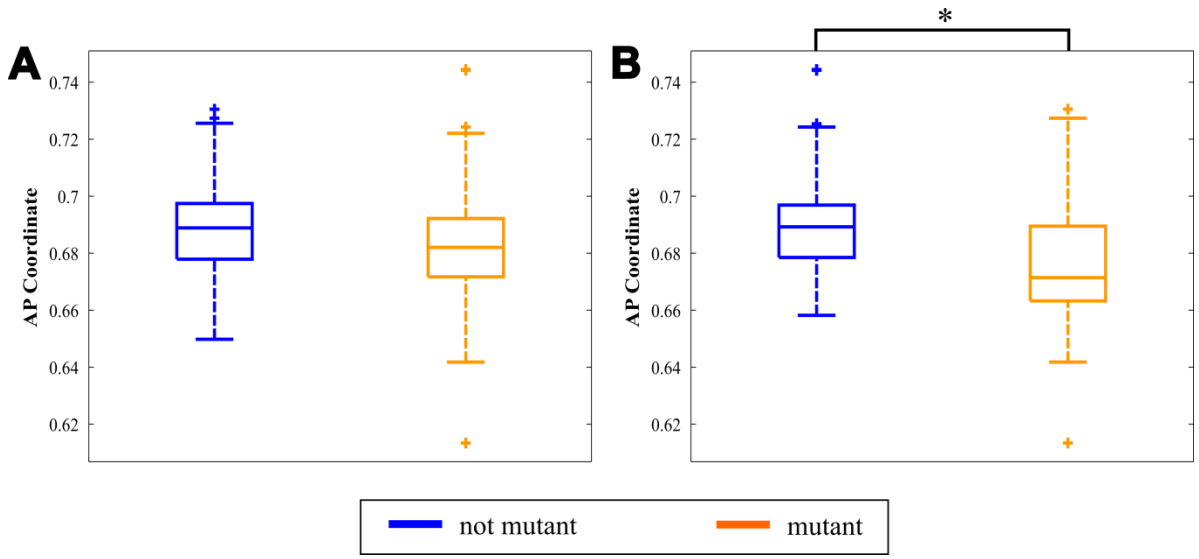

Figure B.2: Example of Association Mapping. (A) Comparison of position of Eve stripe 6 in non-mutant lines (n_{lines} = 8, $n_{embryos}$ = 88), compared to mutant lines (n_{lines} = 5, $n_{embryos}$ = 88) for a non-significant SNP. (B) For a significant SNP (p = 0.045), this SNP shows a correlation (anterior shift in Eve stripe 6) between non-mutant lines (n_{lines} = 8, $n_{embryos}$ = 123) and mutant lines(n_{lines} = 5, $n_{embryos}$ = 53).

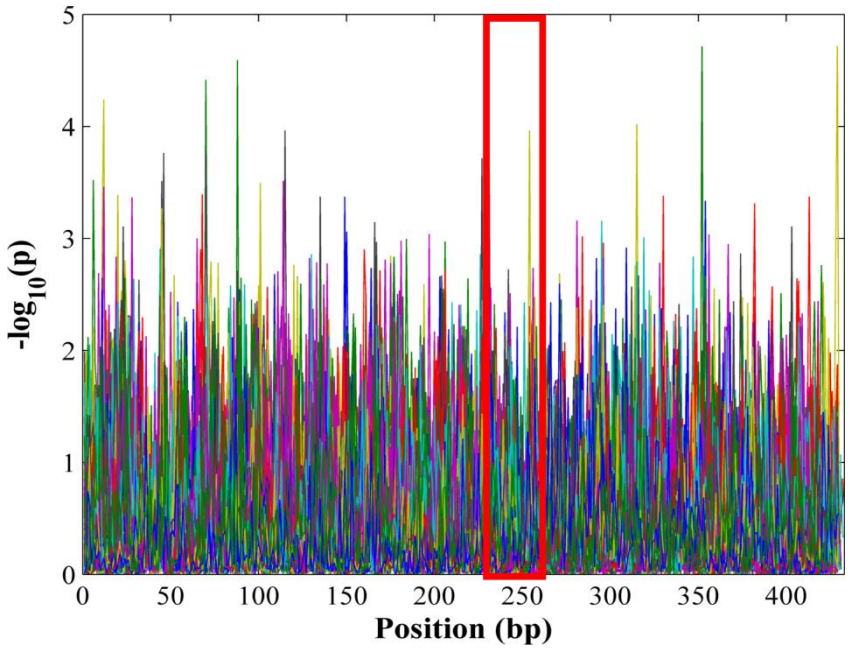

Figure B.3: Position weight matrix analysis to find probably transcription factor binding sites. Analysis for EveA enhancer trap, where each line denotes one transcription factor motif (only motifs shown to be present in one enhancer trap are shown). The region directly surrounding the SNP is denoted by the red box.

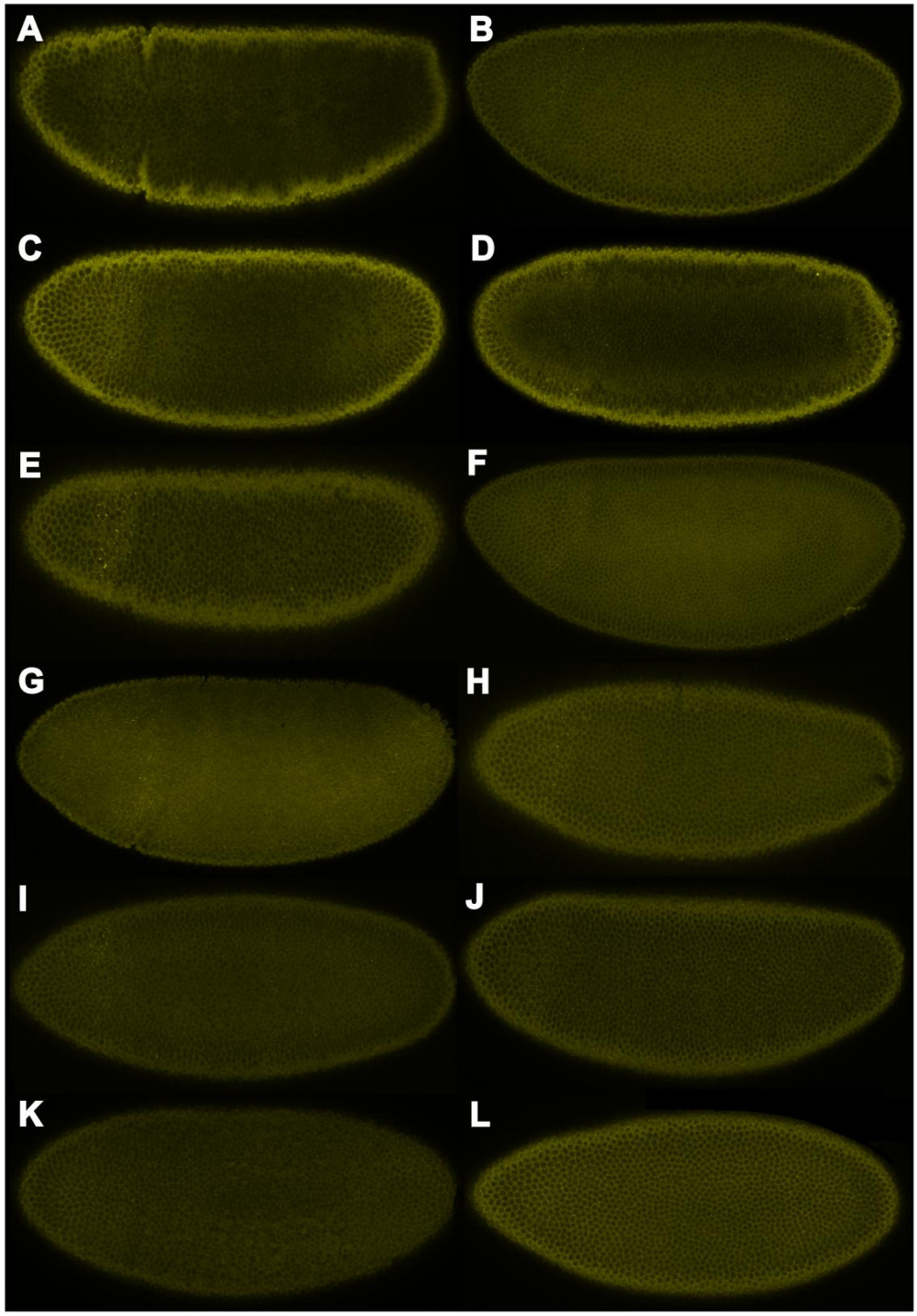

Figure B.4: Expression of *lacZ* due to putative enhancer activity where no expression or expression due to only the minimal promoter is observed. (A) EveD, (B) EveE, (C) EveF, (D) EveG, (E) EveH, (F) EveI, (G) EveJ, (H) EveK, (I) EveL, (J) KrB, (K) KrC, and (L) KrD.

Table B.1: Primers used to amplify enhancers from genomic DNA. Restriction enzyme sites in capital letters. Genomic region of enhancer is shown compared to start of respective gene.

Enhancer		Primer Sequence	Genomic Region	
			Start	*Stop*
KrA	Fwd	cagacatgGAATTCgtcctttaacggtcaacacatag	-473	0
	Rev	gtcagtacAGATCTtgttccttcgctgacagag		
KrB	Fwd	cagacatgGAATTCcggaatttgccaacacacca	-16,526	-16,365
	Rev	gtcagtacAGATCTggtccaagtccgctagcaca		
KrC	Fwd	cagacatgGAATTCacagttagaaggccaaaca	-7,983	-7,703
	Rev	gtcagtacAGATCTccgcataaaagcaaatgctg		
KrD	Fwd	cagacatgGAATTCgtctgagttgagcattagtgag	-4,750	-4,446
	Rev	gtcagtacAGATCTgctGGCGCGCCgaaacgtagagtcaagatcaagg		
EveA	Fwd	cagacatgGAATTCttgagcagttccaatgcct	-19,870	-19,432
	Rev	agctacgaGGCGCGCCgcggtggtttcttacaatagg		
EveB	Fwd	cagacatgGAATTCcctattgtaagaaaccaccgcttgc	-19,452	-18,977
	Rev	gtcagtacAGATCTcctcaacctggaaatgctttgt		
EveC	Fwd	agctacgaGGCGCGCCtgcttccttactcttaatgtttccg	-12,663	-12,311
	Rev	gtcagtacAGATCTtgcattttcctggtaagact		
EveD	Fwd	cagacatgGAATTCaatagcatgtagtggacgag	-19,906	-19,425
	Rev	agctacgaGGCGCGCCaatgcaagcggtggtttct		
EveE	Fwd	cagacatgGAATTCtccgctacgcccagtgactt	-16,641	-16,115
	Rev	agctacgaGGCGCGCCaggcaacctgtgggatatgttggta		
EveF	Fwd	cagacatgGAATTCgactacGGCGCGCCggccaaacaaagcaaacat	-13,559	-13,363
	Rev	gtcagtacAGATCTtagccagaagacctgagaa		
EveG	Fwd	cagacatgGAATTCtgaacctgcaacatatgga	-12,871	-12,632
	Rev	gtcagtacAGATCTgagatgtcggaaacattaagag		
EveH	Fwd	cagacatgGAATTCtaatccttttgcccacgagc	-11,983	-10,886
	Rev	gtcagtacAGATCTcttgactgttttggcgattt		
EveI	Fwd	cagacatgGAATTCcagcagactgatcgaatcatttgtt	-9,619	-9,182
	Rev	gtcagtacAGATCTcaccttccagctgagcgtt		
EveJ	Fwd	agctacgaGGCGCGCCgaaagggcaagggcaaggtca	-8,951	-8,336
	Rev	gtcagtacAGATCTggcgcccgaatccaaatta		
EveK	Fwd	cagacatgGAATTCttaatgggtagcagcgttgccagat	-6,527	-6,069
	Rev	gtcagtacAGATCTtgaatgactttggttccttcggat		
EveL	Fwd	cagacatgGAATTCcaggttgcattgccaatgga	1,925	2,307
	Rev	gtcagtacAGATCTctcgcatcgtttacccatcgt		

Table B.2: Sequence for Mutagenesis Primers. Mutation in capital letters.

Enhancer		Primer Sequence
EveA	Forward	gatatttcttcAgatttgctaaaaacacggaagtaaacaaaagtg
	Reverse	gtttttagcaaatcTgaagaaatatcatttgcaaaatgtcgcaaac
EveB	Forward	cgccgctgaacaaTtaacatctcaatcgcaagc
	Reverse	gagatgttaAttgttcagcggcgcaggtagaatgtg
EveC	Forward	ggcgaattctttTcaatttggtaaatagtggcaactacaatac
	Reverse	ccaaattgAaaagaattcgccaaggaaatcgcttgaag
KrA	Forward	gaaaggaacCCctctagctgtctcattcgcacc
	Reverse	gacagctagagGGgttcctttcaatgcaaaagatatatag

Table B.3: Shifts in positions of *Krüppel* in mutant fly lines. Shifts not statistical significant are not shown ($p > 0.05$).

			Anterior Border	Posterior Border
Laboratory Control		**position (% EL)**	40.4 ± 1.3	55.3 ± 2.2
med	**BS9033**	**position (% EL)**	41.4 ± 1.0	
		p	0.0080	
	BS9006	**position (% EL)**	42.1 ± 1.2	
		p	9.72×10^{-6}	
usp	**BS4660**	**position (% EL)**	42.9 ± 1.0	54.0 ± 1.4
		p	1.53×10^{-7}	0.0221
	BS31414	**position (% EL)**	42.1 ± 0.6	
		p	4.38×10^{-6}	
gcm	**BS31518**	**position (% EL)**	42.3 ± 1.4	56.6 ± 0.6
		p	0.0171	0.0066
otp	**BS57582**	**position (% EL)**	41.4 ± 0.9	
		p	0.0082	

Table B.4: Shifts in the positions of Eve stripes in mutant fly lines. Shifts not statistical significant are not shown ($p > 0.05$).

			Stripe 1	Stripe 2	Stripe 3	Stripe 4	Stripe 6	Stripe 7
Laboratory Control		**position (% EL)**	32.4 ± 1.0	40.1 ± 1.1	47.7 ± 1.2	54.6 ± 1.3	68.9 ± 1.2	78.7 ± 1.6
		position (% EL)					67.9 ± 1.4	77.3 ± 1.2
	BS9033	**p**					0.0300	0.0025
		position (% EL)	33.9 ± 1.4	42.0 ± 1.6	49.2 ± 1.7	55.8 ± 1.8		
med	**BS9006**	**p**	0.0002	5.93 x 10-5	0.0022	0.0138		
		position (% EL)	33.0 ± 0.7	41.4 ± 0.8				77.4 ± 1.5
usp	**BS4660**	**p**	0.0280	4.51×10^{-5}				0.0130
		position (% EL)					70.6 ± 1.5	80.3 ± 1.2
gcm	**BS28913**	**p**					0.0097	0.0033
		position (% EL)						77.2 ± 1.9
otp	**BS57582**	**p**						0.0249

Table B.5: Genes identified from RNAseq analysis whose expression levels are correlated with variations in *Kr* and Eve positioning

CG13362	CG31661	mey	CG42815
e(r)	cold	CR12628	CG12269
Cyp6d2	CG13239	IM3	CG30345
ord	Pvf2	swi2	CG13405
eIF-5A	Sirup	JhI-26	CG17018
RpL28	CG13116	krimp	VhaPPA1-1
CG33160	CG6180	CG8089	pod1
CG12721	CG16957	PGRP-SC2	CG14218
CG14095	Tep1	CG1354	Tao
Acp76A	Cyp303a1	Jon25Bi	e(r)
Rad9	RpS25	CG34298	CR43670
CG4306	CG10013	SelR	Rpb5
CG10749	Hsc70-2	CG6675	CG14117
eIF-4E	CG12269	CG7091	UK114
GNBP3	MtnC	dpr4	CG15412
CG13678	CG5621	CG4449	CG12061
CG10591	TwdlQ	wa-cup	CG43085
CG10672	RpS10a	CG32448	CG40198
lcs	CG15535	CG42819	Pde9

APPENDIX C

Supplementary Information for Chapter 6

Table C.1: Sequences of Oligos Used for Upstream Competing Sequences. Capital letters denotes the sequence designed to anneal to the ribozyme.

Upstream Sequence		**Sequence**
R1	Fwd	cgactgcgcagtcgaaaaaaGCTGGA
	Rev	tcgagTCCAGCttttttcgactgcgcagtcgg
R2	Fwd	cgactgcgcagtcgaaaaaaGCTGGAAG
	Rev	tcgagCTTCCAGCttttttcgactgcgcagtcgg
R3	Fwd	cgactgcgcagtcgaaaaaaGTGGT
	Rev	tcgagACCACttttttcgactgcgcagtcgg
R4	Fwd	cgactgcgcagtcgaaaaaaCTCG
	Rev	tcgagCCAGttttttcgactgcgcagtcgg
R5	Fwd	cgactgcgcagtcgaaaaaaGCTAGAAG
	Rev	tcgagCTTCTAGCttttttcgactgcgcagtcgg
R6	Fwd	cgactgcgcagtcgaaaaaaCGGGTT
	Rev	tcgagAACCGttttttcgactgcgcagtcgg
R7	Fwd	aattccgactgcgcagtcgaaaaaaCTCGAc
	Rev	tcgagTCGAGttttttcgactgcgcagtcgg
R8	Fwd	cgactgcgcagtcgaaaaaaTTGG
	Rev	tcgagCCAAttttttcgactgcgcagtcgg
R9	Fwd	cgactgcgcagtcgaaaaaaCTGGAT
	Rev	tcgagATCCAGttttttcgactgcgcagtcgg
R10	Fwd	cgactgcgcagtcgaaaaaaGGAT
	Rev	tcgagATCCttttttcgactgcgcagtcgg