Measuring and Modeling Enhancers in Perturbed Drosophila Melanogaster Embryos

Measuring and modeling enhancers in perturbed *Drosophila melanogaster* embryos

A dissertation presented

by

Max Valentin Staller

to

The Committee on Higher Degrees in Systems Biology

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the subject of

Systems Biology

Harvard University Cambridge, Massachusetts

October 2014

Dissertation Advisor: Angela H. DePace Max Valentin Staller

Measuring and modeling enhancers in perturbed Drosophila melanogaster embryos

Abstract

The diversity of animal shapes and sizes, colors and textures, or behaviors and habitats all depend on specialized cells. A newly fertilized embryo must build all these specialized cell types, a process called differentiation. Much of differentiation depends on appropriately turning genes on and off in each cell type. Cell type specific control of gene expression is encoded in a type of regulatory DNA called enhancers. I am interested in how enhancers control the cell type specific gene expression that enables specialized cell functions.

Enhancers read in information from regulatory proteins and output a level of gene expression. This conversion from input regulator concentrations to output expression level is a computation. I use quantitative measurement and computational modeling to study how enhancers *compute*. In embryos, many regulatory proteins bind to enhancers, and some will turn an enhancer on, while others will turn it off. This complex process is greatly simplified by employing computational models. These models can test whether all regulators have been identified (and if not, find the missing ones) and quantify the relationships between regulators. The relationships between regulators reflect the underlying molecular mechanisms used in the cell; when several models can

fit the data, perturbation experiments can be used to distinguish the models and underlying mechanisms. However, most computational models of gene expression in animals have not been rigorously validated by perturbation experiments. A major contribution of my thesis work was developing methods for testing models.

To test computational models for how enhancers compute gene expression patterns, I experimentally manipulated the concentrations of regulatory proteins and precisely measured output gene expression patterns. Using the *Drosophila melanogaster* blastoderm embryo, I first developed efficient and scalable techniques for making perturbations to regulatory protein concentrations. This technique revealed a postulated property of development: that embryos mitigate the impact of perturbations by preventing the creation of new cell types. I then used two perturbations to test computational models of an enhancer, finding they were incomplete and discovering new regulatory connections. My work illustrates how computational modeling and quantitative measurement are powerful tools for untangling how regulatory DNA operates in embryos.

Table of contents place holder

tara suggested doing the table of contents last

Acknowledgments

I would not be preparing this Dissertation, nor defending this Ph.D. without the help of many people. First and foremost, I need to thank my advisor Angela DePace for mentoring me through this degree. Among a thousand things, she taught me how to finish a project, how tell a story, and how to mentor a student. Angela goes beyond leading by example to always explain her mentoring intension and strategy. She excels at customizing her mentoring style to each of her students, blending periods of high engagement with space to explore and operate. Some of her philosophies I will take with me include: Love people for their strengths and support them through their weaknesses. The goal of the lab is to help people get where they want to go. Always do science in accordance with your values. Beyond pushing us to do excellent science, she pushes us to be excellent people.

One of Angela's strongest talents is collecting wonderful people to work in her lab. As a result, the DePace lab is a magical place to work, where people deeply engage with each others lives and science. Thank you current and past lab members Kelly Eckenrode, Ian Sudbery, Sarah Saminadin-Peter, Javi Estrada Diez, Jeehae Park, Francheska Lopez Rivera, Clarissa Scholes. Meghan Bragdon for her unflappable enthusiasm and energy. Tara Martin for always being willing to make an idea bigger. Zeba Wunderlich for teaching me how to prioritize and focus on the important questions. My bay mate, Ben Vincent for being the bossy younger brother who kept me on task.

I have been lucky to live in Boston with a fabulous cohort of friends. Hanlin Tang, Julia Liu, Sarah Pfau, Emily Capra, Peter Combs, Sandhya Sinha, Eric Batchelor, Matt Thompson, Evan Kingsley and Namita Bisaria. Monday evening dessert with my

Systems Biology classmates, Moran Cabili Kalmar, Niall Mangan, Paul Cowgill, Danny Lieber, Eric Kelsic, Ashley Wolf, John Bachmann and Steve Hershman, was one of the best parts of graduate school. Leo Ungar and Danny Lieber were the best roommates I could have asked for: through laser tag and cooking curry, shoveling the driveway or adventuring to Trader Joe's, they taught me to love living in Boston.

In graduate school, I have had the privilege to mentor outstanding students: Jue Wang, Virginia Fedrigo, Evi Van Itallie and Adam Carte. Each of them taught me something new about teaching and science. The program in Systems Biology and the Hershel Smith Fellowship supported my summer in Woods Hole at the Embryology course, a time that transformed how I think about ecology, evolution and development. Among the dozens of fantastic instructors from this course, I am particularly grateful to Nipam Patel, Alejandro Sánchez Alvarado, and David Sherwood.

I am grateful to all the teachers who have helped me along the way. In middle school and Gary Vorwald and at Exeter, Anne Rankin, Stanley Lo, Scott Saltman. They made science fun and pulled me in. Russell Weatherspoon both kept an eye on me and imparted many aliquots of wisdom. At Princeton, the Integrate Science program showed me dynamism of interdisciplinary science and sent me down the path to Systems Biology. The professors were unparalleled in their engagement: Bill Bialek, David Botstein, Josh Rabinowitz, Josh Shaevitz, Ned Wingreen, Leonid Kruglyak, Olga Troyanskaya, Eric Wieschaus, Manuel Llinás and Colleen Murphy. The program could not have functioned without the amazing Lewis-Sigler Fellows: Maitreya Dunham, Will Ryu and Amy Caudy.

My decision to pursue science was catalyzed in no small part by the summer I spent living in Princeton with Max Bialek. During the day Sandy Silverman showed the ropes in David Botstein's lab. In evenings, Bill and Char Bialek demonstrated how academics could lead wonderful lives outside of the lab.

For my first exposure to lab work, I am indebted to Jorge Benach, a phenomenal role model and staunch supporter. To Sean Connolly, the graduate student who watched over me in Jorge's lab, I owe all of my best lab habits. To David Botstein, I am especially grateful for his teaching me how to design experiments and for infecting me with his passion for teaching. All three of the labs I have worked in are unified by deeply supportive cultures.

At Harvard many people kept an eye on my through graduate school. Uri Alon taught me that science is a form of self-expression akin to art. Pam Silver helped me navigate rotations. Johann Paulsson, Vamsi Mootha, and Jeremy Gunawardena sat on my exam committees. The members of Dissertation Advisor Committee, Ting Wu, Tim Mitchison and Roy Kishony were always there to help me through the ups and downs of graduate school, helping me to quit projects and seize new opportunities. My fortuitous collaboration, Chapter 2, with Norbert Perrimon was a phenomenal crash course in scientific writing. Periodic dinners with Natasha Staller and Gary Ruvkun, always offered long term perspective and helpful advice. I must emphatically thank Sam Reed for protecting Systems Biology Ph.D. students from Harvard bureaucracy and for making the program function.

Through the highest peaks and lowest crevasses of graduate school, my staunchest supporter has been my girlfriend, Ashley Wolf. Her pragmatism and intuition

for how the world operates has been an endless help for navigating professional and bureaucratic obstacles. She combines grace with a warm heart to be a reservoir of support that helps me restart when I get stuck.

Finally, I must thank my family for always being there. Valentin, for, among other things going to Harvard and hanging out with me for four year, and being the best wingman I could ask for. Right when I thought I could never get away from my experiments, he led us on a cross-country road trip that let me pivot and reset. Yzzy, for inspiring some part of my adventure in science, but more for taking care of all of us and reminding us of our humanity. Mom and Pop, for encouraging and enabling this adventure in science, for your constant willingness to listen, and for always fully supporting my decisions.

Chapter 1: Introduction

A newly fertilized egg must divide and differentiate into a vast array of different cell types. To achieve distinct behaviors and morphologies, cells activate different sets of genes (Davidson, 2010). In animals, the majority of tissue specific and developmental gene regulation is localized to regulatory DNA called enhancers (Levine et al., 2014). Enhancers can be located far away from their target genes, either upstream or downstream. They contain binding sites for key regulatory proteins called transcription factors, which can either activate or repress gene expression. Importantly, some configurations of bound transcription factors activate gene expression and other other configurations repress gene expression. By acting as scaffolds, enhancers read the concentrations bound transcription factors and output a level of expression. This conversion from transcription factor concentration to output expression is a computation. In this sense, enhancers are like a logic gate in a computer. The primary role of enhancers is to compute cell type specific gene expression. Understanding how enhancers compute is essential for linking genotype to phenotype.

In complex developmental systems, deciphering how enhancers "compute" requires computational models. Computational models formalize our assumptions and allow us to explore the logical outcome of those assumptions (Gunawardena, 2014). In this thesis, our models make testable hypotheses that we use to prioritize experiments. We use simple models because they require fewer assumptions and are easier to interpret. Comparing the performance of similar models can isolate the effect of changing one assumption. Our goal is to break our models; breaking models reveals incorrect assumptions and teaches us something new.

To formally describe how enhancers compute, I will use the input/output function. An enhancer input/output function is a multidimensional dose response curve: it maps the concentrations of input regulators to the output level of gene expression (Figure 1.1). This modeling approach therefore “black boxes” the exact mechanism of gene activation and focuses on how enhancers process information from input transcription factors to set output levels. Input/output functions have also been called *cis*-regulatory input functions, gene regulation function or gene regulatory functions (Rosenfeld, 2005; Mayo et al., 2006).

Enhancer input/output functions can be directly measured or inferred by computational modeling. In bacteria, a series of beautiful papers measured input/output functions for a set of promoters by systematically varying environmental signals (Setty et al., 2003; Mayo et al., 2006; Kaplan et al., 2008). The *lac* promoter, with only 3 binding sites, performed a complex computation (Setty et al., 2003). The analogous experiment in animal embryos is to systematically vary input transcription factors for one enhancer and measure activity. Initially, I attempted this experiment, but the interconnectivity of the network made it difficult to create the necessary blank slate embryo. An alternative approach is to infer an computational model for the input/output function from endogenous data (Sharp and Reinitz, 1998; Janssens et al., 2006; Ilsley et al., 2013). This approach is fast and easy, but when two inputs are highly correlated, these correlations can confound inference. Reciprocally, if two inputs never occur together, it is not possible to detect interactions between them. A further advantage of fitted models is that they are straight forward to test with experimental perturbations.

In this thesis, I use quantitative experimental perturbations to test computational models of the input/output function of an enhancer. The first challenge was making perturbations in an efficient way. Towards this end, we developed and characterized methods for using short hairpin RNAs (shRNAs) to deplete transcripts in the early embryo (Chapter 2). With this technique in hand, we characterized one perturbation in detail by measuring the activity of the segmentation network in every cell of a perturbed embryo (Chapter 3). We found that all gene expression pattens changed in concert as the network robustly specified cell fates. Finally, we used two perturbations to test computational models of an enhancer input/output function (Chapter 4). By collecting new data to test models, we added new connections to a classic gene regulatory network. In the process, we uncovered the first example of how two enhancers can use different input regulators to create the same output pattern. We resolved a long standing confusion over regulatory redundancy in this locus by showing there are two distinct, parallel computations.

Long term goals in the field

When we sequence a new genome, we can identify protein coding genes, but it is still not possible to identify regulatory DNA, especially enhancers, from sequence alone. It is even harder to predict when and where in the animal enhancers will be active. For proteins, knowledge of the genetic code and splicing signals enables identification of protein coding genes in new genomes. By analogy, a mechanistic understanding of gene regulation will likely help the community read regulatory DNA in new genomes.

This is clearly a pressing problem: the vast majority of human genetic variants associated with disease map to regulatory DNA (Maurano et al., 2012). Each of these regions contains many sequence changes, and a major challenge in personalized medicine will be to build models that can predict which mutations in regulatory DNA will affect gene expression and which are benign. Building models of how regulatory DNA computes will help solve these problems.

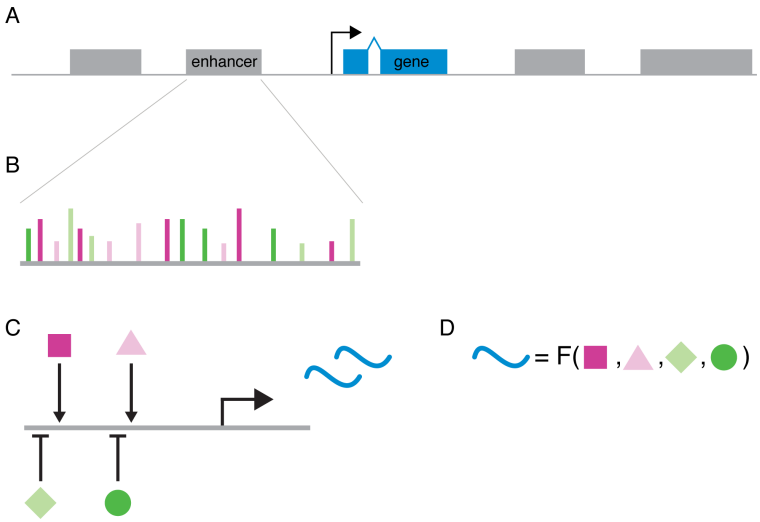


Figure 1.1 We use computational models and quantitative measurements to examine how enhancer compute.
A) A typical animal locus contains multiple enhancers. **B)** Each enhancer is composed of binding sites for transcription factors. **C)** Enhancers compute by reading the concentrations of bound transcription factors to and outputting a level of transcription. **D)** We use computational models to formalize the input/output function of an enhancer.

Drosophila embryos are an ideal model system

Our general strategy to decipher the function of regulatory DNA is to experimentally test computational models for how enhancers compute. Towards this

goal, we need a system where we can efficiently make perturbations and measure all the inputs and outputs, ideally in single cells. In addition, we need a modeling framework to formalize the input/output function and contextualize our results.

As a model system for studying how enhancers compute, I chose the *Drosophila melanogaster* blastoderm embryo. The DePace lab has an imaging infrastructure for measuring the expression of a gene in every cell of the embryo and combining data from many embryos to create an average picture called a gene expression atlas. This average embryo contains gene expression values for dozens of genes over six time points in every cell of the animal. Anterior posterior patterning in the fly embryo is control by a transcriptional network where all the genes and many regulatory connections are known. At the blastoderm stage the system is dominated by transcription factors and DNA, with a relatively smaller role for chromatin as many marks are only just becoming detectable (Chen et al., 2013). We can measure the positions and relative levels of all the transcription factors and we have an arsenal of genetic tools for making perturbations.

There is a rich history of modeling gene expression in WT embryos (Ilsley et al., 2013; Jaeger et al., 2013). By comparison efforts to validate model predictions with additional experiments have been quite limited. The primary goal of this thesis is to use quantitative measurements of perturbed embryos to test computational models of how enhancers compute cell type specific gene expression patterns. I chose the models described in Ilsley et al, as these were developed in collaboration with the DePace lab based on our cellular resolution expression data.

Engineering *trans* perturbations in the embryo

We developed shRNA methods to efficiently make *trans* perturbations in embryos (Chapter 2). Traditionally, enhancers are studied with perturbations in *cis* or in *trans*. *cis* perturbations are mutations to the enhancer sequence, usually the deletion or mutation of individual transcription factor binding sites. In contrast, *trans* perturbations remove an input transcription factor using a mutant or over-express a transcription factor. In fly embryos, *trans* perturbations are traditionally used to narrow down a list of candidate regulators and *cis* mutations used to confirm direct interactions. The primary drawback of *cis* perturbations is that the space of possible mutations is enormous and testing each one is labor intensive. The primary drawback of *trans* perturbations is they have both direct and indirect effects: in fly embryo, the segmentation network is highly interconnected, so removing one gene changes the positions and levels of many other genes. As a result, the ideal *trans* experiment is to remove one gene from the network and measure the perturbed levels of every other gene in every cell of the animal. This experiment is impossible in most animal systems, but we demonstrate feasibility in the fly blastoderm embryo in Chapter 3.

The first step towards testing computational models was efficiently making *trans* perturbations in embryos. We wanted to collect large quantities of perturbed embryos. The simplest approach was to use classic mutant alleles, but these alleles live in sick stocks, and in practice it was very difficult to collect enough embryos. In addition, I wanted to generate double and triple mutant embryos, and for genes on different chromosomes, genetic techniques are inefficient.

The next logical technique for making *trans* perturbation was RNA interference (RNAi), but when I started graduate school, genetically encoded RNAi did not work in the embryo. Fortunately, Norbert Perrimon and the Transgenic RNAi Project (TRiP) at Harvard Medical School had just developed an RNAi technique for depleting genes in the female ovary (Ni et al., 2011). RNAi hijacks host antiviral mechanisms to degrade target mRNAs (Wilson and Doudna, 2013). When a 21 bp mRNA is loaded into the RNA interference silencing complex (RISC), it will cleave any mRNAs that are perfect matches (Wilson and Doudna, 2013). There are many different ways to load the 21mer into RISC. In fly cell culture and somatic tissue, double stranded RNAs and long hairpins have been used very successfully (Mohr and Perrimon, 2012a; Mohr and Perrimon, 2012b; Mohr et al., 2014). However, long hairpins are not effective in the female germline (Ni et al., 2011). The solution discovered in the Perrimon lab was to use short hairpin RNAs (shRNAs) embedded in the backbone of an endogenous micro RNA (Ni et al., 2011). We were interested in expressing shRNAs in the ovary to deplete early zygotic transcripts, and the Perrimon lab was interested in depleting essential genes in the ovary, so we collaborated to apply this technique to key maternal and maternal/zygotic genes.

Determining the embryonic function of essential genes that are expressed both in the ovary and in the embryo is an important but technically challenging problem. The mother loads the egg with all the mRNAs and proteins necessary for the first few hours of life; the zygotic genome is largely silent until the maternal to zygotic transition (MZT) during the blastoderm stage. All embryonic genes are either maternally deposited, zygotically expressed or both. Mutants for maternal genes can be isolated from female

sterile screens and mutants for zygotic genes can be isolated from embryonic lethal screens, but for the third class, maternal/zygotic genes, mutants cannot be easily isolated because if a gene is zygotically essential, the fly dies before creating ovaries, making it difficult to study the role in the ovary. An early method for studying this class of genes was pole cell transplantation, where one takes presumptive germ cells from donor embryos and transfers them to recipients one embryo at a time(Lehmann and Nusslein-Volhard, 1987). People also created chimeras by inducing mitotic recombination with X-rays in a background with a dominant female sterile mutant *(ovoD)*, but fecundity and embryo yield were low (Perrimon et al., 1984). A major improvement came with the ability to induce mitotic recombination with FLP recombinase, which generates chimeras with relatively high efficiency but remains labor intensive (Chou and Perrimon, 1996). As a result, while embryonic lethal screens in *D.melanogaster* have been taken to saturation, when we started this work, only ~10% of the genome had been screened for a role in the ovary (Perrimon et al., 1996; Luschnig et al., 2004). shRNA depletion is an attractive alternative to these approaches. These techniques can complete the parts list of maternal/zygotic genes. These techniques can also disentangle cases when the maternal and zygotic contributions of one gene perform different functions.

shRNA depletion also has a number of other attractive features that make it a useful complement to existing genetic tools (Mohr and Perrimon, 2012b; Mohr et al., 2014). First, collecting pure populations of mutant embryos is very labor intensive for maternal sterile and zygotic mutant alleles (De Renzis et al., 2007). Second, shRNA depletion is genetically dominant, so all the eggs are affected, enabling bulk

biochemical analysis. For example, the Perrimon lab is leveraging this technique to deplete kinases and measure changes in the phosphorylation landscape of the proteome by mass-spec. The ability to collect large quantities of affected embryos enabled us to use shRNAs to build a gene expression atlas of a genetically perturbed embryo, described in Chapter 3.

Chapter 2 describes how to use shRNAs to deplete gene activities in the early fly embryo. The paper was published in *Genetics* in January 2013. It was initially designed as a "how to manual" for using shRNA depletion in the embryo, detailing when they are effective and when they are not. We conclude with a cartoon model of when shRNAs are effective and attempted some unsuccessful experiments to improve the temporal efficacy window. One section is devoted to a screen for maternal effect genes and mid embryogenesis zygotic genes performed in the Perrimon Lab and the TRiP. Our ongoing efforts to improve shRNA efficacy are described in Chapter 5.

Characterizing the effects of a *trans* perturbation

trans perturbations enable us to test models either by creating new combinations of input regulators or by moving a set of regulators to a new part of the embryo. In the first case, new combinations let us test for interactions between regulators. In the second case, moving regulators lets us test if a set of regulators is sufficient to predict the perturbed output, potentially revealing if there was a hidden, correlated regulator. Since we could most effectively deplete maternal-effect genes, we chose *bicoid (bcd),* a maternally deposited transcription factor at the top of the segmentation cascade. *bcd* activates transcription of anterior genes and represses translation of *cad,* inhibiting

posterior cell fates. Embryos laid by mother expressing shRNAs against *bcd* (*bcd* RNAi embryos) have a strong *bcd* phenotype. The resulting *bcd* RNAi atlas contains gene expression measurements for 13 genes in the segmentation network in every cell over six time points in the blastoderm embryo. It is the first 3D data set of a perturbed *Drosophila embryo,* capturing both direct and indirect effects of *bcd* depletion in a format well-suited for modeling.

Since new combinations of regulators could provide a powerful test of enhancer models, we asked if any new combination of gap gene transcription factors occurred in *bcd* RNAi embryos. We found that all the gene expression patterns changed in concert: based on a simple ON/OFF metric, there were no new combinations of gene expression patterns in the first zygotic layer of the segmentation network in *bcd* RNAi embryos. This simplified metric ignores evidence that many of these transcription factors behave in a concentration dependent manner but it was compatible with our data. Given that a combination of transcription factors is an excellent proxy for a cell fate in this system, we do not see any new cell fates in *bcd* RNAi embryos (Lehmann and Frohnhofer, 1989).

We interpret the absence of new combinations of key transcription factors our findings as strong and early canalization of cell fate. In his 1942 paper, Waddington used two lines of evidence to support his thesis that developmental systems are *canalized* (Waddington, 1942; Waddington, 1957). Each of these lines of evidence has led to a different modern definition of canalization. First, he argued that since normal individuals display low variability, wild type developmental systems must extensively buffer of environmental and genetic stress. This line of evidence has led to the definition

that developmental systems reach reproducible outcomes in the face of perturbations. Second, he observed that organisms have discrete tissues and intermediate tissues types are confined to pathologies. This idea has led to the definition that developmental systems build discrete cell fates. We focused on the second definition of canalization: that cells form discrete, recognizable fates.

It is important to distinguish canalization at the level of a gene expression pattern, and the level of individual cells. Patterns in *bcd* RNAi are not canalized: they all move and contain different numbers of cells compared to WT. Individual cells, however, tell a different story: they are canalized. We believe the perspective of the cell is more appropriate. We did not see any new combinations of transcription factors so under our definition, cell fate is canalized. The perturbed expression patterns emerge from different numbers of normal cells in the wrong places.

Waddington's 1957 essay on canalization fits perfectly with current thinking in developmental systems biology (Waddington, 1942; Waddington, 1957). He opens by using a system of differential equations as an analogy for how developmental systems evolve over time. He describes cells moving through a multi-dimensional state space (analogous to gene expression) toward attractors, their final fates. This thinking was heavily influenced by his friendship with the French mathematician, Rene Tomé (Dusa McDuff, Waddington's daughter, personal communication). Notably, the iconic cartoon of canalization, a marble rolling down the epigenetic landscape, immediately follows a figure of a phase plane that today would be more at home in complex systems than developmental biology (Figure 1.2). The modern stem cell field has lionized Waddington and the right cartoon, but largely ignored the very abstract and quantitative reasoning

he used to build the idea. High dimensional gene expression datasets and dynamical modeling have enabled a revisit of this perspective using WT variability (Manu et al., 2009a; Manu et al., 2009b; Gursky et al., 2011). Our dataset uses analysis of end point cell fates to complement this analysis and offer another perspective for marveling at the robustness of developmental canalization.

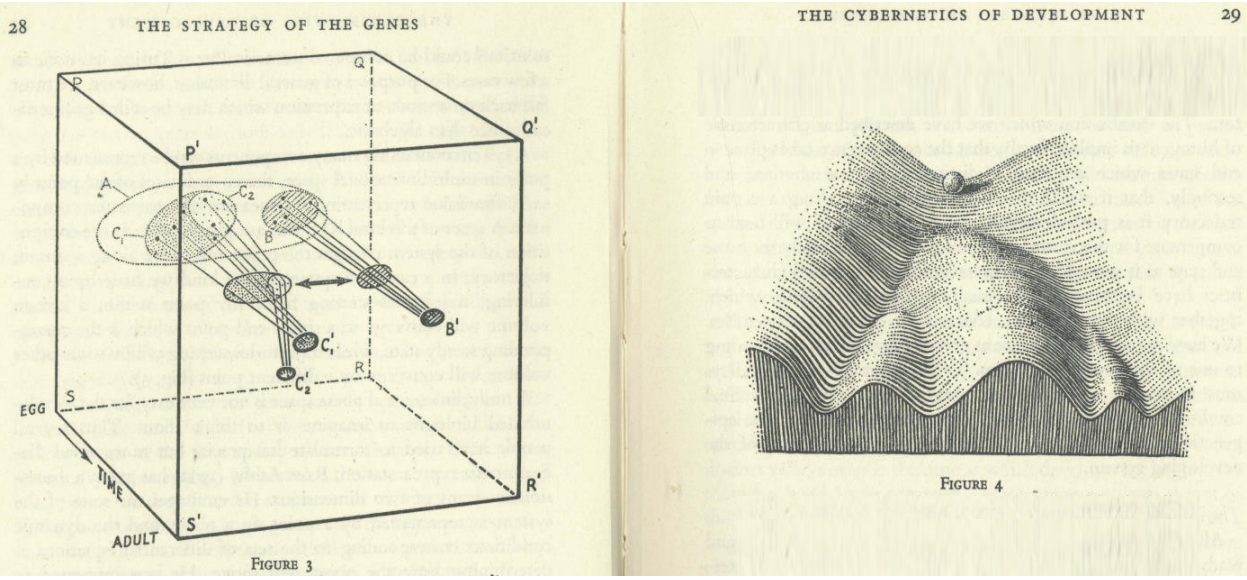


Figure 1.2: Two figures reproduced from Conrad Waddington's 1957 book.

Using trans perturbations to probe computational models of enhancer input/ output functions

After establishing a way to make *trans* changes (Chapter 2) and characterizing one in detail (Chapter 3), we arrive at our original goal in Chapter 4: using *trans* perturbations to test computational models of how enhancers compute. We focused on validating a computational model that predicted Hb both activated and repressed *even-skipped* stripes 3 and 7 (Ilsley et al., 2013). Although it is convenient to think of

transcription factors as either activators or repressors, examples of bifunctional regulators are common (Shore and Nasmyth, 1987; Struhl et al., 1992; Sauer and Jäckle, 1993; Deng et al., 2010; Di Stefano et al., 2014). *hb* is a known bifunctional transcription factor, but it was not clear if this dual regulation was important for positioning *eve* stripes 3 and 7, and we wanted to identify how the bifunctional regulation was encoded in the locus.

The ability to both repress and activate distinct targets is essential to hb's function. The early papers that genetically investigated *hb* function showed it activated some targets and repressed others, and there was evidence that it activated *Kr* at low concentrations and repressed it at high concentrations (Struhl et al., 1992; Schulz and Tautz, 1994). In *Drosophila,* maternal *hb* mRNA is uniformly deposited, but translation is repressed in the posterior by *nanos (nos)* and *pumilio (pum)*(Lawrence, 1992). This maternal *hb* is dispensable, but if it is translated in the posterior, it represses posterior fates (Lehmann and Nusslein-Volhard, 1987; Hulskamp et al., 1989; Irish et al., 1989).

In the past, other groups have shown that Hb bifunctionality improves computational models, but experimental validation has been limited. Although a model that uses Hb as a repressor can fit the WT gap gene patterns in 1D (Jaeger et al., 2004), this same model does not work in 3D (Hengenius et al., 2011). A similar model that allowed for Hb monomers and dimers to have opposite regulatory signs was able to fit gap gene expression patterns in 3D (Bieler et al., 2011). A toy model where Hb dimerized on the DNA similarly argued for bifunctionality (Papatsenko and Levine, 2008). At odds with the modeling, experiments performed on the *eve3+7* enhancer construct argued against Hb bifunctional regulation and suggested that Hb was only a

repressor (Struffi et al., 2011). It is important to note that the modeling generally used endogenous pattern while the experiments used reporters that isolated the *eve3+7* enhancer.

An important theme in Chapter 4 is that the pattern driven by an enhancer reporter (reporter pattern) and the endogenous gene expression pattern (endogenous pattern) can be different. These discrepancies have long been known but are often ignored (Barolo, 2012). We used the models to contextualize the differences between the reporter pattern and the endogenous pattern. We found that one model predicted the reporter pattern and the other model predicted the endogenous pattern. This finding implied that the endogenous pattern and the reporter pattern arose from different input/ output functions. Guided by the modeling, we confirmed this prediction with additional experiments.

Another lesson from this story is that within one locus, different computations can use distinct sets of input regulators to generate the same output. This concept runs counter to the prevailing view that most loci the endogenous expression pattern is the sum of the activities of each enhancer. Indeed, the *eve* locus has been the exemplar of this view. There are many cracks in this concept, including examples of non-additive regulatory interactions between enhancers and shadow enhancers, regulatory sequences in the same locus that drive overlapping expression patterns (Barolo, 2012). In Chapter 4, we use computational models to disentangle two overlapping computations, each localized to a different enhancer.

The *eve* locus is often invoked as the text-book example of a modular locus, but *eve* stripe 7 regulation never fit neatly into this picture. There are five early *eve* enhancers, each of which controls one or two stripes. In reporter assays, the *eve3+7* enhancer drove strong stripe 3 and 7 expression and could not be truncated in a way that separated these activities (Small et al., 1996). The *eve2* enhancer sometimes also drove some stripe 7 activity, and longer versions of this enhancer drove more stripe 7 (Goto et al., 1989; Harding et al., 1989; Small et al., 1991; Stanojevic et al., 1991; Small et al., 1992; Janssens et al., 2006). One of the original papers suggested that stripe 7 activity was distributed across a large region encompassing both the *eve2* and *eve3* enhancers (Goto et al., 1989). It is worth noting that logistic models trained on endogenous stripe 2 predict some stripe 7 expression and logistic models trained on endogenous stripe 3 also predict significant stripe 7 expression (Ilsley et al., 2013). Despite the willingness to acknowledge that *eve* stripe 7 regulation was distributed, no one seriously thought that there might be two different input/out functions generating stripe 7 with different regulatory logic.

We found that Hb does both repress and activate eve stripe 7, but each activity is isolated to a separate enhancer. The *eve3+7* enhancer is repressed by Hb. An expanded version of the *eve2* enhancer which we call *eve2+7* is activated by Hb. This interpretation synthesizes and explains many seemingly contradictory results about stripe 7; while the textbook pictures show an *eve3+7* enhancer and *eve2* enhancer, experts are quick to admit the shortcomings of this cartoon (Alberts et al. 2002) (Mike Levine, personal communication; Stephen Small, personal communication). We argue that these two enhancers should be reclassified as shadow enhancers for *eve* stripe 7

regulation. We think that separating activation and repression might be a general strategy for encoding bifunctional regulation in the genome. This findings is a novel form of regulatory redundancy: two enhancers use two distinct sets of input regulators to set the same output pattern. Binding site analysis has proposed that shadow enhancers might use different regulations (Kazemian et al., 2010), but it had never been experimentally shown. The modeling revealed that multiple computations can have similar outputs. We are continuing to investigate this mechanism as elaborated in Chapter 6.

This thesis contains the full arch of our efforts to use *trans* perturbations to test computational models of enhancer input/output functions. In Chapter 2, we develop and deploy shRNAs to deplete transcripts in the early embryo. In Chapter 3, we built a gene expression atlas of a *bcd* depleted embryo and found early canalization of cell fate. In Chapter 4, we use this *bcd* RNAi gene expression atlas to test computational models of *eve* stripe 7 regulation, revealing novel regulatory redundancy in classically modular locus. Models let us study regulatory DNA by formalizing assumptions and generating predictions. We use perturbation experiments and quantitative measurements to test these model predictions and, indirectly, the validity of the underlying assumptions. While using shRNAs to perturb the embryo, we found strong and immediate canalization of cell fate. Although canalization constraints the nature of shRNA perturbations, we show this kind of perturbation is still useful for testing models. Our story illustrates how adding numbers to the arrows in the cartoon model of *eve* stripe 7 regulation revealed missing arrows. Unexpectedly, this locus uses two distinct parallel computations to position one

part of the gene expression pattern. Formalizing cartoon models with computational models can reveal new biology.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). Chapter 21. Development of Multicellular Organisms. In *Molecular Biology of the Cell* New York: Garland Science.

Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* **34**, 135-141.

Bieler, J., Pozzorini, C. and Naef, F. (2011). Whole-Embryo Modeling of Early Segmentation in Drosophila Identifies Robust and Fragile Expression Domains. *Biophysj* **101**, 287-296.

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C. and Zeitlinger, J. (2013). A global change in RNA polymerase II pausing during the Drosophila midblastula transition. *eLife* **2**, e00861-e00861.

Chou, T. B. and Perrimon, N. (1996). The autosomal FLP-DFS technique for generating germline mosaics in Drosophila melanogaster. *Genetics* **144**, 1673-1679.

Davidson, E. H. (2010). The regulatory genome: gene regulatory networks in development and evolution.

De Renzis, S., Elemento, O., Tavazoie, S. and Wieschaus, E. F. (2007). Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. *PLoS biology* **5**, e117.

Deng, Z., Cao, P., Wan, M. M. and Sui, G. (2010). Yin Yang 1: a multifaceted protein beyond a transcription factor. *Transcription* **1**, 81-84.

Di Stefano, B., Sardina, J. L., van Oevelen, C., Collombet, S., Kallin, E. M., Vicent, G. P., Lu, J., Thieffry, D., Beato, M. and Graf, T. (2014). C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506**, 235-239.

Goto, T., Macdonald, P. and Maniatis, T. (1989). Early and late periodic patterns of *even skipped* expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413-422.

Gunawardena, J. (2014). Models in biology: 'accurate descriptions of our pathetic thinking'. *BMC Biology*

Gursky, V. V., Panok, L., Myasnikova, E. M., Manu, Samsonova, M. G., Reinitz, J. and Samsonov, A. M. (2011). Mechanisms of gap gene expression canalization in the Drosophila blastoderm. *BMC Systems Biology* **5**, 118.

Harding, K., Hoey, T., Warrior, R. and Levine, M. S. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. *The EMBO journal* **8**, 1205-1212.

Hengenius, J. B., Gribskov, M., Rundell, A. E., Fowlkes, C. C. and Umulis, D. M. (2011). Analysis of Gap Gene Regulation in a 3D Organism-Scale Model of the Drosophila melanogaster Embryo. *PLoS ONE* **6**, e26797.

Hulskamp, M., Schroder, C., Pfeifle, C., Jackle, H. and Tautz, D. (1989). Posterior segmentation of the Drosophila embryo in the absence of a maternal posterior organizer gene. *Nature* **338**, 629-632.

Ilsley, G. R., Fisher, J., Apweiler, R., DePace, A. H. and Luscombe, N. M. (2013). Cellular resolution models for even skipped regulation in the entire Drosophila embryo. *eLife* **2**, e00522.

Irish, V., Lehmann, R. and Akam, M. (1989). The Drosophila posterior-group gene nanos functions by repressing hunchback activity. *Nature* **338**, 646-648.

Jaeger, J., Manu and Reinitz, J. (2013). Drosophila blastoderm patterning. *Current opinion in genetics & development*

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Vanario-Alonso, C. E., Samsonova, M. et al. (2004). Dynamic control of positional information in the early Drosophila embryo. *Nature* **430**, 368-371.

Janssens, H., Hou, S., Jaeger, J., Kim, A.-R., Myasnikova, E., Sharp, D. and Reinitz, J. (2006). Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. *Nature genetics* **38**, 1159-1165.

Kaplan, S., Bren, A., Zaslaver, A., Dekel, E. and Alon, U. (2008). Diverse Two-Dimensional Input Functions Control Bacterial Sugar Genes. *Molecular Cell* **29**, 786-792.

Kazemian, M., Blatti, C., Richards, A., McCutchan, M., Wakabayashi-Ito, N., Hammonds, A. S., Celniker, S. E., Kumar, S., Wolfe, S. A., Brodsky, M. H. et al. (2010). Quantitative analysis of the Drosophila segmentation regulatory network using pattern generating potentials. *PLoS biology* **8**,

Lawrence, P. A. (1992). The making of a fly: the genetics of animal design. Blackwell Scientific Publications, London.

Lehmann, R. and Frohnhofer, H. G. (1989). Segmental polarity and identity in the abdomen of Drosophila is controlled by the relative position of gap gene expression. *Development* **107 Suppl**, 21-29.

Lehmann, R. and Nusslein-Volhard, C. (1987). hunchback, a gene required for segmentation of an anterior and posterior region of the Drosophila embryo. *Dev Biol* **119**, 402-417.

Levine, M., Cattoglio, C. and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13-25.

Luschnig, S., Moussian, B., Krauss, J., Desjeux, I., Perkovic, J. and Nüsslein-Volhard, C. (2004). An F1 genetic screen for maternal-effect mutations affecting embryonic pattern formation in Drosophila melanogaster. *Genetics* **167**, 325-342.

Manu, Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M. et al. (2009). Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. *PLoS Computational Biology* **5**, e1000303.

Manu, Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M. et al. (2009). Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. *PLoS biology* **7**, e1000049.

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J. et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195.

Mayo, A. E., Setty, Y., Shavit, S., Zaslaver, A. and Alon, U. (2006). Plasticity of the cis-regulatory input function of a gene. *PLoS Biol* **4**, e45.
Mohr, S. E. and Perrimon, N. (2012). RNAi screening: new approaches, understandings, and organisms. *Wiley Interdiscip Rev RNA* **3**, 145-158.
Mohr, S. E., Hu, Y., Kim, K., Housden, B. E. and Perrimon, N. (2014). Resources for Functional Genomics Studies in Drosophila melanogaster. *Genetics* **197**, 1-18.
Mohr, S. E. and Perrimon, N. (2012). RNAi screening: new approaches, understandings, and organisms. *Wiley interdisciplinary reviews. RNA* **3**, 145-158.
Ni, J. Q., Zhou, R., Czech, B., Liu, L. P., Holderbaum, L., Yang-Zhou, D., Shim, H. S., Tao, R., Handler, D., Karpowicz, P. et al. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat Methods* **8**, 405-407.
Papatsenko, D. and Levine, M. S. (2008). Dual regulation by the Hunchback gradient in the Drosophila embryo. *Proceedings of the National Academy of Sciences* **105**, 2901-2906.
Perrimon, N., Engstrom, L. and Mahowald, A. P. (1984). The effects of zygotic lethal mutations on female germ-line functions in Drosophila. *Dev Biol* **105**, 404-414.
Perrimon, N., Lanjuin, A., Arnold, C. and Noll, E. (1996). Zygotic lethal mutations with maternal effect phenotypes in Drosophila melanogaster. II. Loci on the second and third chromosomes identified by P-element-induced mutations. *Genetics* **144**, 1681-1692.
Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. and Elowitz, M. B. (2005). Gene regulation at the single-cell level. *Science* **307**, 1962-1965.
Sauer, F. and Jäckle, H. (1993). Dimerization and the control of transcription by Krüppel. *Nature* **364**, 454-457.
Schulz, C. and Tautz, D. (1994). Autonomous concentration-dependent activation and repression of Kruppel by hunchback in the Drosophila embryo. *Development* **120**, 3043-3049.
Setty, Y., Mayo, A. E., Surette, M. G. and Alon, U. (2003). Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A* **100**, 7702-7707.
Sharp, D. H. and Reinitz, J. (1998). Prediction of mutant expression patterns using gene circuits. *Bio Systems* **47**, 79-90.
Shore, D. and Nasmyth, K. (1987). Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell* **51**, 721-732.
Small, S., Blair, A. and Levine, M. S. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. *The EMBO journal* **11**, 4047-4057.
Small, S., Blair, A. and Levine, M. S. (1996). Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. *Developmental biology* **175**, 314-324.
Small, S., Kraut, R., Hoey, T., Warrior, R. and Levine, M. S. (1991). Transcriptional regulation of a pair-rule stripe in Drosophila. *Genes & Development* **5**, 827-839.
Stanojevic, D., Small, S. and Levine, M. S. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science* **254**, 1385-1387.
Struffi, P., Corado, M., Kaplan, L., Yu, D., Rushlow, C. and Small, S. (2011). Combinatorial activation and concentration-dependent repression of the Drosophila

even skipped stripe 3+7 enhancer. *Development (Cambridge, England)* **138**, 4291-4299.

Struhl, G., Johnston, P. and Lawrence, P. A. (1992). Control of Drosophila body pattern by the hunchback morphogen gradient. *Cell* **69**, 237-249.

Waddington, C. H. (1942). Canalization of Development and the inheritance of acquired characters. *Nature* 563-565.

Waddington, C. H. (1957). The strategy of the genes; a discussion of some aspects of theoretical biology. **Allen & Unwin**, London.

Wilson, R. C. and Doudna, J. A. (2013). Molecular mechanisms of RNA interference. *Annu Rev Biophys* **42**, 217-239.

Chapter 2: Depleting Gene Activities in Early *Drosophila* Embryos with the "Maternal-Gal4-shRNA" System

Max V. Staller*, Dong Yan§, Sakara Randklev§,, Meghan D. Bragdon*, Zeba B. Wunderlich*, Rong Tao§, Lizabeth A. Perkins§ , Angela H. DePace* and Norbert Perrimon§,**,1**

*Department of Systems Biology, Harvard Medical School

§Department of Genetics, Harvard Medical School

**Howard Hughes Medical Institute

Boston, Massachusetts 02115, USA.

Running Title: Maternal-Gal4 – shRNA method

Key Words: RNAi, Gal4-UAS, shRNAs, *Drosophila,* embryogenesis

Originally published as

Staller, M.V. et al., 2012. Depleting Gene Activities in Early Drosophila Embryos with the "maternal-Gal4 - shRNA" System. Genetics.

Abstract

In a developing *Drosophila melanogaster* embryo, mRNAs have a maternal origin, a zygotic origin or both. During the maternal-zygotic transition, maternal products are degraded and gene expression comes under the control of the zygotic genome. To interrogate the function of mRNAs that are both maternally and zygotically expressed, it is common to examine the embryonic phenotypes derived from female germline mosaics. Recently, the development of RNAi vectors based on short hairpin RNAs (shRNAs) effective during oogenesis has provided an alternative to producing germline clones. Here, we evaluate the efficacies of: 1. maternally-loaded shRNAs to knockdown zygotic transcripts; and 2. maternally-loaded Gal4 protein to drive zygotic shRNA expression. We show that, while Gal4-driven shRNAs in the female germline very effectively generate phenotypes for genes expressed maternally, maternally loaded shRNAs are not very effective at generating phenotypes for early zygotic genes. However, maternally loaded Gal4 protein is very efficient at generating phenotypes for zygotic genes expressed during mid-embryogenesis. We apply this powerful and simple method to unravel the embryonic functions of a number of pleiotropic genes.

Introduction

During *Drosophila* oogenesis, the mother loads the oocyte with the RNAs and proteins necessary to support embryonic development until zygotic transcription begins approximately two hours after fertilization. Based on their expression patterns, three classes of genes can be distinguished: maternally expressed genes (referred to as "Mat"), zygotically expressed genes (referred to as "Zyg"), and genes expressed both maternally and zygotically (referred to as "Mat&Zyg") (for overview see Lawrence, 1992). Characterization of the roles of Mat genes during embryonic development has classically been performed following the examination of the phenotypes of embryos laid by females carrying homozygous viable female sterile mutations. Examples of Mat genes include those that establish the antero-posterior (*bicoid (bcd), nanos (nos), torso (tor)*) and dorsal-ventral (*dorsal (dl))* axes (Lawrence, 1992). Zyg genes have been identified among mutations associated with embryonic lethality, including those that interpret the maternally encoded positional information, such as gap (e.g., *giant (gt), Krüppel (Kr), knirps (kni)*), pair-rule (e.g., *fushi-tarazu (ftz), even-skipped (eve), odd-skipped (odd)*) and segment polarity (e.g., *engrailed (en), wingless (wg), hedgehog (hh)*) genes (see review by St. Johnston and Nusslein-Volhard, 1992).

While many Mat and Zyg genes have been well characterized, the contributions of Mat&Zyg essential genes to embryonic development have yet to be fully described. Examining the null embryonic phenotypes of Mat&Zyg essential genes is technically challenging because embryos need to be derived from mutant germlines; i.e., the functions cannot be examined from heterozygous mothers as the maternal contribution, in most cases, masks their early zygotic functions, and homozygous mutant females

cannot be recovered as they are dead. A solution to this problem has been the creation of germline mosaics whereby eggs are collected from females with mutant homozygous germlines in an otherwise wild type soma. The most commonly used method to produce female germline mosaics is the *FLP-FRT ovoD* germline clone (GLC) technique (Chou and Perrimon, 1996). Using this strategy, FLP-FRT mediated mitotic recombination in an *ovoD* dominant female sterile background generates homozygous germline clones for candidate Mat&Zyg mutations in otherwise somatically heterozygous mutant females.

An example of a Mat&Zyg gene that yields diverse phenotypes when it is depleted at different stages of development is the *D-Raf* serine-threonine kinase (Perrimon et al., 1985; Ambrosio et al., 1989; see review by Duffy and Perrimon, 1994). *D-raf* mutant offspring derived from heterozygous females die during larval-pupal development. However, embryos derived from *D-raf* mutant GLCs exhibit two classes of phenotypes: embryos that receive a wild type (WT) paternal copy display a "terminal class" phenotype, with the acron and telson missing, because maternally-derived D-raf gene product acts downstream of maternally-derived Torso gene product, a receptor tyrosine kinase (RTK) that activates the Zyg genes *tailless (tll)* and *huckebein* (*hkb)*. Embryos that do not receive a paternal copy show poor cuticle development, reflecting the role of D-raf downstream of another RTK, the Epidermal Growth Factor Receptor (EGFR), which is required for proper epidermal differentiation. While the EGFR phenotype can be paternally rescued, the terminal phenotype cannot, reflecting the early activity of Torso signaling and the later function of EGFR signaling. The *D-raf* example illustrates how different embryonic phenotypes can be observed depending on

the level of either maternal or zygotic gene activity present at a specific developmental stage.

Recently, we established an alternative approach to GLCs based on RNA interference (RNAi). We generated vectors employing short hairpin RNAs (shRNAs) which, when expressed during oogenesis using a maternal Gal4 driver, reproduced the phenotypes of Mat, Zyg, and Mat&Zyg genes (Ni et al., 2011). Using RNAi to study early embryonic phenotypes is an attractive strategy as it requires fewer and simpler crosses than the *FLP-FRT ovoD* method. Moreover, the easy production of *maternal-Gal4>>UAS-shRNA* females facilitates large-scale screening and the generation of large numbers of mutant embryos that can be used for phenotypic and biochemical analyses.

Extending the *maternal-Gal4>>UAS-shRNA* technique from oogenesis into early embryonic development is complicated by the maternal-zygotic transition (MZT), a period when maternal mRNAs are degraded and gene expression comes under the control of the zygotic genome (Tadros and Lipshitz, 2009). This constraint led us to evaluate in detail the use of the Gal4-UAS system to drive shRNA expression in early embryos. Specifically, we determined whether maternal loading of shRNAs into embryos could deplete zygotic RNAs and to what extent maternally provided Gal4 could be used to express zygotic shRNAs at sufficient levels to generate mutant phenotypes (Figure 2.1). Our results indicate that while Gal4-driven shRNAs in the female germline targeting maternal transcripts are extremely effective at generating phenotypes consistent with strong knockdown, maternally loaded shRNAs targeting zygotic transcripts are not very effective at yielding phenotypes. However, maternally loaded

Gal4 protein is very efficient at activating zygotic *UAS-shRNA* constructs and generating phenotypes for genes expressed during mid embryogenesis. We illustrate these features of the "maternal-Gal4 - shRNA" system and apply the method to the identification of a number of new zygotic lethal loci with specific maternal effect phenotypes.

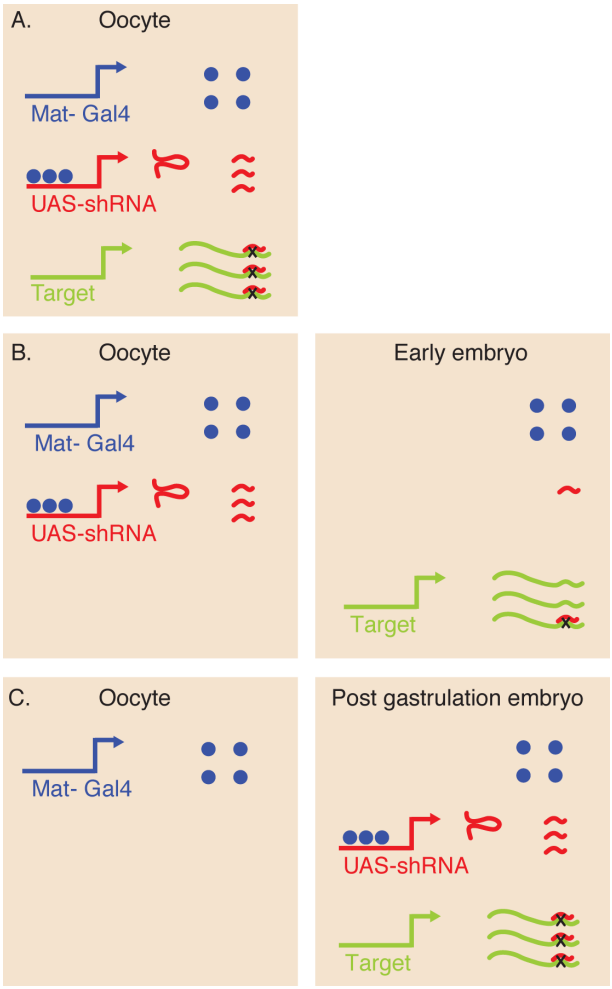


FIGURE 2.1: Strategies for knockdown of maternal and zygotic transcripts. A) Depletion of a maternal transcript following expression of shRNAs in the female germline. The maternal Gal4 driver (blue) activates shRNAs (red), which deplete target transcripts (green). B) Depletion of a zygotic transcript by loading the embryo with maternally derived shRNAs. C) Depletion of a zygotic transcript following zygotic activation of shRNAs by maternally loaded Gal4 protein. Strategies A and B correspond to F2 phenotypes in Table 1 while strategy C corresponds to an F1 phenotype.

Materials and Methods

***Drosophila* strains:** Two different maternal Gal4 drivers (*maternal-Gal4*) were used: (1) Maternal Triple Driver Gal4 (*MTD-Gal4)*: *(P(otu-Gal4::VP16.R)1, w[*]; P(Gal4-nos.NGT)40; P(Gal4::VP16-nos.UTR)CG6325[MVD1])*, described in Petrella et al. (2007) (Bloomington Stock 31777), a gift from L. Cooley. These flies are homozygous for three Gal4 transgenes that together drive expression through all of oogenesis. *P(otu-Gal4::VP16.R)* contains the *ovarian tumor (otu)* promoter and *fs(1)K10* 3' untranslated region (UTR) and drives strong expression beginning in stage 1 egg chambers. *P(Gal4-nos.NGT)* contains the *nanos (nos)* promoter and 3' UTR, driving expression throughout the germarium. *P(Gal4::VP16-nos.UTR)* contains the *nos* promoter and *αTubulin84E* 3′ UTR and drives expression through oogenesis. (2) Maternal-Tubulin-Gal4 (*mat-tub-Gal4*) driver: *y w; P(mat-mat-tub-Gal4)mat67; P(mat-mat-tub-Gal4)mat15* (line 2318) is from D. St. Johnston and F. Wirtz-Peitz. This line is homozygous for two insertions of a construct containing the maternal tubulin promoter from *αTub67C* and the 3' UTR from *αTub84B*. The difference between *MTD-Gal4* and *mat-tub-Gal4* driver lines is that mat-*tub-Gal4* does not drive expression during early oogenesis in the germarium. This difference is useful, as in some cases early oogenesis defects that can be detected with *MTD-Gal4* can be bypassed using *mat-tub-Gal4*, thus allowing the production of eggs (Yan et al., in preparation). Timing aside, the two drivers led to similar embryonic phenotypes and were used interchangeably in this study. Finally, all mutant alleles were obtained from the Bloomington *Drosophila* stock center: *hb[12]* (#1755), *kni[1]* (#1783), *Kr[2]* (#1601), *eve[1]* (#1599), *hkb[2]* (#5457), *twi[1]* (#2381), *fkh[6]* (#545), *en[7]* (#1820), *ftz[11]* (#1841), *hh[21]* (#5338), *sna[1]* (#25127), and *wg[l-17]* (#2980).

The two UAS-shRNA vectors used in this screen are described in Ni et al. (2011). VALIUM20 is effective in both the soma and female germline, and VALIUM22 is more potent in the female germline and less efficient in the soma. The constructs used in this study were generated at the Transgenic RNAi Project (TRiP) at Harvard Medical School and integrated into the genome at either the attP2 (chromosome III) or attP40 (chromosome II) landing sites, as previously described (Ni et al., 2011). Details on the lines used in this study can be found in Table 1 and on the TRiP web site (http://www.flyrnai.org).

Testing for embryonic phenotypes: To determine F1 phenotypes, ~10 *maternal-GAL4* females were crossed with ~5 *UAS-shRNA* homozygous or heterozygous males and embryos collected at 27°. For the F2 phenotype analyses, *maternal-GAL4>>UAS-shRNA* females were recovered from the previous cross and mated to either their siblings or *UAS-shRNA* homozygous males. In the few cases where F1 crosses failed to give progeny (see Table 1), *maternal-GAL4>>UAS-shRNA* flies were generated by crossing *maternal-GAL4* males with *UAS-shRNA* females. Note that all crosses were performed at 27° as Gal4 is more potent at higher temperatures. We avoided testing the flies at 29° because of some male sterility issues at this temperature.

The percentage of embryos hatching was determined by lining up approximately 200 0-24 hr embryos and counting the dead (brown) and hatched eggs after at least 24 hrs. When lethality was observed, cuticles were prepared to examine patterning defects. Unhatched cuticles were prepared and mounted in Hoyer's mounting media. Where

noted, a Z-stack of 3-6 dark field images was acquired and computationally flattened using the Helicon Focus software (HeliconSoft).

Design of new scaffold shRNA vectors: A number of stable maternally deposited mRNAs have been identified by Votruba (2009). Hairpin pre-miRNA sequences for miR-275 and miR-92a were downloaded from miRbase (Kozomara and Griffiths-Jones, 2011), the shRNAs where inserted into the 21bp that normally become the mature miRNA, and the complementary portion of the hairpin made into a perfect match. All oligos used are listed in Appendix A Table S1. Note that all of the pre-miRNA hairpin sequence is included in the oligos and that no other changes were made to the VALIUM20 backbone. Complementary oligos were annealed and cloned into the *Nhe*I and *Eco*RI sites of VALIUM20 and injected into the attP2 landing site. Injections were performed by Genetic Services, Inc. (GSI) (http://www.geneticservices.com).

***in situ* hybridization:** Embryos from *MTD-GAL4>>UAS-shRNA-hb* mothers were collected for 8 hrs, fixed in heptane and formaldehyde for 25 mins and stained with dinitrophenol (DNP) probes against *hb,* and fluorescently detected by horseradish peroxidase/tyramide deposition of Cy3 (Perkin Elmer) as described in Fowlkes et al. (2011). Images were acquired by laser scanning microscopy with 2-photon excitation at 750 nm (Luengo Hendriks et al., 2006). Briefly, the sytox green nuclear stain was used to automatically identify nuclei and the Cy3 signal in each nucleus was quantified (Luengo Hendriks et al., 2006). Analysis of expression domain boundaries was performed in MatLab (Mathworks) using the PointCloud toolbox from the Berkeley *Drosophila* Transcription Network Project (BDTNP, http://bdtnp.lbl.gov). Embryo length

was normalized and expression boundaries were detected by finding the inflection point in the pattern. WT data was downloaded from www.bdtnp.lbl.gov (Fowlkes et al. 2008).

Results and Discussion

shRNAs expressed in the female germline effectively knockdown Mat genes

To extend our previous finding that shRNAs expressed during oogenesis effectively knockdown maternally deposited transcripts, we tested a number of UAS-shRNA lines targeting various Mat genes. shRNA lines were produced against *bcd, tor, nos* and *dl,* and all exhibited embryonic phenotypes commensurate with strong mutant alleles (Table 2.1, Figure 2.2). These data suggest that shRNAs driven by mat-GAL4 are very effective at depleting the relevant transcripts in the female germline.

Maternally loaded shRNAs are not very effective at knocking down early acting Zyg genes

Next, we tested whether maternally loading shRNAs was efficient at knocking down Zyg genes. We generated UAS-shRNA lines against 30 of the earliest known zygotic genes that are not expressed during oogenesis (Table 2.1). Embryos derived from *maternal-Gal4>>UAS-shRNAs* females crossed to sibling males heterozygous for *UAS-shRNA* were examined for embryonic phenotypes. Strikingly, only the shRNA line that targeted *decapentaplegic* (*dpp*) showed embryonic lethality, with 100% of the F2 embryos exhibiting a ventralized phenotype (Table 2.1, *dpp*-F2 phenotype in Figure 2.2). Although we cannot be certain that all the UAS-shRNA lines are effective at knocking down the targeted transcripts, these results indicate that most shRNAs delivered from the mother to the embryo do not sufficiently deplete early zygotic transcripts to generate embryonic phenotypes detectable in the cuticle. Regardless, the

phenotype for zygotic *dpp* transcripts with maternal shRNAs indicates that maternally derived shRNAs can work (see also below results from *hunchback (hb))*.

Our ability to detect a cuticle phenotype for *dpp* most likely reflects the haplo-insufficiency associated with this gene (Spencer et al., 1982) that renders it more sensitive to knockdown. Importantly, depletion of *dpp* suggested the possibility that some of our shRNA constructs were ineffective not because the hairpin did not work, but because an insufficient amount of maternally-derived shRNA was present in early embryos. Thus, we tested whether reducing by half the amount of zygotic gene product in embryos derived from *maternal-Gal4>>UAS-shRNA* females could reveal phenotypes. Crossing *maternal-Gal4>>UAS-shRNA* females to mutant heterozygous males created embryos with the same amount of maternally deposited shRNA but (presumably) half the number of zygotic transcripts for the targeted gene. We looked for phenotypes in sensitized backgrounds for the following genes: *Kr, kni,* (gap); *hkb, forkhead (fkh)* (terminal); *eve, ftz* (pair-rule); *twist (twi), snail (sna)* (dorsal-ventral); *wg, hh,* and *en* (segment polarity), and were able to detect clear phenotypes for shRNAs targeting *Kr* and *twi*. In the case of *twi,* ~50% of the embryos showed the expected twisted phenotype (Figure 2.2). For *Kr,* 25% of the embryos showed a mild gap segmentation phenotype detectable by the absence of the second abdominal segment (A2) (Figure 2.2). Similarly, for *ftz* we observed ~30% lethality and a mild phenotype where one thoracic segment was missing. In addition, for the segment polarity genes *hh* and *wg*, we found rare embryos with cuticle defects similar to those of classic mutant alleles (Figure 2.2). Altogether, these results indicate that maternally loaded shRNAs

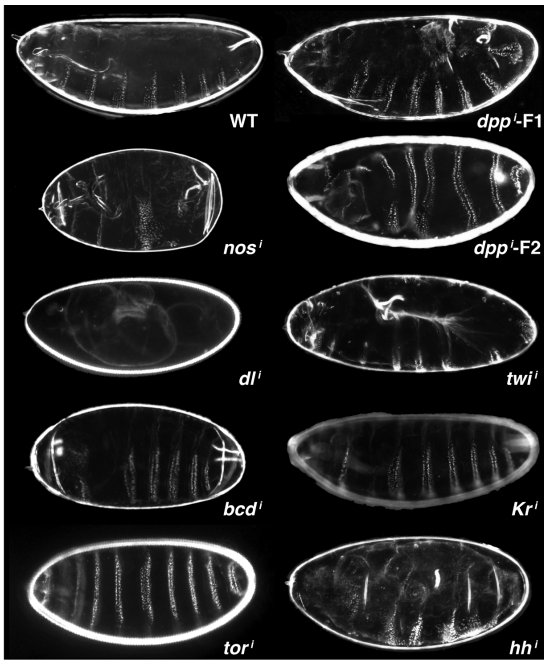


FIGURE 2.2: Embryonic phenotypes associated with knockdown of Mat and Zyg genes.
For *nos, dl, bcd, tor,* and *dpp-F2*, *mat-tub-Gal4>>UAS-shRNA* mothers were crossed to *UAS-shRNA* males. All phenotypes resemble strong classic alleles. *dpp-F1* embryos were obtained from crossing *mat-tub-Gal4* females to *UAS-shRNA* males. For *Kr, twi, hh, ftz and wg*, *mat-tub-Gal4>>UAS-shRNA* mothers were crossed to males heterozygous for a strong mutant allele of the target gene. Only a subset (see text) of embryos from these crosses had cuticle phenotypes. The phenotypes for *twi*, *hh,* and *wg* resemble classic mutants. For *Kr,* the main defect is the absence of the A2 segment (arrowhead), which is a smaller gap than seen in classic mutant embryos. The same phenotype was observed with two shRNA lines (GL01322 and GL01324). For *ftz,* the embryos are missing two anterior segments, a weaker phenotype than is seen in classic mutant embryos. Description of the mutant phenotypes and references for each gene tested can be found at http://flybase.org/. WT refers to a wild type cuticle. The "i" superscript refers to the RNAi-induced phenotypes.

for early zygotically expressed genes are more efficient in a sensitized heterozygous mutant background.

New shRNA backbones for depletion of early zygotic transcripts

The shRNA sequences in VALIUM20 are embedded in the *miR-1* backbone that is not expressed during oogenesis and early embryos (Ruby et al., 2007). To test

whether shRNAs would be more effective when expressed in the backbone of a miRNA normally expressed during late oogenesis and embryogenesis, we generated transgenic lines targeting the *otu, Notch (N), bcd, Kr, gt, wg,* and *arm* genes in the backbone of *miR-275* and *miR-92a*, as both had been shown previously to be some of the most stable miRNAs present in unfertilized embryos (Votruba, 2009). Although shRNAs targeting *bcd*, *otu, N* and *arm* generated phenotypes comparable to the original lines in the *miR-1* design (Table 2.1), shRNAs against *Kr, gt,* and *wg* did not. Thus, backbones of miRNAs expressed or not during oogenesis do not appear to make a significant difference. Further studies that quantify the respective amounts of shRNAs produced with the various designs and that determine the stability of the shRNAs will be needed to evaluate whether the system can be improved further.

Maternally loaded Gal4 protein can trigger zygotic expression of shRNAs

To our surprise, maternally deposited Gal4 protein can activate zygotic expression of *UAS-shRNAs* early enough and strongly enough to generate cuticle phenotypes. We observed significant F1 lethality (60%) in embryos derived from crossing *mat-tub-Gal4* females with *UAS-shRNA-dpp* homozygous males (*dpp*-F1 phenotype in Figure 2.2). These embryos showed variable germ band retraction and head defect phenotypes reminiscent of weak *dpp* alleles, (Spencer et al., 1982; Irish and Gelbart, 1987). In addition, a number of shRNAs targeting other genes also led to F1 embryonic lethality and in some cases cuticle phenotypes (see “F1 Phenotype” column in Table 2.1, Figure 2.3). Two striking examples are *armadillo* (*arm,* the *D. melanogaster* beta-Catenin homolog) and *N*. All embryos derived from *mat-tub-Gal4*

females crossed to *UAS-shRNA-arm*, but not from the reciprocal cross, showed the stereotypical segment polarity phenotype reflecting the role of beta-Catenin in Wg signaling (Peifer et al., 1991) (Figure 2.3). Similarly, most F1 embryos (95%) from *mat-tub-Gal4* females crossed to *UAS-shRNA-N* (line HMS0009), but not from the reverse cross, showed a neurogenic phenotype (Figure 2.3). Note that the VALIUM22 line against *N* (GL00092) showed lower F1 lethality (10%), most likely reflecting the difference between the VALIUM20 and VALIUM22 expression vectors (Ni et al., 2011; Materials and Methods). Interestingly, crossing *mat-tub-Gal4>>UAS-shRNA-N* females to sibling males resulted in 75% neurogenic embryos, with the remaining quarter of the progeny surviving. This fraction is consistent with the quarter of embryos without a *UAS-shRNA-N* transgene surviving, and reminiscent of the previously reported paternal rescue of the Notch maternal effect phenotype (Lehmann et al., 1981). Together, these data suggest how, for genes expressed after gastrulation, maternal Gal4 can activate zygotically delivered shRNAs to strongly deplete target transcripts.

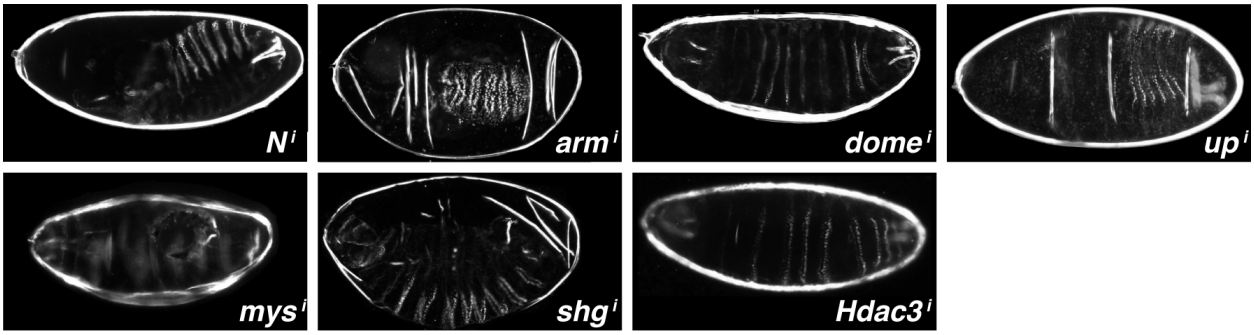


FIGURE 2.3: Zygotic phenotypes revealed by the expression of zygotic shRNAs by maternally loaded Gal4 protein.
For some genes, high rates of F1 lethality and specific embryonic phenotypes were detected when *maternal-Gal4* females were crossed to *UAS-shRNA* males. These included *armadillo (arm), Notch (N), domeless (dome), shotgun (shg), myospheroid (mys), upheld (up),* and *Histone deacetylase 3 (Hdac3).* Additional shRNA lines associated with F1 phenotypes are listed in Table 1.

Varying *UAS-shRNA* copy number to reveal different discrete phenotypes

The ability of maternal Gal4 to activate shRNAs in both the germline and the zygote has implications for detecting and interpreting embryonic phenotypes associated with the knockdown of Mat&Zyg genes. An instructive example is the case of *rolled* (*rl*), the *Drosophila* MAPK/ERK serine/threonine kinase that acts downstream of RTKs such as Tor and EGFR. Previous studies have shown that these RTKs activate a sequential signaling cascade of the D-Raf, D-MEK, and MAPK/Rl kinases (Duffy and Perrimon, 1994; Li, 2005). However, while the roles of D-Raf and D-MEK in Tor signaling have been well characterized by the analysis of their GLC phenotypes (Duffy and Perrimon, 1994; Li, 2005), Rl has only been implicated in Tor signaling by the ability of a *rl* loss-of-function mutation to suppress a gain-of-function Tor mutation (Brunner et al., 1994). Strikingly, different classes of embryonic cuticles are observed depending on the genotypes of the males that are crossed to *MTD-Gal4>>UAS-shRNA-rl* females. If we crossed *MTD-Gal4>>UAS-shRNA-rl* females to WT males, 50% of the embryos showed terminal defects (the torso "terminal class" phenotype) (Figure 2.4A1), while the other half showed poor cuticle development (the EGFR mutant phenotype) (Figure 2.4A2). On the other hand, 100% of the embryos derived from *MTD-Gal4>>UAS-shRNA-rl* females crossed to *UAS-shRNA-rl* homozygous males showed poor cuticle development, similar to those shown in Figure 2.4A2. These distributions indicate that the presence of zygotic *UAS-shRNA-rl* influences the phenotype of embryos derived from *MTD-Gal4>>UAS-shRNA-rl* females. Embryos with either one or two copies of the *UAS-shRNA-rl* transgene show poor cuticle development reflecting a role of Rl in EGFR

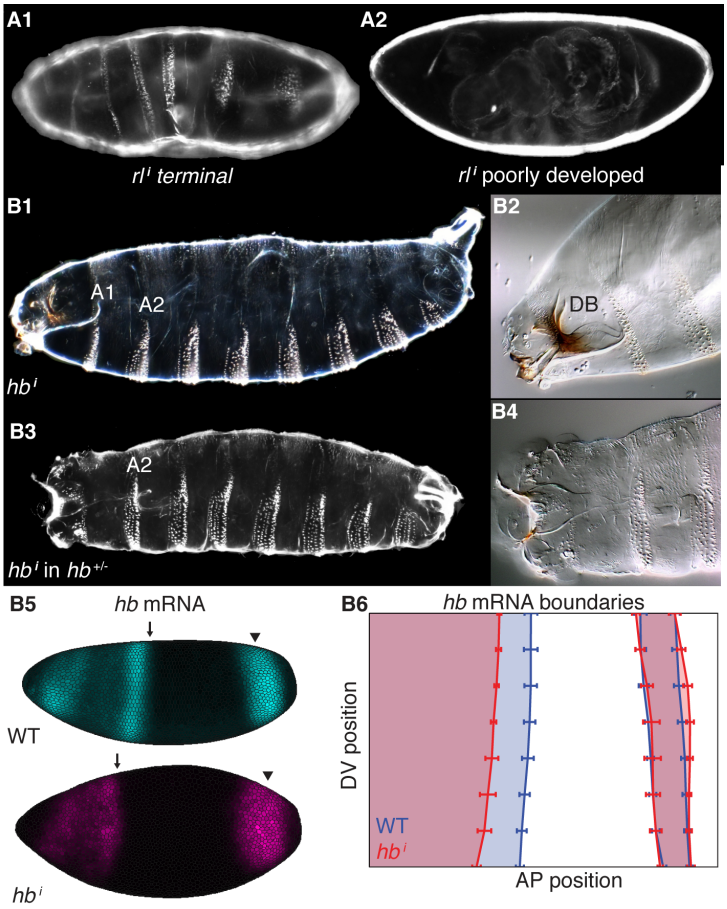


FIGURE 2.4: Embryonic phenotypes associated with *rolled* and *hunchback* shRNAs.
A. *rolled*. When *MTD-Gal4>>UAS-shRNA-rl (GL215)* females were crossed to WT males, the embryos showed differentiated cuticles with terminal defects (A1). However, when crossed to *UAS-shRNA-rl* homozygous males all embryos show poor cuticle development (A2). These phenotypes reflect the role of Rl/MAPK in the Tor and EGFR RTK pathways, respectively (see text). **B.** *hunchback*. Embryos from *MTD-Gal4>>UAS-shRNA-hb* mothers crossed to WT fathers are missing the T2 and T3 thoracic segments, while abdominal segmentation is normal (B1). (B2) shows the head of embryo in (B1). Note that the dorsal bridge (DB) is present and appears normal. When we crossed *MTD-Gal4>>UAS-shRNA-hb* females to WT males, we could not distinguish between embryos with zero or one copy of the *UAS-shRNA-hb* transgene. Similarly, when we crossed *MTD-Gal4>>UAS-shRNA-hb* females to *UAS-shRNA-hb* homozygous males, we could not distinguish between embryos with one or two copies of the *UAS-shRNA-hb* transgene; all three classes of embryos resembled the one shown in Figure 4B1 and 4B2. Together these results demonstrate that zygotically expressed shRNAs do not contribute meaningfully to this phenotype. However, when *MTD-Gal4>>UAS-shRNA-hb* mothers were crossed to *hb[12]/+* males, half of the embryos showed a more severe phenotype (B3). In addition to lacking T2 and T3, these embryos lack the A1 abdominal segment and head structures (B4). Computational representation of *hb* mRNA (maternal and zygotic) *in situ* hybridizations in mid blastoderm stage embryos. The arrow indicates the shift in the anterior expression domain, and the arrowhead indicates that the posterior pattern has not shifted (B5). mRNA expression domain boundaries in embryos from *MTD-Gal4>>UAS-shRNA-hb* females (B6). The vertical lines show the posterior boundary of the anterior expression domain and both boundaries of the posterior domain for each class. The posterior expression domain is unchanged in the *hb* RNAi embryos, while the anterior pattern shifts anteriorly by 10% egg length. Error bars indicate standard deviations.

signaling. In contrast, paternally rescued embryos without a *UAS-shRNA* transgene develop a terminal class phenotype consistent with RI acting downstream of Tor. Altogether, these results are reminiscent of the phenotypes observed from *D-raf* GLCs (see Introduction) and demonstrate that the presence of the shRNA transgene in the embryo needs to be carefully followed to interpret the mutant phenotypes. Importantly, when *MTD-Gal4>>UAS-shRNA* females are crossed to *UAS-shRNA* males, some embryos will carry two and others a single *UAS-shRNA* transgene, which may also account for differences in the severity of embryonic phenotypes. Thus, varying the copy number of zygotic *UAS-shRNA* transgenes provides a useful way to generate phenotypic series and uncover when pleiotropic genes are used in development.

hunchback depletion illustrates the temporal efficacy of *Mat-Gal4* mediated RNAi

Analysis of the Mat&Zyg gene *hb* provided another example of how shRNA depletion of different pools of mRNAs allows the visualization of distinct embryonic phenotypic classes. Maternally deposited *hb* mRNA is selectively translated in the anterior and degraded in posterior regions, while zygotic *hb* is expressed in an anterior domain and a posterior stripe. Embryos lacking both zygotic and maternal *hb* exhibit a more severe phenotype than zygotic mutant embryos, but maternal *hb* is dispensable, as embryos derived from a *hb* homozygous germline can be rescued by a single paternal copy (Lehmann and Nüsslein-Volhard, 1987). Strikingly, embryos derived from *MTD-Gal4>>UAS-shRNA-hb* females exhibit an unusual embryonic lethal phenotype where all the abdominal segments form properly, but two thoracic segments are missing (Figure 2.4B1,B2). This phenotype strongly resembles that of embryos that lack

maternal *hb* and have reduced zygotic *hb* (Simpson-Brose et al., 1994). This phenotype is similar across embryos with zero, one, or two copies of the *UAS-shRNA-hb* transgene, indicating that zygotically expressed shRNAs do not contribute to this phenotype (see legend Figure 2.4).

To examine the distribution of *hb* mRNA after knockdown, we stained for *hb* mRNA by *in situ* hybridization. Compared to WT, the position of the posterior stripe is unchanged in embryos derived from *MTD-Gal4>>UAS-shRNA-hb* females (Figure 2.4B5,B6). In contrast, the anterior expression pattern shifts anteriorly by 10% egg length (EL), and in these embryos, *eve* and *ftz* are each expressed in six stripes rather than their normal seven (data not shown). This defect is consistent with the proposed role of maternal *hb* in working with *bcd* to activate zygotic *hb* robustly and precisely (Porcher et al., 2010).

The observation that embryos derived from *MTD-Gal4>>UAS-shRNA-hb* have a stronger phenotype than those from *hb* germline chimeras (Lehmann and Nüsslein-Volhard, 1987) suggests that some of the maternally loaded shRNAs persisted long enough to knockdown some zygotic *hb* transcripts. Consistent with this model, crossing *MTD-Gal4>>UAS-shRNA-hb* females to *hb/+* males created a second, more severe phenotypic class missing many head structures, as well as the T2, T3, and A1 segments (Figure 2.4B3,B4). This second class resembles embryos that have substantially reduced zygotic expression of anterior *hb* (Wimmer et al., 2000). Together with the data from *dpp* and the heterozygous mutants, these results suggest that the poor knockdown of early zygotic genes stems from our inability to deliver enough shRNAs at the appropriate time.

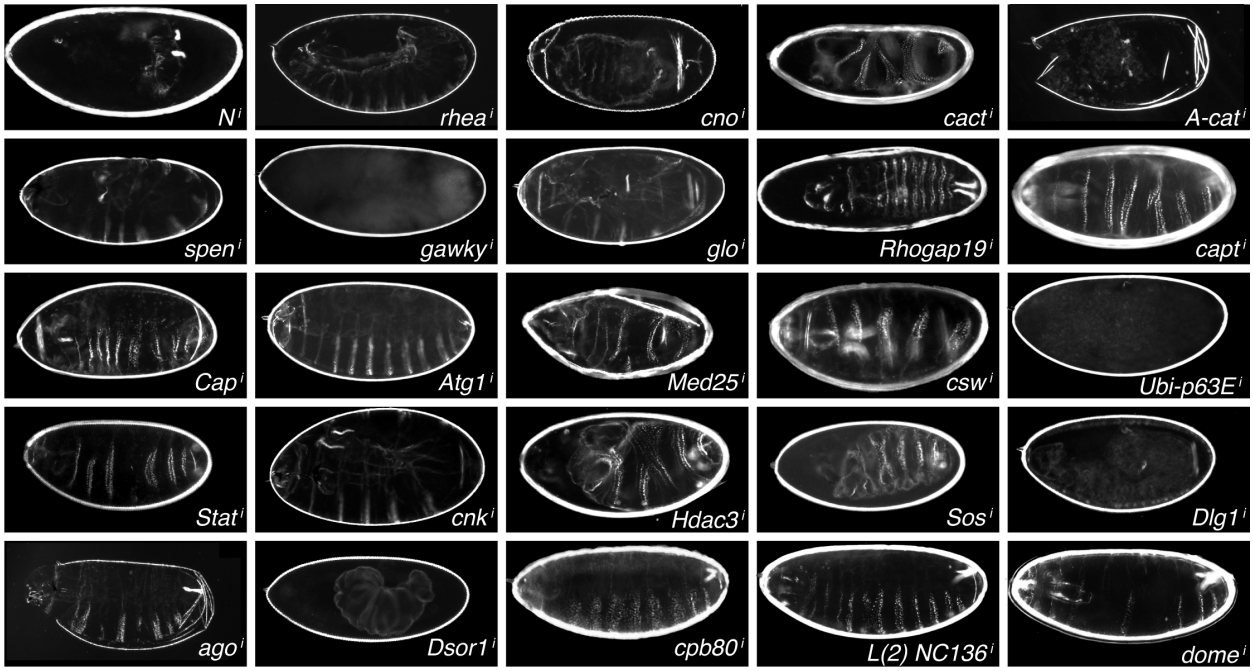


FIGURE 2.5: Embryonic phenotypes associated with Mat&Zyg genes.
F2 embryonic phenotypes of embryos derived from *maternal-Gal4>>UAS-shRNA* females crossed to *UAS-shRNA* males. Details on the shRNA lines associated with F2 phenotypes can be found in Table 1 and the text.

A genetic screen for new Mat&Zyg genes

To date, only about 10% of the genes in *D. melanogaster* have been examined for their maternal functions through the production of GLCs (Perrimon et al., 1989; Perrimon et al., 1996). To demonstrate the efficacy of the “maternal-Gal4 - shRNA” method to characterize the maternal effect of zygotic lethal mutations, we screened more than 1,000 shRNA lines available at the Transgenic RNAi Project (TRiP) in either the VALIUM20 or VALIUM22 vector (see Materials and Methods and the TRiP web site at www.flyrnai.org), and systematically characterized their F1 and F2 phenotypes (Figures 2.3, 2.5 and Table 2.1). A number of shRNAs targeting known genes illustrate the specificity and efficacy of the shRNA lines. These include *domeless, shotgun,*

myospheroid, (Figure 2.4)*, canoe, cactus, alpha-catenin, corkscrew,* and *Stat92E, connector enhancer of ksr, Son of sevenless, discs large 1,* and *Downstream of Raf1* (Figure 2.5). In addition, we recovered novel phenotypes for many genes, in particular the ventralized phenotype associated with *Histone deacetylase 3* (Figure 2.5), the segment polarity phenotypes of *alpha-catenin* (Figure 2.5)*,* the morphogenesis defects associated with *upheld* (Figure 2.4) and *split ends* (Figure 2.5)*,* and the segmentation defects of *archipelago* (Figure 2.5). Additional information on the screened lines is available at www.flyrnai.org/RSVP. Although further analyses, such as the test of additional independent *UAS-shRNA* lines against the same gene or rescue experiments, will need to be done to confirm that these phenotypes are associated with a knockdown of the intended gene, we note that when we observe a phenotype with an *UAS-shRNA* line against a known gene, it matches with the known loss of function phenotype. This agreement most likely reflects the fact that few genes have very specific mutant cuticle phenotypes, reducing the chance that a phenotype is caused by an off target effect.

Table 2.1: Phenotypic analysis of shRNA lines. Unless indicated as "reverse cross" maternal-Gal4 females were crossed to shRNA males. F1 *maternal-Gal4>>UAS-shRNA* females were crossed to sibling males heterozygous for the *UAS-shRNAs*. The % represents the fraction of unhatched eggs. No indicates that embryos have normal viability. NT: Not tested.

Line	Gene name	Vector	Gal4 line	F1 Phenotype	F2 Phenotype
Mat Genes					
HMS00930	nanos (nos)	VALIUM20	mat-tub-Gal4	No	100%, nanos
GL00407	bicoid (bcd)	VALIUM22	mat-tub-Gal4	No	100%, bicoid
GL01320	bicoid (bcd)	VALIUM22	MTD-Gal4	No	100%, bicoid
HMS00727	dorsal (dl)	VALIUM20	mat-tub-Gal4	No	100%, dorsalized
GL00610	dorsal (dl)	VALIUM22	mat-tub-Gal4	No	100%, dorsalized
GL00222	torso (tor)	VALIUM22	mat-tub-Gal4	No	80%, weak torso
HMS00021	torso (tor)	VALIUM20	mat-tub-Gal4	No	100%, torso
Zyg genes					
HMS00595	engrailed (en)	VALIUM20	mat-tub-Gal4	No	No
HMS01312	even skipped (eve)	VALIUM20	mat-tub-Gal4	No	No
HMS01105	giant (gt)	VALIUM20	mat-tub-Gal4	No	No
GL01317	giant (gt)	VALIUM22	MTD-Gal4	No	No
GL01318	giant (gt)	VALIUM22	MTD-Gal4	No	No
GL01319	giant (gt)	VALIUM22	mat-tub-Gal4	No	No
HMS00492	hedgehog (hh)	VALIUM20	mat-tub-Gal4	No	No
HMS01216	huckebein (hkb)	VALIUM20	mat-tub-Gal4	No	No
HMS01184	knirps (kni)	VALIUM20	mat-tub-Gal4	No	No
HMS01106	Kruppel (Kr)	VALIUM20	mat-tub-Gal4	No	No
GL01322	Kruppel (Kr)	VALIUM22	MTD-Gal4	No	No
GL01323	Kruppel (Kr)	VALIUM22	mat-tub-Gal4	No	No
GL01324	Kruppel (Kr)	VALIUM22	MTD-Gal4	No	No
HMS01186	runt (run)	VALIUM20	mat-tub-Gal4	No	No
HMS01108	sloppy paired 2 (slp2)	VALIUM20	mat-tub-Gal4	No	No

Table 2.1 (Continued)					
Line	**Gene name**	**Vector**	**Gal4 line**	**F1 Phenotype**	**F2 Phenotype**
HMS01313	hairy (h)	VALIUM20	mat-tub-Gal4	No	No
HMS01317	twist (twi)	VALIUM20	mat-tub-Gal4	No	No
HMS01215	brother of odd with entrails limited (bowl)	VALIUM22	mat-tub-Gal4	No	No
HMS01122	crocodile (croc)	VALIUM20	mat-tub-Gal4	No	No
HMS01150	Dichaete (D)	VALIUM20	mat-tub-Gal4	No	No
HMS01103	forkhead (fkh)	VALIUM20	mat-tub-Gal4	No	No
HMS01104	fushi tarazu (ftz)	VALIUM20	mat-tub-Gal4	No	No
HMS01552	knirps like (knl)	VALIUM20	mat-tub-Gal4	No	No
HMS01315	odd-skipped (odd)	VALIUM20	mat-tub-Gal4	No	No
HMS01314	ocelliless (oc) / orthodenticle (otd)	VALIUM20	mat-tub-Gal4	No	No
HMS01167	schnurri (shn)	VALIUM20	mat-tub-Gal4	No	No
HMS01107	sloppy paired 1 (slp1)	VALIUM20	mat-tub-Gal4	No	No
HMS01252	snail (sna)	VALIUM20	mat-tub-Gal4	No	No
HMS00794	wingless (wg)	VALIUM20	mat-tub-Gal4	No	No
HMS00844	wingless (wg)	VALIUM20	mat-tub-Gal4	No	No
HMS01109	zerknult 1 (zen)	VALIUM20	mat-tub-Gal4	No	No
HMS01124	zerknult 2 (zen2)	VALIUM20	mat-tub-Gal4	No	No
HMS00545	outstretched (os)	VALIUM20	mat-tub-Gal4	No	No
HMS01316	tailless (tll)	VALIUM20	mat-tub-Gal4	No	No
HMS00922	paired (prd)	VALIUM20	mat-tub-Gal4	No	No
HMS01443	teashirt (tsh)	VALIUM20	mat-tub-Gal4	No	No
JF02455	decapentaplegic (dpp)	VALIUM20	mat-tub-Gal4	60%, retraction defects	100%, ventralized
Mat&Zyg Genes					
HMS00317	alpha Catenin (a-cat)	VALIUM20	mat-tub-Gal4	No	99%, blobbed or segment polarity
HMS00111	archipelago (ago)	VALIUM20	mat-tub-Gal4	No	50%, pair rule phenotype

Table 2.1 (Continued)					
Line	**Gene name**	**Vector**	**Gal4 line**	**F1 Phenotype**	**F2 Phenotype**
HMS01414	armadillo (arm)	VALIUM20	mat-tub-Gal4	100%, segment polarity	NT
HMS01414	armadillo (arm) (reverse cross)	VALIUM20	mat-tub-Gal4	No	100%, segment polarity
GL00047	Autophagy-specific gene 1 (Atg1)	VALIUM22	mat-tub-Gal4	No	80%, some head defects
HMS00084	cactus (cact)	VALIUM20	mat-tub-Gal4	No	100%, ventralized
GL00627	cactus (cact)	VALIUM22	mat-tub-Gal4	No	100%, ventralized
HMS00239	canoe (cno)	VALIUM20	mat-tub-Gal4	No	100%, dorsal open
GL00633	canoe (cno)	VALIUM22	mat-tub-Gal4	No	100%, dorsal open
HMS00052	cap binding protein 80 (CBP80)	VALIUM20	mat-tub-Gal4	100% cuticles WT	No eggs
HMS00810	capulet (capt)	VALIUM20	mat-tub-Gal4	No	99%, some U-shaped
HMS00318	Chromosome -associated protein (Cap)	VALIUM20	mat-tub-Gal4	No	100%, mostly blobbed
HMS00238	connector enhancer of ksr (cnk)	VALIUM20	mat-tub-Gal4	No	80%, weak terminal class
HMS00012	corkscrew (csw)	VALIUM20	mat-tub-Gal4	No	100%, weak corkscrew
JF02287	discs large 1 (dlg1)	VALIUM20	mat-tub-Gal4	No	100%, blobbed and dorsal open
HMS00647	domeless (dome)	VALIUM20	mat-tub-Gal4	No	90%, JAK/STAT variable
HMS01293	domeless (dome)	VALIUM20	mat-tub-Gal4	100%, head defects	NT
HMS00145	Downstream of raf1 (Dsor1)	VALIUM20	mat-tub-Gal4	No	100%, terminal defects
HMS00128	Elongin C (Elongin C)	VALIUM20	mat-tub-Gal4	No	100%, white eggs
HMS00105	gawky (gw)	VALIUM20	mat-tub-Gal4	No	100%, abnormal oogenesis, fused filaments, a few brown eggs
HMS00079	glorund (glo)	VALIUM20	mat-tub-Gal4	No	95%, some JAK/STAT
HMS00076	Helicase at 25E (Hel25E)	VALIUM20	mat-tub-Gal4	Many dead embryo with WT cuticle dead L1 dead few adults	100%, abnormal oogenesis
HMS00087	Histone deacetylase 3 (Hdac3)	VALIUM20	mat-tub-Gal4	100%, head defects	100%, ventralized

Table 2.1 (Continued)					
Line	**Gene name**	**Vector**	**Gal4 line**	**F1 Phenotype**	**F2 Phenotype**
GL01321	hunchback (hb)	VALIUM22	mat-tub-Gal4	No	90% head defect, some segmentation defects
HMS00802	lethal (2) NC136 (l(2)NC136)	VALIUM20	mat-tub-Gal4	100% cuticles WT	No eggs
HMS00256	Mediator complex subunit 25 (Med 25)	VALIUM20	mat-tub-Gal4	No	95%, some blobbed some U-shaped
HMS00043	myospheroid (mys)	VALIUM20	mat-tub-Gal4	85%, variable cuticle	NT
HMS00526	Not1 (Not1)	VALIUM20	mat-tub-Gal4	Larval lethal, very fe	NT
HMS00009	Notch (N)	VALIUM20	mat-tub-Gal4	95%, neurogenic embryos	NT
GL00092	Notch (N)	VALIUM22	mat-tub-Gal4	10%, neurogenic embryos	75%, neurogenic
HMS00009	Notch (N) (reverse cross)	VALIUM20	mat-tub-Gal4	No	NT
HMS00310	pasilla (ps)	VALIUM20	mat-tub-Gal4	85%, cuticles WT, f	NT
HMS01662	PDGF- and VEGF-receptor related (Pvr)	VALIUM20	mat-tub-Gal4	No	100%, of embryos unhatched, WT cuticle
HMS00187	Proteasome beta3 subunit (Prosbeta3)	VALIUM20	mat-tub-Gal4	Some brown eggs, r	NT
HMS00856	rhea (rhea)	VALIUM20	mat-tub-Gal4	100%, cuticles WT	NT
HMS00799	rhea (rhea)	VALIUM20	mat-tub-Gal4	100%, cuticles WT	100%, dorsal cuticle defects
HMS00352	RhoGAP19D (RhoGAP19D)	VALIUM20	mat-tub-Gal4	No	50% embryos anterior holes
HMS00968	Ribosomal protein S15Aa (Rps15Aa)	VALIUM20	mat-tub-Gal4	80%, cuticles WT,	NT
HMS00173	rolled (rl)	VALIUM20	mat-tub-Gal4	No	100%, blobbed, terminal defects
GL00215	rolled (rl)	VALIUM22	MTD-Gal4	No	100%, blobbed, terminal defects
HMS00693	shotgun (shg)	VALIUM20	mat-tub-Gal4	80%, some dorsal c	NT
HMS01009	Sirt6 (Sirt6)	VALIUM20	mat-tub-Gal4	80%, cuticles WT,	NT
HMS00274	small nuclear ribonucleoprotein 70K (snrp70K)	VALIUM20	mat-tub-Gal4	80%, variable, few la	NT
HMS00149	Son of sevenless (sos)	VALIUM20	mat-tub-Gal4	No	100%, weak terminal class

Table 2.1 (Continued)					
Line	**Gene name**	**Vector**	**Gal4 line**	**F1 Phenotype**	**F2 Phenotype**
HMS00276	split ends (spen)	VALIUM20	mat-tub-Gal4	No	99%, of embryos with U-shaped and head defects
HMS00035	Stat92E (dStat)	VALIUM20	mat-tub-Gal4	No	100%, JAK/STAT phenotype
HMS00580	trithorax (trx)	VALIUM20	mat-tub-Gal4	Few dead embryos,	NT
HMS01417	tumbleweed (tum)	VALIUM20	mat-tub-Gal4	80%, cuticles WT,	NT
HMS00284	Ubiquitin-63E (Ubi-p63E)	VALIUM20	mat-tub-Gal4	No	100%, No development
HMS00743	upheld (up)	VALIUM20	mat-tub-Gal4	100%, head defects segments compressed	NT
HMS01618	zipper (zip)	VALIUM20	mat-tub-Gal4	No	100%, abnormal oogenesis, few abnormal eggs
HMS02519	Kruppel (Kr)	VALIUM20 -miR92a	mat-tub-Gal4	WT	WT
HMS02518	Notch (N)	VALIUM20 -miR92a	mat-tub-Gal4	neurogenic	100% neurogenic
HMS02520	armadillo (arm)	VALIUM20 -miR92a	mat-tub-Gal4	segment polarity	NT
HMS02521	bicoid (bcd)	VALIUM20 -miR92a	mat-tub-Gal4	NT	100%, bicoid
HMS02522	wingless (wg)	VALIUM20 -miR92a	mat-tub-Gal4	WT	WT
HMS02516	giant (gt)	VALIUM20 -miR275	mat-tub-Gal4	NT	WT
HMS02517	ovarian turmor (otu)	VALIUM20 -miR275	mat-tub-Gal4	NT	few eggs
HMS02511	bicoid (bcd)	VALIUM20 -miR275	mat-tub-Gal4	NT	100%, bicoid

Conclusion

We evaluated the efficacy of the Gal4-UAS system to drive shRNA expression in early embryos by performing a number of tests using shRNAs targeting Mat, Zyg, and Mat&Zyg expressed genes. We show that Gal4 driven shRNAs in the female germline efficiently generate mutant phenotypes. In addition, loading the embryo with shRNAs against early zygotic genes was only effective in a few cases (*dpp* and *hb*), possibly because shRNAs are unstable (see model in Figure 2.6). However, the efficacy of additional shRNAs was unmasked by generating heterozygous mutant zygotic backgrounds. To increase the stability of our shRNAs, we generated two new delivery backbones, which, although effective, did not increase the severity of phenotypes recovered. Interestingly, maternally loaded Gal4 protein, in combination with different copy numbers and delivery methods of *UAS-shRNAs*, can be used to knockdown zygotic transcripts in certain time windows and reveal distinct and discrete phenotypes of pleiotropic genes. The system appears especially effective at depleting genes required during mid embryogenesis after gastrulation (4-5 hours after egg laying). A possible way to improve the efficacy of RNAi in embryos would be to cross *maternal-Gal4>>UAS-shRNA* females to males carrying a strong uniformly expressed zygotic Gal4 driver.

The “maternal Gal4 – shRNA” method will allow a number of investigations in *Drosophila* embryos. In particular, the opportunity to collect large pools of homogenous embryos will enable biochemical analyses (R. Sopko and N. Perrimon, unpublished). Further, the technique will be useful for the analysis of regulatory network architecture and *cis*-regulatory element reporter constructs.

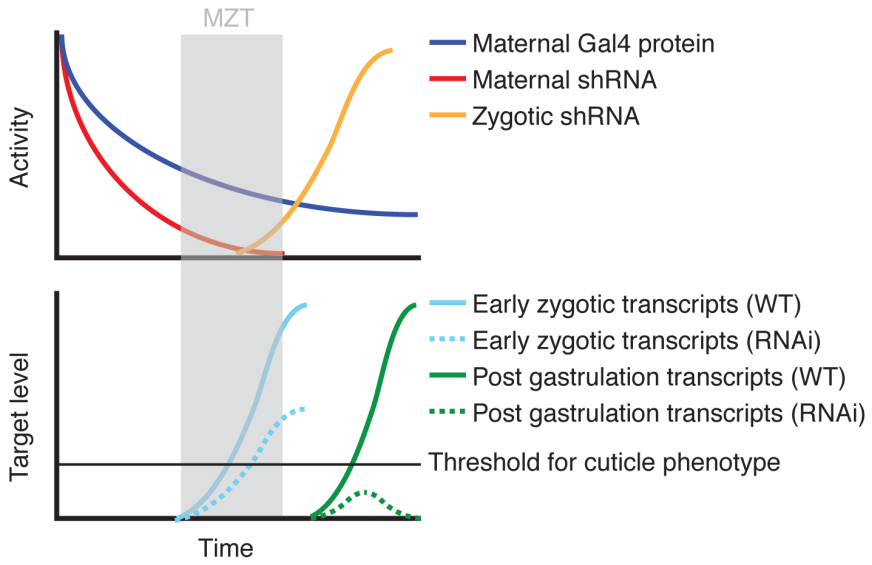


FIGURE 2.6: Model for gene knockdown using the "maternal Gal4 – shRNA" system.
Maternally deposited shRNAs can deplete early zygotic transcripts only modestly, in most cases not enough to reveal a phenotype (red acting on cyan). Zygotically activated shRNAs can effectively deplete target transcripts when they are expressed before the target is activated (orange acting on green). Early patterning genes escape knockdown because maternally loaded shRNAs lose efficacy over time, and zygotically expressed shRNAs are activated too late. Maternal-Zygotic transition (MZT).

Acknowledgments: We are thankful to the Transgenic RNAi facility (TRiP) for providing the shRNA lines used in this study and Lynn Cooley, Daniel St. Johnston and Fredrick Wirtz-Peitz for the maternal Gal4 driver lines. We acknowledge Richelle Sopko and Rich Binari for helpful discussions, Ben Vincent and Clarissa Scholes for close reading of the manuscript, and Tara Lydiard-Martin for help analyzing the *hb* expression patterns. M.S. was supported by the Harvard Herchel Smith and Harvard Merit Fellowships. This work was supported in part by NIH-GM084947 (N.P). N.P is an investigator of the Howard Hughes Medical Institute.

References

Ambrosio, L., A. Mahowald, and N. Perrimon, 1989 Requirement of the Drosophila raf homologue for torso function. Nature **342:** 288-291.

Brunner, D., Oellers, N., Szabad, J., Biggs, W.H.3rd., and E. Hafen, 1994 A gain-of-function mutation in Drosophila MAP kinase activates multiple receptor tyrosine kinase signaling pathways. Cell **76**: 875-888.

Chou, T., and N. Perrimon, 1996 The autosomal FLP-DFS technique for generating germline mosaics in Drosophila melanogaster. Genetics **144:** 1673-1679.

Duffy, J., and N. Perrimon, 1994 The torso pathway in Drosophila: lessons on receptor tyrosine kinase signaling and pattern formation. Dev Biol **166:** 380-395.

Fowlkes, C., C. L. Luengo Hendriks, S. Keränen, G. Weber, O. Rübel, M. Huang, S. Chatoor, A. H. DePace, L. Simirenko, C. Henriquez, A. Beaton, R. Weiszmann, S. Celniker, B. Hamann, D. Knowles, M. Biggin, M. Eisen, and J. Malik, 2008 A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell, **133:** 364–374.

Fowlkes, C., K. Eckenrode, M. Bragdon, M. Meyer, Z. Wunderlich, L. Simirenko, C. L. Luengo Hendriks, S. Keranen, C. Henriquez, D. Knowles, M. Biggin, M. Eisen, and A. H. DePace, 2011 A conserved developmental patterning network produces quantitatively different output in multiple species of Drosophila. PLoS Genet **7:** e1002346.

Irish, V., and W. Gelbart, 1987 The decapentaplegic gene is required for dorsal-ventral patterning of the Drosophila embryo. Genes Dev **1:** 868-879.

Kozomara, A., and S. Griffiths-Jones, 2011 miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res **39:** D152-7.

Lawrence, P. A., 1992 *The making of a fly: the genetics of animal design*. Blackwell Scientific Publications Ltd, Oxford, UK.

Lehmann, R., U. Dietrich, F. Jimenez, and J. A. Campos-Ortega, 1981 Mutations of early neurogenesis in Drosophila. Development Genes and Evolution **190:** 226-229.

Lehmann, R., and C. Nüsslein-Volhard, 1987 hunchback, a gene required for segmentation of an anterior and posterior region of the Drosophila embryo. Developmental biology **119:** 402.

Li, W., 2005 Functions and mechanisms of receptor tyrosine kinase Torso signaling: lessons from Drosophila embryonic terminal development. Dev Dyn **232:** 656-672.

Luengo Hendriks, C. L., S. Keranen, C. Fowlkes, L. Simirenko, G. Weber, A. Depace, C. Henriquez, D. Kaszuba, B. Hamann, M. Eisen, J. Malik, D. Sudar, M. Biggin, and D. Knowles, 2006 Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. Genome Biol **7:** R123.

Ni, J., R. Zhou, B. Czech, L. Liu, L. Holderbaum, D. Yang-Zhou, H. Shim, R. Tao, D. Handler, P. Karpowicz, R. Binari, M. Booker, J. Brennecke, L. Perkins, G. Hannon, and N. Perrimon, 2011 A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nat Methods **8:** 405-407.

Peifer, M., C. Rauskolb, M. Williams, B. Riggleman, and E. Wieschaus, 1991 The segment polarity gene armadillo interacts with the wingless signaling pathway in both embryonic and adult pattern formation. Development **111:** 1029-1043.
Perrimon, N., L. Engstrom, and A. P. Mahowald, 1985 A pupal lethal mutation with a paternally influenced maternal effect on embryonic development in Drosophila melanogaster. Developmental biology **110:** 480-491.
Perrimon, N., L. Engstrom, and A. Mahowald, 1989 Zygotic lethals with specific maternal effect phenotypes in Drosophila melanogaster. I. Loci on the X chromosome. Genetics **121:** 333-352.
Perrimon, N., A. Lanjuin, C. Arnold, and E. Noll, 1996 Zygotic lethal mutations with maternal effect phenotypes in Drosophila melanogaster. II. Loci on the second and third chromosomes identified by P-element-induced mutations. Genetics **144:** 1681-1692.
Petrella, L., T. Smith-Leiker, and L. Cooley, 2007 The Ovhts polyprotein is cleaved to produce fusome and ring canal proteins required for Drosophila oogenesis. Development **134:** 703-712.
Porcher, A., A. Abu-Arish, S. Huart, B. Roelens, C. Fradin, and N. Dostatni, 2010 The time to measure positional information: maternal hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription. Development **137:** 2795-2804.
Ruby, J., A. Stark, W. Johnston, M. Kellis, D. Bartel, and E. Lai, 2007 Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res **17:** 1850-1864.
Simpson-Brose, M., J. Treisman, and C. Desplan, 1994 Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. Cell **78:** 855-865.
Spencer, F. A., F. M. Hoffmann, and W. M. Gelbart, 1982 Decapentaplegic: a gene complex affecting morphogenesis in Drosophila melanogaster. Cell **28:** 451-461.
St. Johnston, D., and C. Nusslein-Volhard, 1992 The origin of pattern and polarity in the Drosophila embryo. Cell **68:** 201-219.
Tadros, W., and H. Lipshitz, 2009 The maternal-to-zygotic transition: a play in two acts. Development **136:** 3033-3042.
Votruba, S. M., 2009 miRNAs in the *Drosophila* egg and early embryo. Department of Molecular Genetics **Master of Science:** 1-75.
Wimmer, E., A. Carleton, P. Harjes, T. Turner, and C. Desplan, 2000 Bicoid-independent formation of thoracic segments in Drosophila. Science **287:** 2476-2479.

Chapter 3: A gene expression atlas of a *bicoid*-depleted *Drosophila* embryo reveals early canalization of cell fate

Short Title

Canalization in a gene expression atlas of a *bicoid*-depleted *Drosophila* embryo

Running Title

Canalization in *bicoid* RNAi embryos

Max V. Staller[1], Charless C. Fowlkes[2], Meghan D.J. Bragdon[1], Javier Estrada[1], Zeba Wunderlich[1], Angela H. DePace[1,]

[1]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts
[2]Department of Computer Science, University of California Irvine, Irvine, California

Abstract

In developing embryos, gene regulatory networks drive cells towards discrete terminal fates, a process called canalization. We studied the behavior of the anterior-posterior segmentation network in *Drosophila melanogaster* embryos by depleting a key maternal input, *bicoid (bcd),* and measuring gene expression of the network at cellular resolution. This method results in a gene expression atlas containing the levels of mRNA or protein expression of thirteen core patterning genes over six time points for every cell of the blastoderm embryo. This is the first cellular resolution dataset of a genetically perturbed *Drosophila* embryo that captures all cells in 3D. We describe the technical developments required to build this atlas and how the method can be employed and extended by others. We also analyze this novel dataset to characterize the degree of cell fate canalization in this network and when it occurs. We find that in two layers of this gene regulatory network, following depletion of *bcd,* individual cells rapidly canalize towards normal cell fates. This result supports the hypothesis that the segmentation network topology directly canalizes cell fate, rather than an alternative hypothesis where cells are initially misspecified and later eliminated by apoptosis. Our gene expression atlas provides a high resolution picture of a classic perturbation and will enable further computational modeling of canalization and gene regulation in this transcriptional network.

Key Words

Canalization, *Drosophila*, *bicoid*, *even-skipped*, transcriptional network, gene expression atlas

Introduction

Specialization of cell fate underlies the diversity of metazoan form and function. Cell fates are specified robustly and precisely by gene regulatory networks that pattern embryos (Davidson, 2010). How do developmental networks encode these properties? The *Drosophila melanogaster* blastoderm embryo is a premiere system for coupling computational models and quantitative experimental data to test hypotheses about the design of developmental networks (Reinitz and Sharp, 1995; Jaeger et al., 2004a; Jaeger et al., 2004b; Poustelnikova et al., 2004). Anterior-posterior patterning of the embryo is controlled by the well-characterized segmentation network (Lawrence, 1992; St Johnston and Nüsslein-Volhard, 1992; Jaeger et al., 2013). Computational models of this network have tested the sufficiency of known connections, proposed new connections, and tested network-level properties (Wunderlich and DePace, 2011; Jaeger et al., 2013).

Gene expression atlases enable the study of network properties. These atlases combine measurements of mRNA or protein expression from many individual embryos into an average embryo; the resulting data are in a computationally amenable format with high resolution in space and time. The first such dataset, the FlyEx Database (Poustelnikova et al., 2004; Pisarev et al., 2009) was a 1D anterior-posterior atlas that triggered a renaissance in computational modeling of fly patterning and transcriptional control (Jaeger et al., 2004b; Janssens et al., 2006; Segal et al., 2008; He et al., 2010). The 3D atlas built by the Berkeley Drosophila Transcription Network Project (Keränen et al., 2006; Luengo Hendriks et al., 2006; Luengo Hendriks et al., 2007; Fowlkes et al., 2008) enabled similar approaches in every cell of the embryo (Bieler et al., 2011;

Hengenius et al., 2011; Umulis and Othmer, 2012; Ilsley et al., 2013; Samee and Sinha, 2013). Extending 3D atlas building methods to other species enabled quantitative comparison of transcriptional circuits (Fowlkes et al., 2011; Wunderlich et al., 2012).

The existing wildtype (WT) atlases allow for fitting and cross validation of computational models, but a gold standard for computational models is whether they can predict behavior under genetic perturbation. This strategy has been difficult to apply because of limited quantitative data for mutant embryos. It is common to simulate the effect of a mutation and qualitatively compare the computational model predictions to published images in the literature (Papatsenko and Levine, 2008; Ilsley et al., 2013; Kim et al., 2013). However, it is difficult to accurately simulate mutant embryos because both direct and indirect effects are prevalent. To validate computational models, it is clearly preferable to have direct quantitative measurements of entire gene regulatory networks in mutant embryos (Kozlov et al., 2012; Janssens et al., 2013; Surkova et al., 2013).

Here we present a 3D gene expression atlas of a *Drosophila* blastoderm embryo depleted of the maternal transcription factor *bicoid (bcd)*. We overcame two technical challenges: first, collecting enough embryos for high throughput imaging, and second, controlling phenotypic variability. To solve the first problem, we used the maternal Gal4 shRNA system (Staller et al., 2013). shRNA depletion is genetically dominant, avoids labor intensive sorting of mutant females, and enables biochemical analysis in future work. The second problem, phenotypic variation, is shared by shRNA depletion and mutant alleles (Waddington, 1942). We reduced variability both experimentally and by curating our dataset so that the resulting atlas is a meaningful representation of the dominant phenotypic class. Both of these technical developments will be applicable to

building gene expression atlases of additional genetic perturbations in the future. We anticipate that the gene expression atlas we describe here will be a valuable resource for computational analysis and modeling of gene regulation in *Drosophila* blastoderm embryos.

Our goal in building this atlas was not to investigate *bcd* behavior per se, but to determine how individual cells respond to a dramatic perturbation of the segmentation network. Bcd protein activates head cell fates and represses posterior cell fates (Lawrence, 1992). Deleting *bcd* leads to duplication of posterior structures in the anterior, a strong perturbation of cell fate specification. More subtle perturbations, such as variations in Bcd levels, have been useful for computational modeling of segmentation network behavior (Manu et al., 2009a; Manu et al., 2009b; Gursky et al., 2011). Our *bcd*-depleted gene expression atlas combines data for thirteen key segmentation genes and seven reporters for enhancers that respond to these genes into a single morphological framework for 6 time points in stage 5 (blastoderm). This atlas captures the direct and indirect effects of *bcd* depletion on the segmentation network for every cell of the embryo.

We used our atlas of the *bcd*-depleted embryo to investigate canalization of cell fate in individual cells. In his 1942 paper, Conrad Waddington used genetic and embryological evidence to support the idea that development canalizes cell fate (Waddington, 1942; Waddington, 1957). Each of these lines of evidence has developed into a different modern definition of canalization. First, WT individuals are phenotypically highly reproducible while mutant populations are more variable; this genetic evidence leads to one definition, that developmental systems buffer genetic and environmental

perturbations to create stereotyped individuals, reducing phenotypic variability over time. Second, differentiated cells and tissues are distinct; this embryological evidence leads to the second definition, that developmental systems create discrete cell fates, avoiding hybrids. We focused primarily on the second definition of canalization: developmental systems create discrete cell fates.

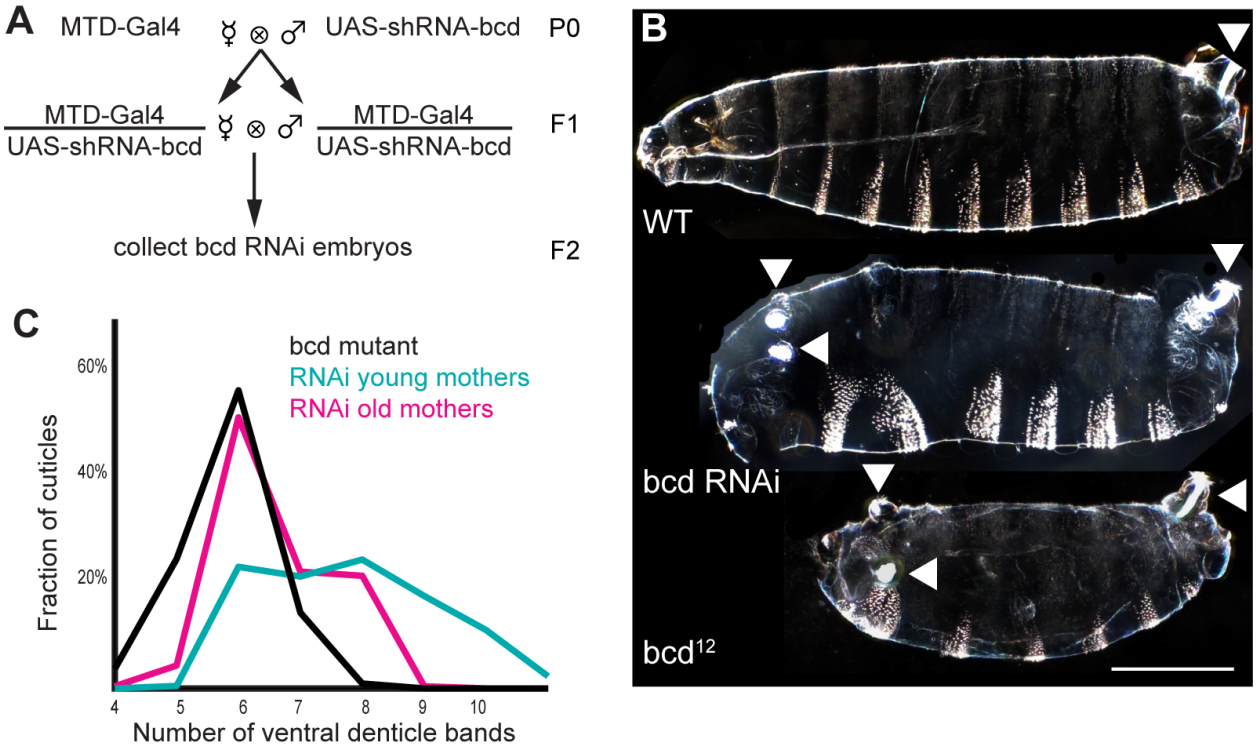
To examine canalization, we used cellular gene expression profiles as a proxy for cell fate (Waddington, 1957). We defined cellular gene expression profiles (and thus cell fate) as combinations of key transcription factors. This strategy allows us to ask if any new cell fates emerge in the *bcd*-depleted embryo. We examined two layers of the segmentation network, the gap genes and the pair-rule genes. We found that all gap gene cell fates present in the *bcd*-depleted embryo were also present in WT. For the pair-rule genes, we observed that extensive early overlap of *even-skipped (eve)* and *fushi-tarazu (ftz)* mRNA expression domains in *bcd*-depleted embryos resolved into mutually exclusive domains. These two results support the hypothesis that canalization is directly encoded by the segmentation network and emerges at the blastoderm stage. The techniques we describe can be readily applied to other genetic perturbations, and the *bcd* RNAi gene expression atlas we present will be a useful resource for computational modeling of gene regulation in the embryo.

Results

Maternal Gal4 shRNA knockdown of *bcd* phenocopies mutant alleles.

To collect the large quantities of embryos necessary to build a gene expression atlas, we used the "maternal Gal4 shRNA" system to deplete *bcd* mRNA in the female germ line (Ni et al., 2011; Staller et al., 2013). shRNAs are genetically dominant, a feature that avoids labor intensive sorting of mutant females to ensure all embryos are affected. The fly husbandry is simple, enabling biochemical and functional genomic analysis. The technique is extendable to other genes, may prove more consistent than classic mutant alleles, and is inducible in specific tissues, an advantage over CRISPR-Cas9 genome editing (Ren et al., 2013). We crossed *maternal triple driver Gal4 (MTD-Gal4)* females with *UAS-shRNA-bcd* males and collected embryos laid by *MTD-Gal4/ UAS-shRNA-bcd* females (Figure 3.1A).

Both mutant and shRNA embryos exhibited phenotypic variability, which needed to be controlled prior to building a gene expression atlas (Figs 2.1B, S1). Embryos laid by *MTD-Gal4/UAS-shRNA-bcd* females (*bcd* RNAi embryos) had a distribution of phenotypes that overlaped the distribution of embryos laid by *bcd* mutant females (*bcd* mutant embryos) (Figs3.1B, Appendix B S1) (Frohnhöfer and Nüsslein-Volhard, 1986). We quantified phenotypic variability by counting the number of ventral denticle bands on each cuticle (Figure 3.1C). The primary determinant of phenotypic strength and variability in *bcd* RNAi embros was the age of the mothers: older mothers laid embryos with stronger and less variable phenotypes (Figure 3.1C). This improvement may stem from a slowing of oogenesis in older females, permitting the shRNAs more time to deplete targets (Ni et al., 2011). By day 8, as measured by qPCR, <10% of *bcd*



transcripts remained (Appendix B Figure S2B). To balance the need to reduce variability

Figure 3.1: Gal4-driven shRNA against *bcd* in the female germline phenocopies *bcd* mutant alleles
(A) Crossing scheme for generating *bcd* RNAi embryos (methods). (B) Top: dark field image of a WT larval cuticle. The white patches of bristles on each segment are the ventral denticle bands. Middle: *bcd* RNAi cuticle. Bottom: *bcd* mutant cuticle. The *bcd* RNAi embryo has several key features of a classic *bcd* mutant, including the the absence of all head and thoracic structures, and the unextended ectopic filzkörper (arrowheads). All cuticles are oriented with anterior to the left and ventral on the bottom. Scale bar 200 microns. (C) The strength of knockdown increases and the phenotypic variability decreases as the *MTD-Gal4/UAS-shRNA-bcd* mothers age. Mutant n = 216; old mothers n = 253, day 15; young mothers n = 217, day 3. Coefficients of variation are: mutant = .127, old mothers = .136, and young mothers = .180.

against declining fecundity, we collected embryos after aging the flies in cages for at least 11 days, at which point >90% of embryos passed our threshold for a strong *bcd* phenotype: 8 or fewer denticle bands (all abdominal) and ectopic tail structures (Appendix B Figure S2A).

Cuticle preparations provide a fast and easy way to identify sources of variability. We tested temperature, shRNA sequence, maternal driver, paternal genotype and the number of *UAS-shRNA-bcd* transgenes in each embryo, but none contributed strongly to phenotypic variability (Appendix B Figure S2C,D). The absence of any paternal or zygotic effects enabled introduction of enhancer *lacZ* reporters into the atlas (methods; Appendix B Table S1).

Building a gene expression atlas of a *bicoid* depleted embryo

To build a gene expression atlas, many individually stained embryos are registered together using a common gene expression pattern (also known as a fiduciary marker). Registration requires a template embryo, which captures both average embryo morphology (cell number and cell density) and the expression pattern of the fiduciary marker. *bcd* RNAi embryos differ in morphology and fiduciary marker expression. We therefore needed to build a new template. We built our *bcd* RNAi registration template using 249 embryos stained only for *ftz* mRNA (Fowlkes et al., 2008; Fowlkes et al., 2011). At late time points, some embryos expressed an extra *ftz* stripe, and these individuals were excluded from the dataset. In principle, many genes could serve as a fiduciary marker. In WT either *eve* or *ftz* was used (Fowlkes et al., 2008). We chose *ftz* because the probe is very reliable.

Characteristics of the *bcd* RNAi gene expression atlas

The *bcd* RNAi atlas includes 1817 embryos with mRNA stains for *caudal (cad), Kruppel (Kr), knirps (kni), giant (gt), hunchback (hb), fork head (fkh), huckebein (hkb),*

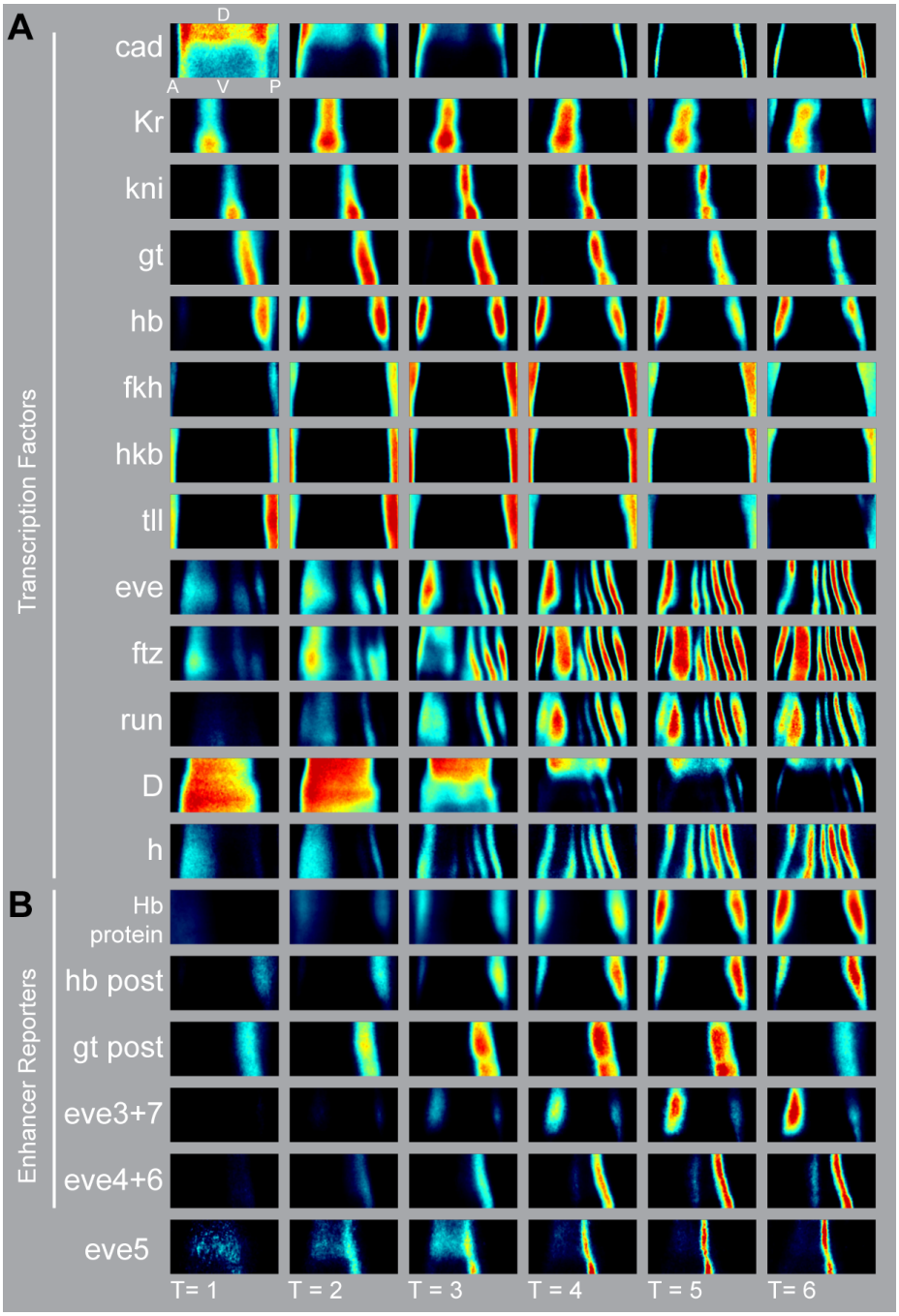


Figure 3.2: The gene expression atlas of a *bcd* depleted embryo highlights the expansion of trunk patterns, duplication of posterior patterns and loss of anterior patterns

(A) Heat maps for mRNA expression patterns of 13 genes in the *bcd* RNAi atlas. Relative mRNA levels scale from no expression (black) to peak expression (red). We also collected Hb protein data. We partition the data into 6 ~10 min cohorts that span all of stage 5 using a morphological marker (methods). (B) Heat maps of five reporter constructs included in the atlas. Anterior is left, dorsal is top.

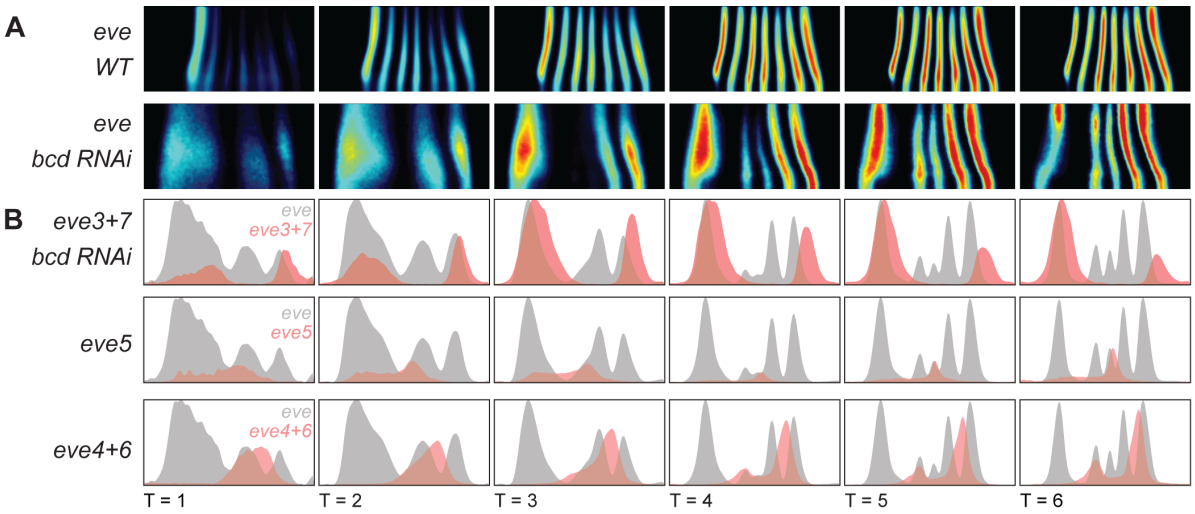


Figure 3.3: Enhancer reporter constructs identify the *eve* stripes in *bcd* RNAi embryos
(A) Heat maps for *eve* mRNA patterns in the WT and *bcd* RNAi gene expression atlases. (B) Line traces of the endogenous *eve* pattern (gray) and the reporter (orange) plot anterior-posterior position on the x-axis and expression level on the y-axis for a single strip along the side of the embryo. The levels of the reporter line traces have been manually scaled to match the corresponding endogenous stripe.

tailless (tll), Dichaete (D), runt (run), hairy (h), even-skipped (eve), and *fushi-tarazu (ftz)* (Figure 3.2; embryos per gene in Appendix B Table S2). In addition, we measured 6 *lacZ* reporter constructs containing the following enhancers: *hb posterior, gt posterior, eve stripe3+7, eve stripe5,* two *eve stripe4+6* enhancers, and whole locus *eve* reporter (gift from Miki Fujioka) (Appendix B Table S1). We also collected embryos carrying reporters for the *eve stripe1, eve stripe2,* and *eve late seven stripe* enhancers, but these sequences drove very little expression in the blastoderm. Finally, we collected protein data for Hb, for which there is a large difference in the mRNA and protein patterns in both WT and *bcd* mutants (Figure 3.2, Appendix B Figure S4). In WT, anterior Hb protein arises from translational regulation of maternal mRNA and *bcd* activated zygotic mRNA. In *bcd* RNAi, there is a broad maternally controlled pattern

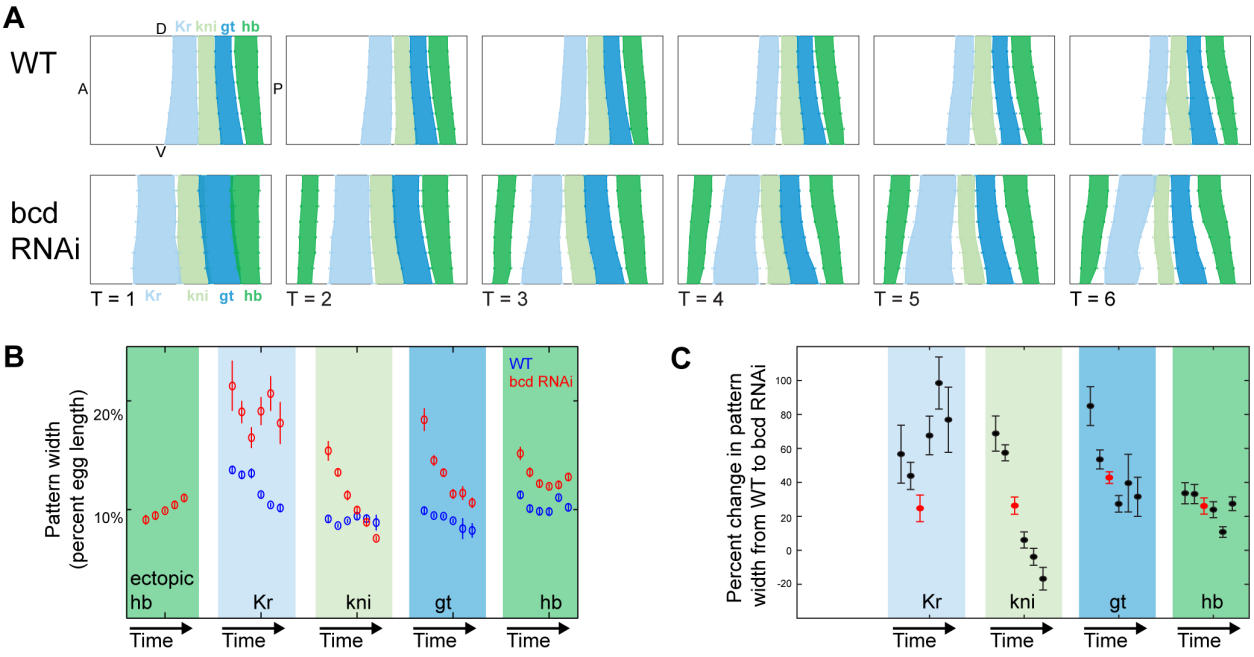


Figure 3.4: The gap gene expression patterns in the trunk expand by different amounts in *bcd* RNAi embryos
(A) The gap gene expression patterns in the trunk in WT and *bcd* RNAi gene expression atlases plotted as unrolled embryos. The pattern boundaries were calculated by finding the inflection point of lateral strips in individual embryos. Error bars are s.e.m.. (B) The widths of each gap gene expression domain change over time in WT (blue) and *bcd* RNAi (red). For each gene, the width of the pattern, calculated from a lateral strip, is plotted over 6 time points. The patterns narrow over time in both genotypes, but more quickly in *bcd* RNAi. Pattern widths plotted as percent egg length (EL). One nuclear diameter is ~1% EL. (C) The percent change in gap gene expression domain widths between WT and *bcd* RNAi, calculated for each time point from C. Time point 3 is indicated in red.

and single zygotic stripe, a duplication of the posterior stripe (Tautz, 1988; Hulskamp et al., 1989; Irish et al., 1989; Struhl et al., 1989; Fowlkes et al., 2008). All gene expression patterns agree with published images (Frasch and Levine, 1987; Nüsslein-Volhard et al., 1987; Tautz, 1988; Hooper et al., 1989; Struhl et al., 1989; Hulskamp et al., 1990; Kraut and Levine, 1991a; Rivera-Pomar et al., 1995), but our high temporal and spatial resolution revealed dynamics not always captured in published images.

There was more variability in the pair-rule gene expression patterns than in the gap gene expression patterns. In 22/98 of *bcd* RNAi embryos, the anterior *eve* stripe split at T=5 and T=6. These embryos were excluded from the atlas. In embryos with a single anterior stripe, the position and width of this stripe varied more than the other stripes (Appendix B Figure S5). Aside from the anterior stipe, the coefficients of variation of *eve* stripe widths were comparable to gap gene widths, indicating that both layers of the network had similar embryo-to-embryo variability (Appendix B Figure S6). The boundaries of both the reporters and endogenous *eve* stripes refined later in *bcd* RNAi than in WT (Figs 3.2, 3.3, Appendix B S5).

The *bcd* RNAi gene expression atlas is of similar or higher quality than the original WT atlas. First, the standard deviation of each gene averaged over all cells and all time points was smaller in the *bcd* RNAi atlas than in the WT atlas for 7 of 10 genes (Appendix B Table S3). Second, for all but a few genes, background expression levels in cells with low expression levels (OFF cells) were lower in *bcd* RNAi, as shown in the histogram of expression levels (Figure 3.5A). The atlas is freely available at depace.med.harvard.edu.

Identifying the perturbed *eve* stripes in *bcd* RNAi embryos

To correspond the five *eve* stripes in *bcd* RNAi embryos with their WT counterparts, we introduced *eve* enhancer reporter constructs into the *bcd* RNAi embryo (Figure 3.3). The *eve* locus contains five enhancers that together drive seven stripes (Goto et al., 1989; Small et al., 1991; Small et al., 1996; Fujioka et al., 1999). To our knowledge, the stripe 4+6 and stripe 5 enhancer reporter constructs have not previously

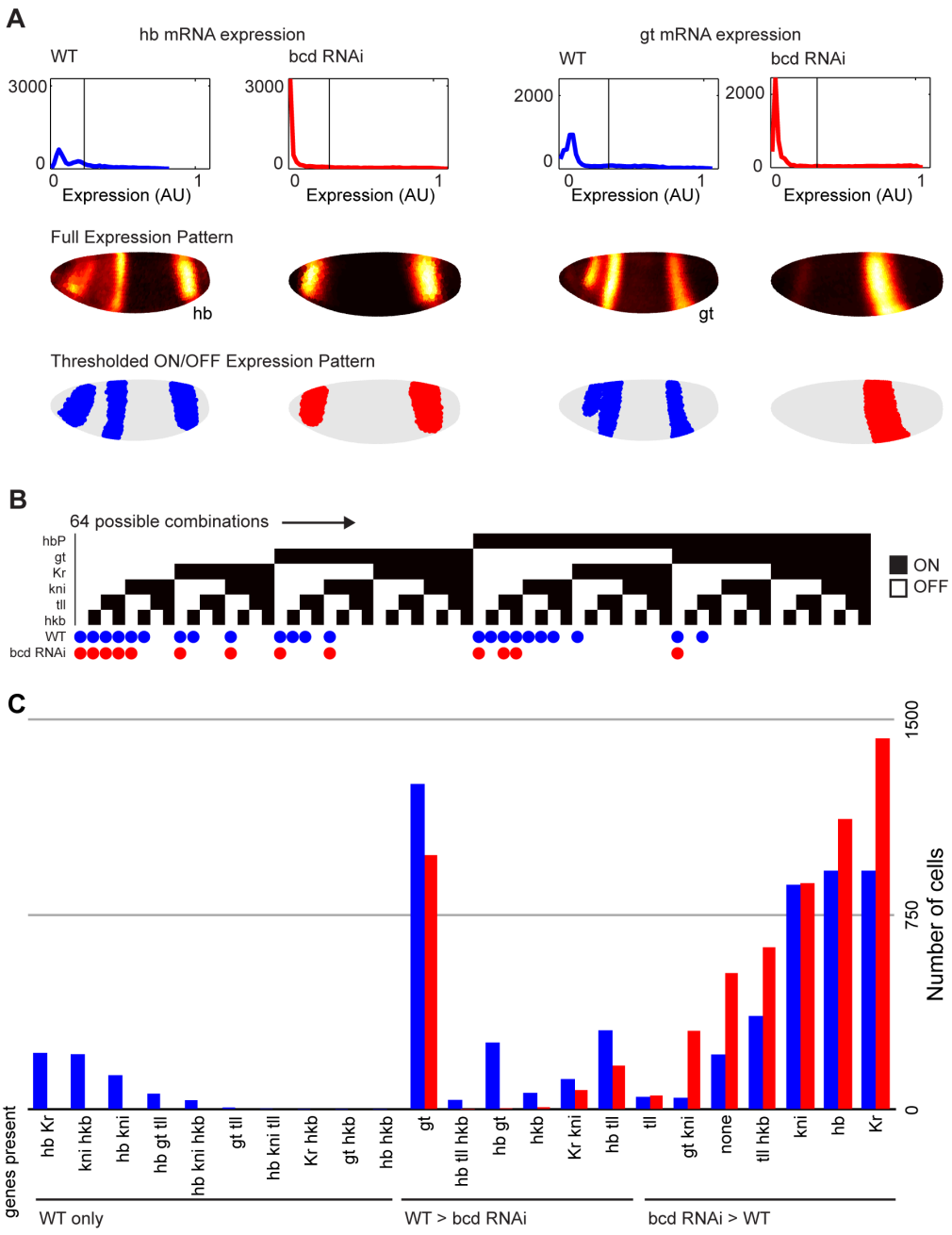


Figure 3.5: There are no new combinations of gap and terminal cellular gene expression profiles in *bcd* RNAi embryos.
(A) For each gene, we thresholded the expression pattern to find ON cells. Histograms of expression levels (top), heat maps of continuous expression patterns (middle), and the thresholded pattern (bottom). (B) Each column represents one of the 64 possible ON/OFF combinations of 6 genes. Filled squares indicate presence of a TF in a combination. There are 23 combinations present in WT (indicated by blue dots) and 13 combinations present in *bcd* RNAi (red dots). No combinations are present only in *bcd* RNAi. (C) The number of cells with each combination in each genotype.

been examined in *bcd* mutant embryos. Consistent with the literature, we found that the five *eve* stripes in *bcd* RNAi embryos correspond to *eve* stripes 3+7, *eve* stripes 4+6 and *eve* stripe 5 (Figure 3.3).

The gap gene expression patterns expand asymmetrically

A prominent feature of the *bcd* RNAi embryo is the asymmetric expansion of the gap gene expression patterns. In WT, the anterior boundary of *Kr* begins at 44% egg length (from the anterior) and the *Kr, kni, gt, hb* and *tll* patterns fill the remaining 56% of the embryo. In *bcd* RNAi, the anterior boundary of *Kr* shifts to begin at 27% egg length, and the gap gene domains expand to fill 73% of the embryo (Figure 3.4). While individual pattern shifts have been noted in the past (Struhl et al., 1989; Hulskamp et al., 1990; Kraut and Levine, 1991a; Rivera-Pomar et al., 1995), our measurements revealed that each pattern expanded by a different amount and had unique dynamics (Figure 3.4B). The asymmetric expansion of the gap genes is an important feature of our dataset that can be used to challenge other computational models of gap gene pattern formation and refinement (Jaeger et al., 2004a; Jaeger et al., 2004b; Bieler et al., 2011; Hengenius et al., 2011; Papatsenko and Levine, 2011).

Cellular gene expression profiles are canalized in *bcd* RNAi embryos

The *bcd* RNAi atlas provides a unique opportunity to examine how and when individual cells canalize cell fate following a strong genetic perturbation. Depletion of *bcd* leads to a complete replacement of the head and thorax with a second set of tail structures. This observation has been interpreted as strong canalization of cell fate

because discrete structures still form (Nüsslein-Volhard et al., 1987). However, this canalization could either be mediated by the segmentation network or by later compensatory processes such as apoptosis. In support of the latter hypothesis, *bcd* RNAi and mutant cuticles are much smaller than the WT cuticle and there is extensive apoptosis in *bcd* mutant embryos, which has been interpreted as selective elimination of misspecified cells (Werz et al., 2005).

To determine whether the segmentation network directly canalizes cell fate, we compared gene expression profiles of individual cells in *bcd* RNAi and WT embryos. The gene expression profile of a cell prefigures its eventual cell fate (Lehmann and Frohnhofer, 1989; Lawrence, 1992; St Johnston and Nüsslein-Volhard, 1992). We defined a cell fate as a binary gene expression profile where each gene is either ON or OFF. All the gene expression patterns in *bcd* RNAi are different from WT, but these perturbed patterns could, in principle, arise from new combinations of WT cell fates or from new cell fates. The emergence of a new combination of genes in *bcd* RNAi embryos would indicate a new cell fate.

We analyzed combinations of genes in the first zygotic layer of the network: the gap and terminal genes *Kr, hb, gt, kni, tll* and *hkb*. For each of the six regulators, we thresholded expression to classify cells as ON or OFF (Figure 3.5A, methods, Appendix B Table S4), giving 2^6 (64) possible ON/OFF combinations. There are three possible outcomes: a cell fate can be present only in WT, present only in *bcd* RNAi, or present in both.

By our simple definition, no new cell fates were created in *bcd* RNAi embryos. The third time point is illustrative of the general trend: of the possible 64 combinations,

23 cell fates were present in WT embryos (Figure 3.5B). In WT, there were no combinations with 4 or more genes and only 4/20 possible combinations of 3 genes, consistent with the strong mutual repression between some of the gap genes (Jaeger, 2011). All cells in *bcd* RNAi embryos belonged to 13 cell fates, all of which were present in WT. For the 10 WT cell fates lost in *bcd* RNAi, most cells were located in the anterior (Appendix B Figure S7). Virtually all of the shared cell fates changed in abundance between genotypes, with 6 more abundant in WT, and 7 more abundant in *bcd* RNAi (Figs 3.5C, Appendix B S8). We also compared gene expression profiles using Hb protein in place of *hb* mRNA because these patterns differ (Appendix B Figure S4). We again found that no new combinations arose in *bcd* RNAi (Appendix B Figure S9). We conclude that the dramatic changes in gap gene expression patterns result from changes in the proportion of cells with WT fates. This result supports the hypothesis that the first zygotic layer of the segmentation network directly canalizes cells towards normal fates.

When we examined other time points and other ON/OFF thresholds we continued to see canalization of cell fate. For T=3, over a wide range of ON/OFF thresholds, we found that all combinations of genes in *bcd* RNAi were also present in WT (Appendix B Figure S8). When we analyzed the data using Hb protein instead of Hb mRNA, there were no new combinations within a more limited range of thresholds, because the Hb protein data is harder to faithfully partition into ON and OFF cells (Appendix B Figure S9). At other times and thresholds, we sometimes found a handful of cells with a combination unique to *bcd* RNAi, but in virtually all cases, this combination existed in WT at other thresholds or adjacent time points. These failures to

detect combinations in WT likely arose from the higher background signal in the WT expression data (Figure 3.5A). At T=6 in *bcd* RNAi, this analysis detected a handful of cells with new combinations of the terminal patterns of *tll*, *hkb*, and *Kr*, but this effect is likely an artifact of the low quality WT T=6 *hkb* data, as visual inspection revealed these patterns overlap in WT (Appendix B Figure S7C).

To guard against the possibility that the fine registration using the *ftz* stain influenced our interpretation of the data, we repeated this analysis on a coarsely aligned atlas where embryos are aligned without the *ftz* fiduciary marker, and instead using only morphology (Appendix B Figure S10). For T=1-3, there were no additional combinations over the full range of thresholds. For T=4-6 in *bcd* RNAi, the analysis detected 3 additional combinations of *hb, tll*, *hkb*, and *Kr*, each with 1-2 cells at the boundaries of these patterns (Appendix B Figure S10). We conclude the fine registration did not confound our interpretation that cell fate is canalized in *bcd* RNAi embryos.

The pair-rule gene expression boundaries of *eve* and *ftz* are dynamically canalized in *bcd* RNAi embryos

The dynamic refinement of *eve* and *ftz* gene expression patterns differs between WT and *bcd* RNAi embryos. The primary pair-rule genes *eve* and *ftz* define the parasegment boundaries that later establish the compartment boundaries (Martinez-Arias and Lawrence, 1985; Lawrence, 1992). We chose to examine this layer of the network separately from the gap and terminal genes for three reasons: 1) *eve* and *ftz* are regulated by both the gap and maternal genes; 2) these genes may be sensitive to quantitative changes in relative levels of the gap genes not detected by our binary

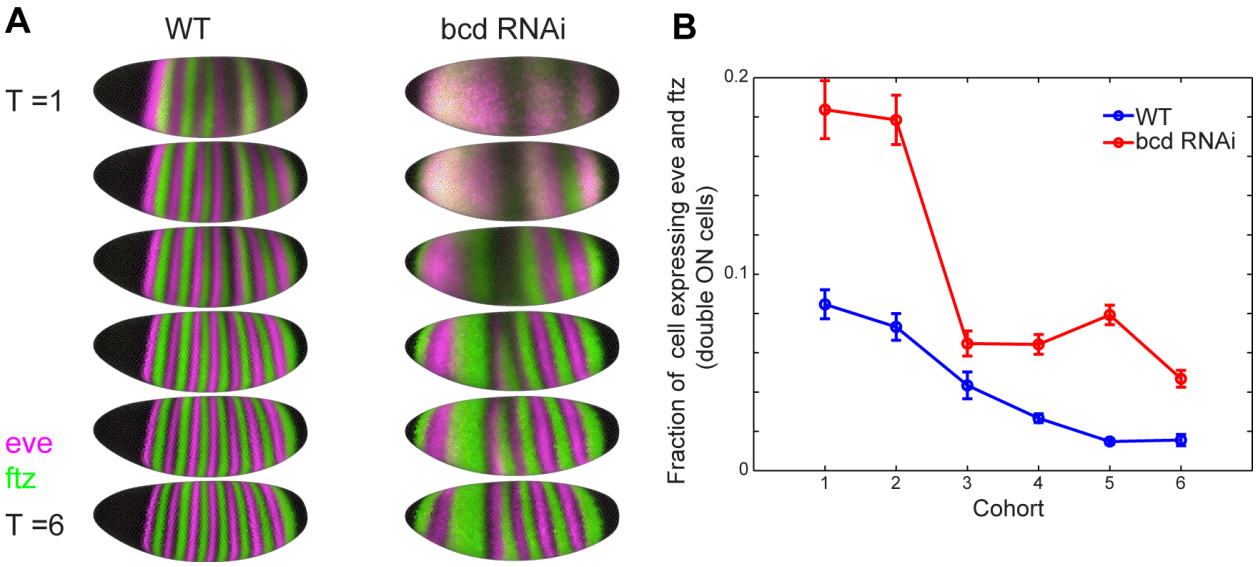


Figure 3.6: The mRNA expression domains of *eve* and *ftz* canalize over the blastoderm stage
(A) *eve* (pink) and *ftz* (green) mRNA patterns in the gene expression atlas for each cohort in WT and *bcd* RNAi. Cells with no expression appear black and cells expressing both *eve* and *ftz* appear white. (B) We quantified the fraction of cells that express both *eve* and *ftz* in individual embryos. For each embryo, we thresholded each expression pattern to be ON or OFF, and counted the fraction of cells where both genes were ON. WT n = 113, *bcd* RNAi n = 287.

combination analysis; and 3) while the initial gap gene patterns appear in stage 4, before we started collecting data, our stage 5 data captured the emergence and refinement of *eve* and *ftz* expression. In WT, these two gene expression patterns are mutually exclusive for virtually the entire blastoderm stage (Figure 3.6). In *bcd* RNAi, some individual embryos had extensive overlap of these two patterns. To quantify this difference, we examined individual embryos stained for *eve* and *ftz*, thresholded each gene separately to be ON or OFF, and counted the fraction of cells with both genes ON. In WT <10% of cells express both genes in the first temporal cohort and this fraction quickly dropped. In contrast, in *bcd* RNAi embryos in the first two cohorts, ~20% of cells expressed both *eve* and *ftz.* Beginning with the third cohort, the fraction of cells in *bcd*

RNAi embryos expressing both genes dropped sharply (Fig 3.6B). The shape of the trend does not depend on the threshold used to assign cells as ON and OFF or the *in situ* hapten. (Appendix B Figure S11). In individual embryos, early *eve/ftz* overlap resolves into mutually exclusive stripes, another manifestation of canalization in the segmentation network.

Discussion

We used the maternal Gal4 shRNA system to build a gene expression atlas of an embryo depleted of *bcd*, a maternally deposited transcription factor critical for anterior-posterior patterning of the *Drosophila melanogaster* embryo. This is the first 3D cellular resolution atlas of a genetic perturbation. The technical innovations we describe can be readily applied to build gene expression atlases of additional genetic perturbations. The specific dataset we present will be broadly useful for testing computational models of the segmentation network. Our data also reveals that the segmentation network directly canalizes cell fates: there were no new combinations of gap and terminal gene expression following *bcd* depletion. In the next layer of the network, the pair-rule genes *eve* and *ftz* initially overlapped, but eventually established sharp parasegment boundaries. We conclude that the anterior-posterior patterning network robustly specifies cell fates following the loss of a key maternal input.

Extension of the technique and utility of the dataset

Imaging techniques are uniquely positioned to capture how development unfolds in space and time (Megason, 2009). Gene expression atlases combine the spatiotemporal expression patterns of many genes in the same morphological framework, enabling computational analyses of gene regulatory networks (Jaeger et al., 2004b; Janssens et al., 2006; Papatsenko and Levine, 2008; Manu et al., 2009a; Manu et al., 2009b; Bieler et al., 2011; Gursky et al., 2011; Kozlov et al., 2012; Ilsley et al., 2013; Kim et al., 2013). Although imaging of single embryos is limited to 4-5 channels (Kosman et al., 2004; Dubuis et al., 2013), atlases can overcome this limit by

registering together data from many individual embryos stained for different genes but with a common fiduciary marker. This method is flexible and extendable: adding genes to the dataset simply requires co-staining with the fiduciary marker and imaging at high-resolution on a 2-photon microscope. The software for data visualization, image processing and atlas building are freely available (http://bdtnp.lbl.gov). Here we extended atlas building methods to genetically perturbed embryos, overcoming multiple technical challenges: collecting sufficient numbers of embryos, reducing population variability and building an appropriate registration template.

In the long term, other spatially and temporally resolved methods for quantitating gene expression may emerge. For example, mRNA sequencing can be performed on cryo-sliced embryos (Combs and Eisen, 2013) or *in situ* (Lee et al., 2014), though the latter remains prohibitively expensive for most labs. In the meantime, gene expression atlases are an accessible technique for examining the ensemble behavior of gene regulatory networks in single cells. By combining data for many genes into a unified morphology, atlases enable computational modeling and analysis. In particular, we anticipate that the asymmetric expansion of expression patterns in the *bcd* RNAi embryo will provide a useful challenge for computational models of the gap gene network.

Phenotypic variability in *bcd* RNAi embryos can be controlled, and may be useful in the future.

To build the gene expression atlas, we controlled the variability in the distribution of phenotypes in *bcd* RNAi embryos using specific experimental conditions and manual

curation. In the future, variability may be useful for studying other emergent properties of the network. For example, partially penetrant mutants helped constrain mathematical models of signal integration (Corson and Siggia, 2012). Some of our phenotypic variability may stem from inconsistent shRNA knockdown (Mohr and Perrimon, 2012), but the mutant data suggests much of the variability must emerge from the network response to *bcd*-depletion (Figure 3.1C). Increased variability in mutant phenotypes is common (Waddington, 1942; Wieschaus et al., 1984), and recent examination of gene expression patterns in *tll, Kr, kni,* and *Kr/kni* mutants concluded that there was more molecular variability in mutant embryos than in WT embryos (Janssens et al., 2013; Surkova et al., 2013). In *bcd* RNAi embryos, the variability in anterior *eve* stripe expression may explain the distribution of cuticle denticle bands. To enable study of this variability, we have made the data from individual embryos with split anterior *eve* stripes publicly available (depace.med.harvard.edu).

The segmentation network canalizes cell fate in *bcd* RNAi embryos

While it was known that cell fates were canalized in *bcd* mutant embryos by hatching, it was not clear if this canalization occurred immediately, due to the segmentation network, or later, due to downstream compensatory processes. The extensive apoptosis in *bcd* mutant embryos was proposed to be due to removal of misspecified cells (Werz et al., 2005). Misspecification can either imply the presence of too many cells of a given type, or the emergence of new types. Our analysis is the first direct evidence that the segmentation network prevents the creation of new cell fates in the absence of a maternal input. Several lines of evidence predicted this canalization

including: cytoplasmic transplantation experiments (Nüsslein-Volhard et al., 1987), the coordinated shifts in gene expression patterns following changes in *bcd* dosage (Driever and Nüsslein-Volhard, 1989; Struhl et al., 1989; Liu et al., 2013), and the molecular canalization of gene expression patterns in WT (Manu et al., 2009a; Manu et al., 2009b; Gursky et al., 2011). We have shown that canalization occurs early and strongly, resulting in changes in the abundance of most cell fates, but not the creation of new fates. This canalization is likely enforced by the abundant cross repression in the gap gene network (Jäckle et al., 1986; Kraut and Levine, 1991b; Jaeger et al., 2004b; Jaeger, 2011; Papatsenko and Levine, 2011; Sokolowski et al., 2012).

We propose that the increased apoptosis in *bcd* mutants does not eliminate cells with new fates, but instead compensates for enlarged compartments. The *eve* and *ftz* stripes set compartment size, and large compartments experience increased cell death (Namba et al., 1997; Hughes and Krause, 2001). The wide second *ftz* stripe (Figure 3.2) is approximately where the most apoptosis is observed in *bcd* mutant embryos (Werz et al., 2005). According to our analysis, cells undergoing apoptosis do not have new fates at the blastoderm stage. Rather, they reside in a compartment that is too large, and this increased compartment size may trigger cell death.

Dynamic canalization establishes sharp *eve* and *ftz* parasegment boundaries

We observed canalization of parasegment boundaries by examining the expression patterns of *eve* and *ftz* in individual *bcd* RNAi embryos. In the first two time points, ~20% of cells in *bcd* RNAi embryos express both *eve* and *ftz*, but this fraction later plummeted as the patterns resolved into mutually exclusive stripes. Similar early

overlaps of *eve* and *ftz* that resolve to mutually exclusive stripes have recently been reported in *Kr* mutant embryos (Surkova et al., 2013). This resolution of *eve* and *ftz* boundaries is likely mediated by direct repression of *ftz* by *eve* and indirect repression of *eve* by *ftz* through seven-stripe enhancers and may be a general feature of network responses to mutants (Jiang et al., 1991; Manoukian and Krause, 1992; Fujioka et al., 1996; Saulier-Le Drean et al., 1998; Nasiadka and Krause, 1999; Schroeder et al., 2011).

Conclusion

Reexamining a classic genetic perturbation at cellular resolution provided direct evidence that the segmentation network canalizes cell fates early and robustly. Our increased resolution also revealed subtle new features of the network, including the asymmetric expansion of the gap genes and the dynamic canalization of the parasegment boundaries. We anticipate that the *bcd* RNAi gene expression atlas will be useful to the developmental systems biology community by providing a cellular resolution dataset for testing computational models of how individual regulatory circuits position expression domains. These studies also lay important groundwork for our long-term goal of identifying the features of the network architecture that contribute to canalization of cell fate.

Materials and Methods

Fly Work

We depleted *bcd* with *UAS-shRNA-bcd* (TRiP GL00407) and the *maternal triple driver Gal4 (MTD-Gal4)* (Figure 3.1A). For reference, we used bcd^{12} (Bloomington 1755) (Frohnhöfer and Nüsslein-Volhard, 1986; Struhl et al., 1989). For controls we used *maternal-tubulin-Gal4 (mat-tub-Gal4)*, GL01320 *UAS-shRNA-bcd*, and TB184 *UAS-shRNA-GFP* (Appendix B Figure S2) (Neumuller et al., 2012; Staller et al., 2013). For future work with other maternal effect genes, we recommend *mat-tub-Gal4* (Appendix B Figure S2). We crossed virgin *MTD-Gal4/UAS-shRNA-bcd* females to males homozygous for reporter constructs. Enhancer were cloned into the *NotI* and *BglII* sites of *BOY-lacZ* and integrated in attP2 (Groth et al., 2004). Reporter sequences, original references and cloning primers are listed in Table S1.

Preparation of unhatched larval cuticles

Unhatched larval cuticles were mounted in lactic acid (Stern, 2000). We manually counted the number of denticle bands on each cuticle under dark field illumination, rounding up partial segments. For the majority of cuticles shown, a Z-stack of 2-4 images was computationally flattened with Helicon Focus (Helicon soft).

quantitative RT PCR

Embryos were collected for 2 hours and snap frozen in liquid nitrogen. We extracted RNA with Trizol and synthesized cDNA with superscript reverse transcriptase (Life). We used TaqMan probes (Life) with *actin* as a reference.

in situ hybridization

All RNA stains were performed as in (Fowlkes et al., 2011; Wunderlich et al., 2014). Briefly, embryos were collected over 4 hours at 25°C, dechorionated in bleach, fixed in formaldehyde/heptane for 25 minutes, dehydrated with methanol and stored in ethanol at -20°C. We used a digoxigenin (DIG) *ftz* probe, a dinitrophenol (DNP) probe against the gene of interest, and developed them sequentially with a tyramide amplification reaction (Perkin Elmer), with DIG in the coumarin channel and DNP in the Cy3 channel. We kept the amplification in the linear range, as described (Wunderlich et al., 2012). After RNase treatment overnight at 37°C, DNA was stained with Sytox green (Life). Embryos were dehydrated with ethanol, cleared with xylenes and mounted in DePeX (Electron Microscopy Sciences). To acquire Hb protein data, we stained embryos first with *ftz* DNP in the coumerin channel, and stained with guinea pig anti-Hb (a generous gift from John Reinitz (Chicago, Illinois, USA)) and goat anti-guinea pig AlexaFluor 555 (Life).

Image acquisition and manual data curation

We acquired Z-stacks with 2-photon excitation at 750 nm, with 1 micron increments, and simultaneously collected the 3 fluorescent channels. Protein stains were imaged in the same way. We use automated image processing to segment the nuclei and extract expression of the two genes in every cell, creating a pointcloud file for each embryo (Luengo Hendriks et al., 2006). We manually classified embryos into 6 cohorts: 0-3%, 4-8%, 9-25%, 26-50%, 51-75%, and 76-100% membrane invagination, which evenly divide the ~60 min blastoderm stage (Keränen et al., 2006). To remove individual embryos with weak phenotypes from the set of embryos laid by old mothers,

we manually inspected the *ftz* pattern. For time points 4-6, we removed embryos with a narrow second *ftz* stripe or an extra *ftz* stripe. For *eve* stains, we removed any embryos with a split anterior stripe.

Finding expression pattern boundaries

Pointcloud files were manipulated in MatLab (Mathworks) using the pointcloud toolbox (bdtnp.lbl.gov). For each embryo, we created line traces for 16 stripes around the dorsal ventral axis, and found the inflection point in each trace. Similar results were obtained when we used the half maximum of each line trace.

Building the *bcd* RNAi gene expression atlas

To account for a small increase in cell number and changes in cell density, we built a new morphological template for the *bcd* RNAi atlas using 1567 embryos (Fowlkes et al., 2008; Fowlkes et al., 2011). To build a new gene expression registration template we used 249 embryos stained only with DNP *ftz* probes. Embryo alignment is a two step process: first embryos are aligned coarsely to the morphological template, and second they are finely aligned to the registration template using the DIG *ftz* gene expression pattern. This fine scale alignment involves a local warping of each embryo described in detail in (Fowlkes et al., 2008). The degree of local warping tends to be higher at later time points when the patterns are sharper, leading to more reduction in variance, see supplement of Fowlkes et al., 2008. Each gene was normalized separately so that relative levels between time points were preserved, but the absolute levels between atlases are likely different. Cell density maps (Appendix B Figure S3) were generated using the demo_densities function in the pointcloud toolbox.

We have provided the *bcd* RNAi gene expression atlas and a bundled file containing all the individual embryos stained for *eve* and *ftz*, including those that were excluded from the atlas (depace.med.harvard.edu). We recommend visualizing the data with PointcloudXplore (bdtnp.lbl.gov) (Rübel et al., 2006).

Identifying combinations of ON and OFF cells

To create the binary gene expression profile of each cell, we thresholded *Kr, hb, kni, gt, tll,* and *hkb* mRNA at T=3. The ON/OFF threshold was calculated for each gene by creating a histogram, finding the peak of the OFF cell population and adding one standard deviation (e.g. in Appendix B Figure S9B). For *eve* and *ftz*, we determined thresholds for each gene in each embryo and recorded the fraction of cells expressing both. Using the published stains of WT embryos, we found that swapping the haptens (DNP/DIG) did not change the fraction of double ON cells (Appendix B Figure S9A).

Acknowledgements

We thank John Reinitz for the spectacular Hb antibody, Norbert Perrimon for discussions of phenotypic variability, TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) for the maternal drivers and *UAS-shRNA-bcd* line, Miki Fujioka for providing the *eve late seven stripe* enhancer reporter and *eve* whole locus reporter lines, and Tara Lydiard-Martin for making the other *eve* enhancer reporter lines. We thank Becky Ward, Ashley Wolf and members of the DePace Lab for feedback on the manuscript and Ben Vincent for extensive comments on the manuscript. This work was supported in part by the Harvard Herchel Smith Fellowship (MVS), NSF DBI-1053036 (CCF), Jane Coffin Childs Memorial Fund for Medical Research (ZBW), NIH K99HD073191 (ZBW), and NIH U01 GM103804-01A1 (AHD).

Author Contributions

MVS and AHD designed the study. MVS and MDJB performed the experiments. CCF built the templates for the gene expression atlas. ZBW and JE processed raw image stacks into pointcloud files. MVS analyzed the data with input from AHD and ZBW. MVS and AHD wrote the paper.

References

Bieler, J., Pozzorini, C. and Naef, F. (2011). Whole-Embryo Modeling of Early Segmentation in Drosophila Identifies Robust and Fragile Expression Domains. *Biophysj* **101**, 287-296.

Combs, P. A. and Eisen, M. B. (2013). Sequencing mRNA from cryo-sliced Drosophila embryos to determine genome-wide spatial patterns of gene expression. *PLoS ONE* **8**, e71820.

Corson, F. and Siggia, E. D. (2012). Geometry, epistasis, and developmental patterning. *Proc Natl Acad Sci U S A* **109**, 5568-5575.

Davidson, E. H. (2010). The regulatory genome: gene regulatory networks in development and evolution.

Driever, W. and Nüsslein-Volhard, C. (1989). The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature* **337**, 138-143.

Dubuis, J. O., Samanta, R. and Gregor, T. (2013). Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol Syst Biol* **9**, 639.

Fowlkes, C. C., Eckenrode, K. B., Bragdon, M. D., Meyer, M., Wunderlich, Z., Simirenko, L., Luengo Hendriks, C. L., Keranen, S. V., Henriquez, C., Knowles, D. W. et al. (2011). A conserved developmental patterning network produces quantitatively different output in multiple species of Drosophila. *PLoS Genet* **7**, e1002346.

Fowlkes, C. C., Hendriks, C. L. L., Keränen, S. V. E., Weber, G. H., Rübel, O., Huang, M.-Y., Chatoor, S., DePace, A. H., Simirenko, L., Henriquez, C. et al. (2008). A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. *Cell* **133**, 364-374.

Frasch, M. and Levine, M. (1987). Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in Drosophila. *Genes & development* **1**, 981-995.

Frohnhöfer, H. G. and Nüsslein-Volhard, C. (1986). Organization of anterior pattern in the Drosophila embryo by the maternal gene bicoid. *Nature* **324**, 120-125.

Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T. and 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 (Cambridge, England)* **126**, 2527-2538.

Fujioka, M., Miskiewicz, P., Raj, L., Gulledge, A. A., Weir, M. and Goto, T. (1996). Drosophila Paired regulates late even-skipped expression through a composite binding site for the paired domain and the homeodomain. *Development (Cambridge, England)* **122**, 2697-2707.

Goto, T., Macdonald, P. and Maniatis, T. (1989). Early and late periodic patterns of< i> even skipped</i> expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413-422.

Groth, A. C., Fish, M., Nusse, R. and Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775-1782.

Gursky, V. V., Panok, L., Myasnikova, E. M., Manu, Samsonova, M. G., Reinitz, J. and Samsonov, A. M. (2011). Mechanisms of gap gene expression canalization in the Drosophila blastoderm. *BMC Systems Biology* **5**, 118.

He, X., Samee, M. A. H., Blatti, C. and Sinha, S. (2010). Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Computational Biology* **6**,

Hengenius, J. B., Gribskov, M., Rundell, A. E., Fowlkes, C. C. and Umulis, D. M. (2011). Analysis of Gap Gene Regulation in a 3D Organism-Scale Model of the Drosophila melanogaster Embryo. *PLoS ONE* **6**, e26797.

Hooper, K. L., Parkhurst, S. M. and Ish-Horowicz, D. (1989). Spatial control of hairy protein expression during embryogenesis. *Development* **107**, 489-504.

Hughes, S. C. and Krause, H. M. (2001). Establishment and maintenance of parasegmental compartments. *Development* **128**, 1109-1118.

Hulskamp, M., Pfeifle, C. and Tautz, D. (1990). A morphogenetic gradient of hunchback protein organizes the expression of the gap genes Kruppel and knirps in the early Drosophila embryo. *Nature* **346**, 577-580.

Hulskamp, M., Schroder, C., Pfeifle, C., Jackle, H. and Tautz, D. (1989). Posterior segmentation of the Drosophila embryo in the absence of a maternal posterior organizer gene. *Nature* **338**, 629-632.

Ilsley, G. R., Fisher, J., Apweiler, R., DePace, A. H. and Luscombe, N. M. (2013). Cellular resolution models for even skipped regulation in the entire Drosophila embryo. *Elife* **2**, e00522.

Irish, V., Lehmann, R. and Akam, M. (1989). The Drosophila posterior-group gene nanos functions by repressing hunchback activity. *Nature* **338**, 646-648.

Jäckle, H., Tautz, D., Schuh, R., Seifert, E. and Lehmann, R. (1986). Cross-regulatory interactions among the gap genes of Drosophila.

Jaeger, J. (2011). The gap gene network. *Cellular and Molecular Life Sciences* **68**, 243-274.

Jaeger, J., Blagov, M., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Surkova, S., Vanario-Alonso, C. E., Samsonova, M., Sharp, D. H. et al. (2004). Dynamical analysis of regulatory interactions in the gap gene system of Drosophila melanogaster. *Genetics* **167**, 1721-1737.

Jaeger, J., Manu and Reinitz, J. (2013). Drosophila blastoderm patterning. *Current opinion in genetics & development*

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Vanario-Alonso, C. E., Samsonova, M. et al. (2004). Dynamic control of positional information in the early Drosophila embryo. *Nature* **430**, 368-371.

Janssens, H., Crombach, A., Richard Wotton, K., Cicin-Sain, D., Surkova, S., Lu Lim, C., Samsonova, M., Akam, M. and Jaeger, J. (2013). Lack of tailless leads to an increase in expression variability in Drosophila embryos. *Developmental biology*

Janssens, H., Hou, S., Jaeger, J., Kim, A.-R., Myasnikova, E., Sharp, D. and Reinitz, J. (2006). Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. *Nature genetics* **38**, 1159-1165.

Jiang, J., Hoey, T. and Levine, M. (1991). Autoregulation of a segmentation gene in Drosophila: combinatorial interaction of the even-skipped homeo box protein with a distal enhancer element. *Genes Dev* **5**, 265-277.

Keränen, S. V. E., Fowlkes, C. C., Luengo Hendriks, C. L., Sudar, D., Knowles, D. W., Malik, J. and Biggin, M. D. (2006). Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution II: dynamics. *Genome Biology* **7**, R124.

Kim, A.-R., Martinez, C., Ionides, J., Ramos, A. F., Ludwig, M. Z., Ogawa, N., Sharp, D. H. and Reinitz, J. (2013). Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic. *PLoS Genetics* **9**, e1003243.

Kosman, D., Mizutani, C. M., Lemons, D., Cox, W. G., McGinnis, W. and Bier, E. (2004). Multiplex detection of RNA expression in Drosophila embryos. *Science* **305**, 846-846.

Kozlov, K., Surkova, S., Myasnikova, E., Reinitz, J. and Samsonova, M. (2012). Modeling of gap gene expression in Drosophila Kruppel mutants. *PLoS Computational Biology* **8**, e1002635.

Kraut, R. and Levine, M. (1991). Spatial regulation of the gap gene giant during Drosophila development. *Development* **111**, 601-609.

Kraut, R. and Levine, M. S. (1991). Mutually repressive interactions between the gap genes giant and Krüppel define middle body regions of the Drosophila embryo. *Development (Cambridge, England)* **111**, 611-621.

Lawrence, P. A. (1992). The making of a fly: the genetics of animal design.

Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Yang, J. L., Ferrante, T. C., Terry, R., Jeanty, S. S., Li, C., Amamoto, R. et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360-1363.

Lehmann, R. and Frohnhofer, H. G. (1989). Segmental polarity and identity in the abdomen of Drosophila is controlled by the relative position of gap gene expression. *Development* **107 Suppl**, 21-29.

Liu, F., Morrison, A. H. and Gregor, T. (2013). Dynamic interpretation of maternal inputs by the Drosophila segmentation gene network. *Proceedings of the National Academy of Sciences* **110**, 6724-6729.

Luengo Hendriks, C. L., Keränen, S. V., Biggin, M. D. and Knowles, D. W. (2007). Automatic channel unmixing for high-throughput quantitative analysis of fluorescence images. *Optics express* **15**, 12306-12317.

Luengo Hendriks, C. L., Keränen, S. V. E., Fowlkes, C. C., Simirenko, L., Weber, G. H., DePace, A. H., Henriquez, C., Kaszuba, D. W., Hamann, B., Eisen, M. B. et al. (2006). Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. *Genome Biology* **7**, R123.

Manoukian, A. S. and Krause, H. M. (1992). Concentration-dependent activities of the even-skipped protein in Drosophila embryos. *Genes Dev* **6**, 1740-1751.

Manu, Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M. et al. (2009). Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. *PLoS Computational Biology* **5**, e1000303.

Manu, Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M. et al. (2009). Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. *PLoS biology* **7**, e1000049.

Martinez-Arias, A. and Lawrence, P. A. (1985). Parasegments and compartments in the Drosophila embryo. *Nature* **313**, 639-642.

Megason, S. G. (2009). In toto imaging of embryogenesis with confocal time-lapse microscopy. *Methods Mol Biol* **546**, 317-332.

Mohr, S. E. and Perrimon, N. (2012). RNAi screening: new approaches, understandings, and organisms. *Wiley Interdiscip Rev RNA* **3**, 145-158.

Namba, R., Pazdera, T. M., Cerrone, R. L. and Minden, J. S. (1997). Drosophila embryonic pattern repair: how embryos respond to bicoid dosage alteration. *Development* **124**, 1393-1403.

Nasiadka, A. and Krause, H. M. (1999). Kinetic analysis of segmentation gene interactions in Drosophila embryos. *Development* **126**, 1515-1526.

Neumuller, R. A., Wirtz-Peitz, F., Lee, S., Kwon, Y., Buckner, M., Hoskins, R. A., Venken, K. J., Bellen, H. J., Mohr, S. E. and Perrimon, N. (2012). Stringent analysis of gene function and protein-protein interactions using fluorescently tagged genes. *Genetics* **190**, 931-940.

Ni, J. Q., Zhou, R., Czech, B., Liu, L. P., Holderbaum, L., Yang-Zhou, D., Shim, H. S., Tao, R., Handler, D., Karpowicz, P. et al. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat Methods* **8**, 405-407.

Nüsslein-Volhard, C., Frohnhöfer, H. G. and Lehmann, R. (1987). Determination of anteroposterior polarity in Drosophila. *Science* **238**, 1675-1681.

Papatsenko, D. and Levine, M. S. (2008). Dual regulation by the Hunchback gradient in the Drosophila embryo. *Proc Natl Acad Sci U S A* **105**, 2901-2906.

Papatsenko, D. and Levine, M. S. (2011). The Drosophila gap gene network is composed of two parallel toggle switches. *PLoS ONE* **6**, e21145.

Pisarev, A., Poustelnikova, E., Samsonova, M. and Reinitz, J. (2009). FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic acids research* **37**, D560-D566.

Poustelnikova, E., Pisarev, A., Blagov, M., Samsonova, M. and Reinitz, J. (2004). A database for management of gene expression data in situ. *Bioinformatics* **20**, 2212-2221.

Reinitz, J. and Sharp, D. H. (1995). Mechanism of eve stripe formation. *Mechanisms of development* **49**, 133-158.

Ren, X., Sun, J., Housden, B. E., Hu, Y., Roesel, C., Lin, S., Liu, L. P., Yang, Z., Mao, D., Sun, L. et al. (2013). Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. *Proc Natl Acad Sci U S A* **110**, 19012-19017.

Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H. and Jackle, H. (1995). Activation of posterior gap gene expression in the Drosophila blastoderm. *Nature* **376**, 253-256.

Rübel, O., Weber, G. H., Keränen, S. V. E., Fowlkes, C. C., Hendriks, C. L., Simirenko, L., Shah, N. Y., Eisen, M. B., Biggin, M. D. and Hagen, H. (2006). PointCloudXplore: Visual analysis of 3D gene expression data using physical

views and parallel coordinates. **Proceedings of the Eighth Joint Eurographics/ IEEE VGTC conference on Visualization**, 203-210.
Samee, M. A. H. and Sinha, S. (2013). Evaluating thermodynamic models of enhancer activity on cellular resolution gene expression data. *Methods* 1-12.
Saulier-Le Drean, B., Nasiadka, A., Dong, J. and Krause, H. M. (1998). Dynamic changes in the functions of Odd-skipped during early Drosophila embryogenesis. *Development* **125**, 4851-4861.
Schroeder, M. D., Greer, C. and Gaul, U. (2011). How to make stripes: deciphering the transition from non-periodic to periodic patterns in Drosophila segmentation. *Development (Cambridge, England)* **138**, 3067-3078.
Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature* **451**, 535-540.
Small, S., Blair, A. and Levine, M. S. (1996). Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. *Developmental biology* **175**, 314-324.
Small, S., Kraut, R., Hoey, T., Warrior, R. and Levine, M. S. (1991). Transcriptional regulation of a pair-rule stripe in Drosophila. *Genes & Development* **5**, 827-839.
Sokolowski, T. R., Erdmann, T. and Ten Wolde, P. R. (2012). Mutual repression enhances the steepness and precision of gene expression boundaries. *PLoS Computational Biology* **8**, e1002654.
St Johnston, D. and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the Drosophila embryo. *Cell* **68**, 201-219.
Staller, M. V., Yan, D., Randklev, S., Bragdon, M. D., Wunderlich, Z. B., Tao, R., Perkins, L. A., DePace, A. H. and Perrimon, N. (2013). Depleting gene activities in early Drosophila embryos with the “maternal-Gal4-shRNA” system. *Genetics* **193**, 51-61.
Stern, D. L., and E. Sucena. (2000). Preparation of larval and adult cuticles for light microscopy. In *Drosophila protocols* pp. 601-615.
Struhl, G., Struhl, K. and Macdonald, P. M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell* **57**, 1259-1273.
Surkova, S., Golubkova, E., Manu, Panok, L., Mamon, L., Reinitz, J. and Samsonova, M. (2013). Quantitative dynamics and increased variability of segmentation gene expression in the Drosophila Krüppel and knirps mutants. *Developmental biology* **376**, 99-112.
Tautz, D. (1988). Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres. *Nature* **332**, 281-284.
Umulis, D. M. and Othmer, H. G. (2012). The importance of geometry in mathematical models of developing systems. *Curr Opin Genet Dev* **22**, 547-552.
Waddington, C. H. (1942). Canalization of Development and the inheritance of acquired characters. *Nature* 563-565.
Waddington, C. H. (1957). The strategy of the genes; a discussion of some aspects of theoretical biology. **Allen & Unwin**,
Werz, C., Lee, T. V., Lee, P. L., Lackey, M., Bolduc, C., Stein, D. S. and Bergmann, A. (2005). Mis-specified cells die by an active gene-directed process, and inhibition

of this death results in cell fate transformation in Drosophila. *Development* **132**, 5343-5352.

Wieschaus, E., Nusslein-Volhard, C. and Kluding, H. (1984). Kruppel, a gene whose activity is required early in the zygotic genome for normal embryonic segmentation. *Dev Biol* **104**, 172-186.

Wunderlich, Z., Bragdon, M. D. and DePace, A. H. (2014). Comparing mRNA levels using in situ hybridization of a target gene and co-stain. *Methods*

Wunderlich, Z., Bragdon, M. D., Eckenrode, K. B., Lydiard-Martin, T., Pearl-Waserman, S. and DePace, A. H. (2012). Dissecting sources of quantitative gene expression pattern divergence between Drosophila species. *Mol Syst Biol* **8**, 604.

Wunderlich, Z. and DePace, A. H. (2011). Modeling transcriptional networks in Drosophila development at multiple scales. *Curr Opin Genet Dev* **21**, 711-718.

Chapter 4: Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo

Max V. Staller, Ben J. Vincent, Meghan D.J. Bragdon, Zeba B. Wunderlich, Angela H. DePace

Department of Systems Biology, Harvard Medical School, Boston, Massachusetts

Key Words:

enhancer, computational model, dual transcriptional regulators, Drosophila development

Abstract

When bifunctional transcription factors activate and repress target genes within the same cell, these opposing activities must be encoded in regulatory DNA. Here, we use cellular resolution gene expression data and computational modeling to investigate Hunchback (Hb) bifunctionality in *Drosophila* embryogenesis. Previous computational models predicted that Hb both activated and repressed the enhancer controlling *even-skipped (eve)* stripes 3 and 7 *(eve3+7)*. We tested this hypothesis by measuring and modeling *eve* expression under multiple genetic perturbations and found that the *eve3+7* enhancer could not explain endogenous stripe 7 behavior. To explain this discrepancy, we measured the response of an extended *eve* stripe 2 enhancer that drives expression of *eve* stripes 2 and 7 *(eve2+7)*. We found that the behavior of endogenous stripe 7 is explained by the combined behavior of both enhancers, *eve3+7* and *eve2+7*. Bifunctionality arises from Hb activating the *eve2+7* enhancer and repressing the *eve3+7* enhancer. This pair can thus be considered “shadow enhancers” that both direct *eve* stripe 7, but respond to Hb in opposite ways. This example may illustrate a general way of encoding bifunctional regulation in the genome.

Introduction

Transcription factors (TFs) are typically categorized as activators or repressors, but many TFs can act bifunctionally by both activating and repressing expression of their target genes (Shore and Nasmyth, 1987; Struhl et al., 1992; Sauer and Jäckle, 1993; Deng et al., 2010; Di Stefano et al., 2014). In cases where TFs activate and repress targets in the same cells, bifunctionality must be locally encoded in regulatory DNA sequence. Determining the regulatory DNA sequence features that control TF bifunctionality will advance two fundamental challenges of decoding transcriptional networks: predicting expression patterns from regulatory sequence, and deciphering how network topology dictates systems-level properties including gene expression precision and robustness to genetic and environmental perturbations.

Here, we investigate how TF bifunctionality is encoded in regulatory DNA using a classic example: the *Drosophila* segmentation gene, *hunchback (hb)* (Small et al., 1991; Zuo et al., 1991; Struhl et al., 1992; Schulz and Tautz, 1994; Arnosti et al., 1996; Small et al., 1996). Hb both activates and represses the seven-striped *even-skipped (eve)* gene expression pattern by acting on multiple enhancers, genomic regions responsible for tissue specific gene expression (Levine et al., 2014). Hb activates *eve* stripes 1 and 2 and represses stripes 3,4,5,6 and 7 (Small et al., 1991; Small et al., 1996; Fujioka et al., 1999). With our collaborators, we recently developed models of *eve* stripe regulation that suggested, consistent with previous models, that Hb bifunctionally regulates expression of *eve* stripes 3 and 7 (Figure 4.1) (Papatsenko and Levine, 2008; Ilsley et al., 2013). Here, we test this hypothesis further using quantitative expression data in genetically perturbed embryos. We focus on whether Hb acts as both an activator and a

repressor in the annotated *eve3+7* enhancer. We measured both the expression driven by the *eve3+7* enhancer in a reporter construct and the endogenous *eve* expression pattern at high resolution in embryos with perturbed Hb expression levels. We then used these data to challenge our computational models.

We found that Hb bifunctionality is encoded by separate enhancers that both direct *eve* stripe 7 expression. The first is the annotated *eve3+7* enhancer, where Hb acts as a repressor. The second is an extended piece of regulatory DNA encompassing the minimal *eve* stripe 2 *(eve2)* enhancer that drives expression of *eve* stripes 2 and 7 *(eve2+7)* (Goto et al., 1989; Harding et al., 1989; Small et al., 1991). In the *eve2+7* enhancer, Hb acts as an activator. Therefore, *eve* stripe 7 is controlled by a pair of shadow enhancers, separate sequences in a locus that drive overlapping spatiotemporal patterns (Barolo, 2012). Notably, these shadow enhancers respond to Hb in opposite ways and therefore use different regulatory logic. The separation of activation and repression into distinct enhancers may be a general mechanism of encoding TF bifunctionality in the genome.

Results

eve enhancer reporter patterns do not match the endogenous *eve* pattern

To determine if Hb bifunctionality is encoded in the annotated *eve3+7* enhancer, we compared the pattern of the endogenous *eve* stripes to the pattern driven by a *lacZ* reporter construct in two genetic backgrounds. We refer to the resulting patterns throughout the manuscript as “the *eve3+7* reporter pattern” and “the endogenous pattern” (Figure 4.2). We examined both WT embryos and embryos where expression of Hb had been perturbed by removing *bcd*, one of its key regulators, using RNAi *(bcd* RNAi embryos) (Staller et al., 2014). We quantitatively measured expression patterns at cellular resolution using *in situ* hybridization, 2-photon microscopy, and an image processing toolkit developed specifically for *Drosophila* embryos (methods) (Keränen et al., 2006; Luengo Hendriks et al., 2006). We then averaged these data together into gene expression atlases (Fowlkes et al., 2008). Importantly, the reporter construct isolates the activity of the annotated *eve3+7* enhancer while the endogenous pattern integrates the activity of the whole locus.

Our high resolution measurements revealed discrepancies between the endogenous pattern and the *eve3+7* reporter pattern (Figure 4.2). In WT embryos, the *eve3+7* reporter pattern overlaps the endogenous *eve* stripes, but these stripes are broader, have uneven levels, and the peaks lie posterior to the endogenous peaks (Figure 4.2). These discrepancies were more pronounced in *bcd* RNAi embryos than in WT embryos, especially for the anterior stripe (Figure 4.2). When we tested reporters for other *eve* enhancers, we also found that they did not fully recapitulate the endogenous pattern (Appendix C Figs S1, S2).

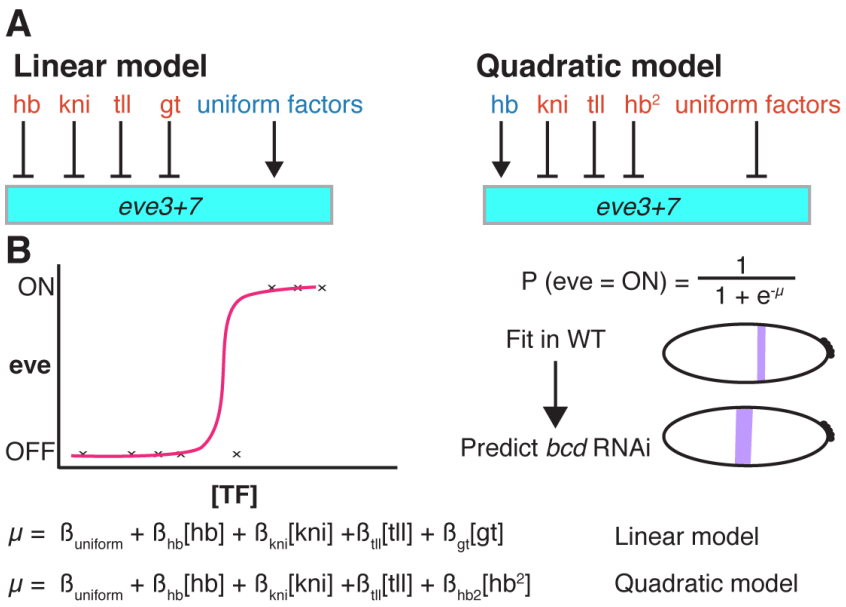


Figure 4.1: The linear and quadratic models formalize two alternative regulator sets for *eve* stripes 3 and 7.
(A) The linear model includes repression (red) by Hb, *knirps (kni), giant (gt),* and *tailless (tll)* and activation (blue) by a constant term that represents spatially uniform factors. The quadratic model includes activation by a linear Hb term and repression by a quadratic Hb term, *kni*, *tll*, and uniform factors. (B) A schematic of the logistic regression framework. Logistic regression calculates the probability the target will be ON based on a linear combination of the concentrations of regulators (μ). We fit models in WT and use the perturbed regulator gene expression patterns to predict the perturbed *eve* patterns in *bcd* RNAi embryos.

To test if the discrepancies between the *eve3+7* reporter pattern and the endogenous pattern resulted from differences in *eve* and *lacZ* mRNA half-lives, we measured the expression driven by an *eve* locus BAC reporter where the coding sequence had been replaced with *lacZ* (a generous gift from Miki Fujioka). In both WT and *bcd* RNAi embryos, the peak positions and widths of the BAC reporter pattern were more faithful to the endogenous *eve* pattern, but still did not match exactly (Figure 4.2, Appendix C S1, S2). Differences between the endogenous and BAC reporter patterns must arise from differences in the transcripts. Differences between the BAC reporter

and the isolated enhancer reporter patterns must arise from regulatory DNA not included in the reporter constructs. These data indicate that the *eve3+7* reporter construct is missing relevant regulatory DNA. Additional regulatory DNA in the endogenous locus may respond to other TFs or respond to the same TFs differently, leading to differences in how the endogenous pattern is computed from the concentrations of its regulators. We therefore hypothesized that the endogenous locus and the *eve3+7* enhancer perform different computations to produce *eve* stripes 3 and 7. We next tested this hypothesis using computational models of *eve* regulation.

Computational models suggest that Hb activates and represses endogenous *eve* stripes 3 and 7, but only represses the *eve3+7* enhancer

With our collaborators, we previously identified two empirical computational models for the expression of *eve* stripes 3 and 7. (Figure 4.1) (Ilsley et al., 2013). These models used logistic regression to directly relate the concentrations of input regulators to output expression in single cells by fitting a parameter for each regulator that reflects both TF strength and TF-DNA interactions (Ilsley et al., 2013). In the linear model, Hb has one parameter and only represses. In the quadratic model, Hb has two parameters, and activates at low concentrations while repressing at high concentrations (Ilsley et al., 2013). Both models perform equally well in WT embryos, but they make different predictions under genetic perturbation. Specifically, the quadratic model predicts published data from *hb* ventral misexpression experiments (see below).

To investigate patterns driven by individual enhancers, these models were fit on data parsed from the entire endogenous *eve* pattern because data for enhancer

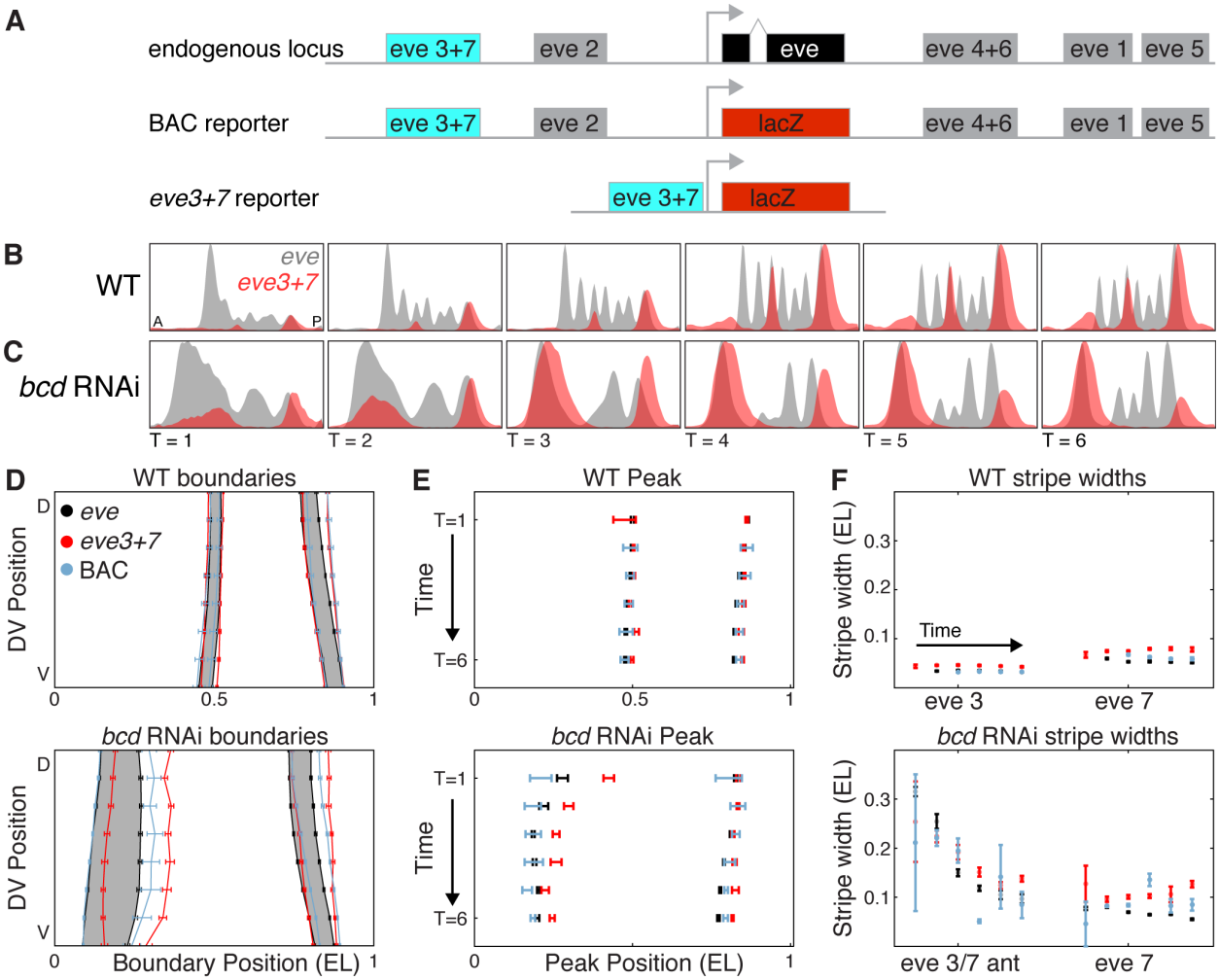


Figure 4.2:The *eve3+7* reporter pattern differs from the endogenous pattern. (A) The *eve* locus contains 5 annotated primary stripe enhancers. The endogenous pattern integrates the activity of the whole locus. The BAC reporter construct also integrates the activity of the whole locus, but the transcript is the same as the *eve3+7* reporter construct. The *eve3+7* reporter construct isolates the activity of the annotated enhancer sequence. (B) WT expression patterns are represented as line traces where anterior-posterior (A-P) position is plotted on the X-axis with expression level on the Y-axis for a lateral strip of the embryo. Endogenous *eve* pattern (gray), *eve3+7* reporter pattern (red). The reporter pattern was manually scaled to match the level of the endogenous pattern. (C) Line traces in *bcd* RNAi embryos. (D) The boundaries of the endogenous pattern (gray), the *eve3+7* reporter pattern (red), and the BAC reporter pattern (blue) at T=3. All error bars are the standard error of the mean. The BAC reporter pattern is more faithful to the endogenous pattern than the *eve3+7* reporter pattern, especially in the anterior of *bcd* RNAi embryos (eve 3/7 ant). The endogenous pattern is shaded for visual clarity. (E) Peak positions of stripes 3 and 7, calculated from the line traces in B and C. The *eve3+7* reporter pattern shows better agreement to the endogenous pattern in WT than in *bcd* RNAi embryos. (F) Stripe widths, calculated from the inflection point of the line traces in B and C. The *eve3+7* reporter pattern is wider than the corresponding endogenous pattern.

reporters was not available (Ilsley et al., 2013). To relate the results to individual enhancers, we employed a standard assumption: the endogenous expression of *eve* stripes 3 and 7 could be attributed to the activity of the annotated *eve3+7* enhancer. Here, we test this assumption explicitly by measuring and modeling the *eve3+7* reporter and endogenous patterns separately.

We compared the performance of the linear and quadratic models in WT and *bcd* RNAi embryos. As input regulators we used Hb protein and *gt, tll* and *kni* mRNA, and we used thresholded endogenous or *eve3+7* reporter mRNA data as our target output patterns (Fig 4.1, methods). All of these regulators, especially Hb, are perturbed in *bcd* RNAi embryos (Appendix C Figure S3) (Staller et al., 2014). We report our modeling of the third time point, which is representative of results for other time points (Appendix C Figure S6), and evaluated model performance by computing the area under the receiver operating characteristic curve (AUC) (Swets, 1988).

We first addressed the endogenous *eve* pattern: we fit our models in WT embryos and used the resulting parameters to predict expression in *bcd* RNAi embryos. In this case, we found that the quadratic model more precisely predicted the perturbed endogenous *eve* pattern. Both models correctly predicted the positional shifts of stripe 7 and a wide anterior stripe, but the quadratic model was more accurate than the linear model (AUC_{linear} = 0.93, AUC_{quad} = 0.98) (Figure 4.3F, Appendix C Figure S4). These analyses indicated that the quadratic model captured the activity of the whole locus by allowing Hb to both activate and repress *eve* stripes 3 and 7.

We next addressed the *eve3+7* reporter pattern: again, we fit our models in WT embryos and used the resulting parameters to predict expression in *bcd* RNAi embryos.

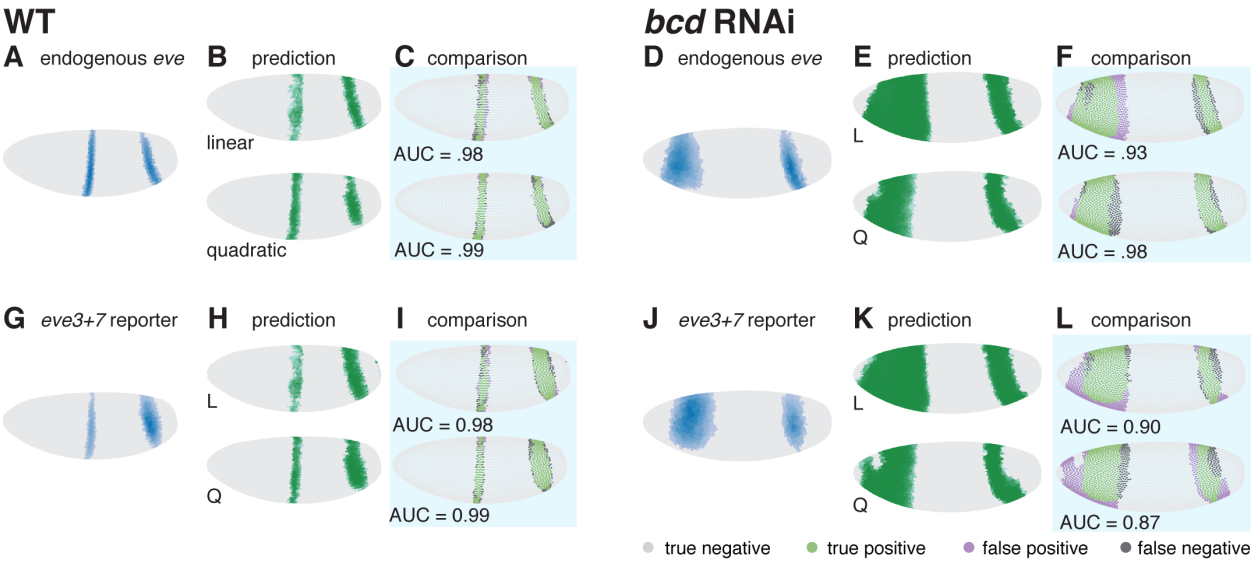


Figure 4.3: In *bcd* RNAi embryos, the quadratic model more accurately predicts the endogenous pattern, and the linear model more accurately predicts the *eve3+7* reporter pattern

(A) The endogenous *eve* pattern in WT embryos. Cells with expression below an ON/OFF threshold (methods) are plotted in gray. For cells above this threshold, darker color indicates higher level. (B) The predictions of the linear and quadratic models in WT embryos. (C) Comparison of model predictions to the endogenous pattern in WT embryos. Green cells are true positives, purple cells are false positives, dark gray cells are false negatives, and light gray cells are true negatives. For visualization, the threshold is set to 80% sensitivity, but the AUC metric quantifies performance over all thresholds. (D) The endogenous *eve* pattern in *bcd* RNAi embryos. (E) The predictions of the linear (L) and quadratic (Q) models in *bcd* RNAi embryos. (F) Comparison of model predictions to the endogenous pattern in *bcd* RNAi embryos. The quadratic model more accurately predicts the endogenous pattern in *bcd* RNAi embryos. (G-L) Same as A-F, respectively, for the *eve3+7* reporter pattern. The linear model predicts the *eve3+7* reporter pattern more accurately in *bcd* RNAi embryos. Model parameters are in Table S1.

To our surprise, the linear model was more accurate than the quadratic model in this case (AUC_{linear} = 0.90, AUC_{quad} = 0.87) (Figure 4.3L). Although neither model captured the dorsal-ventral modulation of the pattern, the linear model accurately predicted the posterior boundary of the anterior stripe. We controlled for several factors that may

confound model performance. We assessed sensitivity to changes in regulator concentrations, refit the models in *bcd* RNAi embryos alone, and refit the models on all of the data, none of which changed our conclusions (Appendix C Figure S5 and S6, Supplemental Note 1). Although these differences in model performance were subtle, the results supported our hypothesis that the *eve3+7* reporter pattern is regulated differently than the endogenous pattern. Specifically, they suggested that the endogenous pattern required bifunctional Hb while the *eve3+7* reporter pattern required only Hb repression.

Based on these results, we hypothesized that Hb activation is encoded in regulatory DNA outside the annotated *eve3+7* enhancer. The differences in model performance were not conclusive of their own accord, prompting us to seek another perturbation to validate this hypothesis. We therefore returned to the perturbation that previously distinguished the linear and quadratic models, ventral mis-expression of *hb* (Papatsenko and Levine, 2008; Ilsley et al., 2013).

hb mis-expression confirms that the endogenous *eve* pattern and the *eve3+7* reporter pattern respond to Hb differently

In Ilsley et al., we preferred the quadratic model because it qualitatively predicted the behavior of a classic mis-expression experiment. Mis-expressing *hb* along the ventral surface of the embryo (*sna::hb* embryos) causes *eve* stripe 3 to retreat and bend and stripe 7 to bend and expand (Clyde et al., 2003). In simulations of this perturbation, the quadratic model predicted this behavior while the linear model did not (Figure 4.4 E and F reproduced from (Ilsley et al., 2013)). We hypothesized that if Hb bifunctionality is

encoded in the whole locus but not in the annotated *eve3+7* enhancer, the endogenous and *eve3+7* reporter patterns would respond differently to *hb* misexpression. To test this hypothesis, we repeated this perturbation and measured both the endogenous *eve* pattern and the *eve3+7* reporter pattern quantitatively at cellular resolution (Figure 4.4).

In *sna::hb* embryos, the endogenous pattern behaved differently from the *eve3+7* reporter pattern. As previously observed, the endogenous *eve* stripe 3 retreated from the ventral Hb domain and bent posteriorly; the endogenous stripe 7 did not retreat at all, and instead expanded and bent anteriorly. Both of these behaviors were predicted by the quadratic model applied to simulated mis-expression data (Figure 4.4 B and E). By contrast, in the *eve3+7* reporter pattern, both stripes 3 and 7 retreated from the ventral Hb domain, consistent with the predictions of the linear model applied to simulated mis-expression data (Figure 4.4 C and F) (Ilsley et al., 2013). Thus, under this mis-expression perturbation, the linear model predicted the behavior of the *eve3+7* reporter pattern while the quadratic model predicted the behavior of the endogenous pattern. The subtle quantitative difference between the two models that we saw in *bcd* RNAi embryos was corroborated by a strong qualitative difference in *sna::hb* embryos. These data indicate that Hb represses the *eve3+7* reporter pattern and both activates and represses the endogenous pattern. Hb repression is encoded in the *eve3+7* enhancer; we next sought to identify the regulatory DNA in the *eve* locus that directs stripe 7 expression, but is activated by Hb.

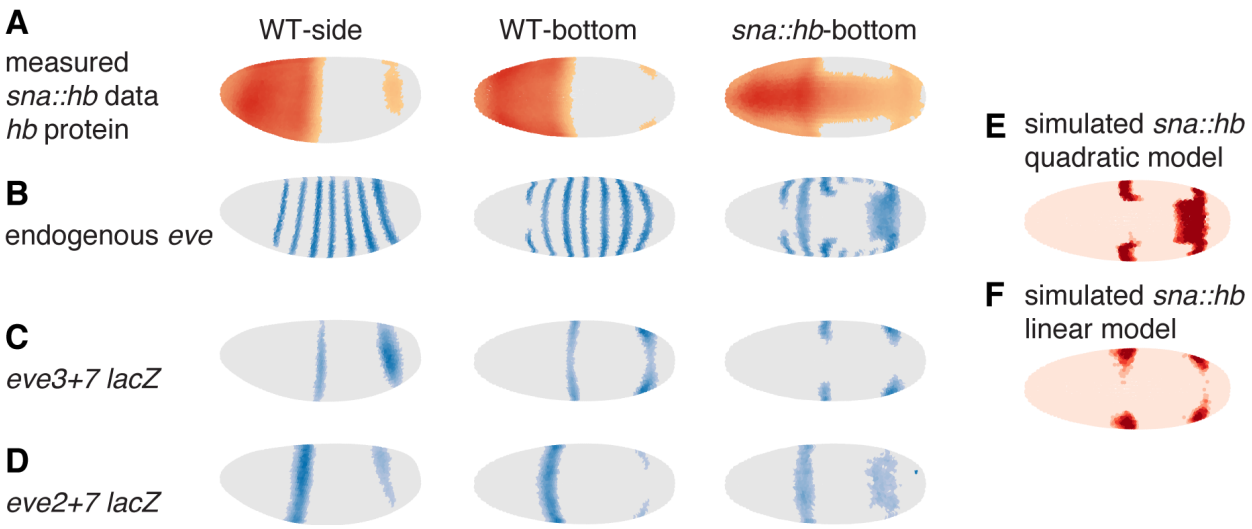


Figure 4.4: In *hb* ventral misexpression *(sna::hb)* embryos, the quadratic model predicts the endogenous pattern while the linear model predicts the *eve3+7* reporter pattern
(A) The distribution of Hb in WT and *sna::hb* embryos, from a lateral view (left) and ventral view (right). The expression level of each regulator is shown for individual cells: cells with expression below an ON/OFF threshold (methods) are plotted in gray. For cells above this threshold, darker colors indicate higher levels. (B) Endogenous *eve* pattern. (C) The *eve3+7* reporter pattern. Both stripes retreat from ectopic Hb. (D) The *eve2+7* reporter pattern. Stripe 7 expands into the ectopic Hb domain. (E-F) Bottom (ventral) view of predictions of the quadratic model (E) and linear (F) models based on simulated *sna::hb* data. OFF cells are light pink and ON cells are red. Reproduced from Ilsley et al. 2013.

Two shadow enhancers encode bifunctional Hb regulation of *eve* stripe 7

We hypothesized that Hb activation of *eve* stripe 7 was encoded in regulatory DNA near the *eve2* enhancer. We focused on the *eve2* enhancer for several reasons: Hb is known to activate the *eve2* enhancer (Arnosti et al., 1996); longer versions of *eve2* drive stripe 7 in some embryos (Goto et al., 1989; Harding et al., 1989; Small et al., 1992; Janssens et al., 2006); orthologous *eve2* enhancers from other species sometimes drive stripe 7 expression (Hare et al., 2008; Peterson et al., 2009); finally, in *sna::hb* embryos the border of the expanded stripe 7 appeared to be set by *Krüppel (Kr)*, a known regulator of *eve2* (Appendix C Figure S7) (Small et al., 1991; Stanojevic

et al., 1991). We measured expression driven by an extended version of the minimal *eve2* enhancer construct that drove a robust stripe 7 pattern (Appendix C Table S2); we call this enhancer reporter construct *eve2+7*. Since both the *eve3+7* and *eve2+7* enhancers drive stripe 7 expression, they can be considered a pair of shadow enhancers for stripe 7 (Barolo, 2012).

We found that Hb activates stripe 7 expression in the *eve2+7* enhancer. In *sna::hb* embryos, the stripe 7 region of the *eve2+7* reporter pattern expanded, recapitulating the bulging behavior observed in the endogenous *eve* pattern (Figure 4.4B and D). We conclude that Hb activates endogenous *eve* stripe 7 through the *eve2+7* enhancer. Taken together, our results indicate that Hb controls *eve* stripe 7 expression by activating and repressing distinct enhancers.

Discussion

We uncovered a pair of shadow enhancers in the *eve* locus that both direct expression of stripe 7 using distinct regulatory logic: one uses Hb as an activator, the other uses Hb as a repressor. We measured expression of the endogenous *eve* locus and transgenic reporter constructs at cellular resolution under two genetic perturbations. We contextualized our results by comparing two computational models with different roles for Hb*:* a linear model where Hb is a dedicated repressor and a quadratic model where Hb is a bifunctional regulator (Ilsley et al., 2013). Guided by the modeling, we found that stripe 7 is encoded by two enhancers: Hb represses the *eve3+7* enhancer and activates the *eve2+7* enhancer. These two shadow enhancers therefore use Hb in different ways to position overlapping patterns. This form of regulatory redundancy may be a general way to encode TF bifunctionality.

Expression patterns driven by reporter constructs do not precisely match the endogenous pattern

"Veteran enhancer-bashers, and those who carefully read the papers, know that 'minimal' enhancer fragments do not always perfectly replicate the precise spatial boundaries of expression of the native gene…" (Barolo, 2012). Our data clearly support this often neglected aspect of enhancer reporter constructs. One explanation offered for such discrepancies is different mRNA half-lives. We controlled for this possibility with a BAC reporter where the *eve* coding sequence was replaced with *lacZ* and found better, but not perfect, agreement with the endogenous pattern. We conclude that transcript sequence features contribute to the differences between reporter and endogenous

patterns, but that additional regulatory DNA in the locus also plays a role. This result highlights the limitations of enhancer reporter constructs for recapitulating behavior of endogenous loci and the importance of using BAC reporters or genomic editing to study loci with multiple enhancers (Venken and Bellen, 2012; Crocker and Stern, 2013; Ren et al., 2013).

eve stripe 7 is encoded by two shadow enhancers

Early efforts to dissect the regulatory architecture of the *eve* locus suggested that stripe 7 activity was distributed over DNA encompassing both the *eve3+7* and *eve2* enhancers (Goto et al., 1989; Harding et al., 1989). We find that there are at least two regions of regulatory DNA that position stripe 7 using different regulatory logic. The classically annotated *eve3+7* enhancer is repressed by Hb (Small et al., 1996; Clyde et al., 2003; Struffi et al., 2011), while the *eve2+7* enhancer, which encompasses the minimal *eve2* enhancer, is activated by Hb. We think that activation of *eve2+7* is direct because it is clear from *in vivo* DNA binding data and binding site mutagenesis that Hb directly activates the minimal *eve2* enhancer (Arnosti et al., 1996; Li et al., 2008). The redundancy in *eve* stripe 7 regulation may confer robustness to genetic or environmental stresses (Frankel et al., 2010; Perry et al., 2010; Dunipace et al., 2011), may increase synchrony or precision (Boettiger and Levine, 2009), may facilitate temporal refinement of patterns (Dunipace et al., 2013), or may arise from genetic drift (Lynch, 2007).

It is likely that these two enhancers are differentially sensitive to additional TFs. Previous studies have revealed the *eve3+7* enhancer is activated by the spatially uniform TFs Stat92E and Zelda, the anterior boundary of stripe 7 is set by Kni

repression, and the posterior boundary is set by Hb repression (Small et al., 1996; Yan et al., 1996; Clyde et al., 2003; Struffi et al., 2011). The minimal *eve2* enhancer is activated by Bcd and Hb, its anterior boundary is set by Slp1 and Gt, and its posterior boundary set by Kr (Stanojevic et al., 1991; Small et al., 1992; Arnosti et al., 1996; Andrioli et al., 2002). Kr appears to set the boundary of both the expanded endogenous *eve* stripe 7 and the *eve2+7* reporter pattern in *sna::hb* embryos (Appendix C Figure S7), but *eve3+7* is not sensitive to Kr, as *eve* stripe 3 sits directly beneath the Kr pattern and this enhancer has no predicted Kr binding sites (Appendix C Figure S7). In agreement with others, we speculate that the anterior boundary of *eve* stripe 7 in *eve2+7* may be set by Gt (Janssens et al., 2006). However, we cannot rule out the possibility that this boundary is set by limiting levels of Hb activation or by Kni repression.

The molecular mechanism by which Hb represses *eve3+7* and activates *eve2+7* remains unclear. One hypothesis is that other TFs convert Hb from a repressor into an activator. For example, there is experimental evidence for activator synergy between *bcd* and *hb* (Simpson-Brose et al., 1994) and Hb/Cad activator synergy has been proposed based on computational work (Kim et al., 2013). Another proposed mechanism is that monomeric Hb is an activator, but Hb dimers are repressors (Papatsenko and Levine, 2008; Bieler et al., 2011). High and low Hb concentrations may also be correlated with some other spatially varying factor in the embryo, such as phosphorylation by the MAPK pathway in the poles (Kim et al., 2010), but data from *sna::hb* embryos reduces the likelihood of this possibility. Testing these hypotheses will

require quantitative data in additional genetic backgrounds and mutagenesis of individual binding sites in the two enhancers.

The quadratic model is a superposition of two computations

Models are not ends in and of themselves, but merely means to formalize assumptions and develop falsifiable hypotheses (Wunderlich and DePace, 2011; Gunawardena, 2014). The quadratic model, which includes Hb as a bifunctional TF, accurately predicts the behavior of the locus in all cells of WT and perturbed embryos, but it does not predict the behavior of either individual enhancer. The interpretation in Ilsley et al. that Hb bifunctionality is a feature of the canonical *eve3+7* enhancer was based on a common assumption: that the endogenous expression pattern could be attributed to the annotated enhancer. Here we show that Hb bifunctionality is encoded in separate enhancers. The quadratic model works because it combines the critical features of the *eve3+7* enhancer and the *eve2+7* enhancer, effectively behaving as a superposition of the two activities. In the future, we plan to develop computational models of each enhancer and uncover how these two (or more) activities are combined.

Conclusion

We tested the hypothesis that Hb bifunctionality is encoded in the *eve3+7* enhancer and discovered that it is actually encoded in two separate enhancers that respond to Hb in opposite ways. We show that expression patterns driven by annotated enhancers differ from the endogenous pattern, especially under perturbation, and that these differences can be due to relevant, yet unannotated, regulatory DNA. The stripe 7

shadow enhancers reside in a classic example of a modular locus (Maeda and Karch, 2011), implying that regulatory complexity may be pervasive. Since the enhancers are active in the same cells, Hb bifunctionality must be encoded in their DNA sequences. This example provides an opportunity to uncover sequence features governing Hb bifunctionality, which will improve our ability to interpret regulatory DNA and infer connections in gene regulatory networks.

Materials and Methods

Fly Work

The *bcd* RNAi gene expression atlas is described in Staller et al. 2014 (submitted) and available at depace.med.harvard.edu. Briefly, we combined short hairpin RNA knockdown of *bcd* with *in situ* hybridization and 2-photon imaging and automated image segmentation (Fowlkes et al., 2008; Fowlkes et al., 2011; Wunderlich et al., 2012; Staller et al., 2013). Hb protein stains used a guinea pig anti-*hb* from John Reinitz (University of Chicago, IL). Embryos were partitioned into six time points using the degree of membrane invagination (0-3%, 4-8%, 9-25%, 26-50%, 51-75%, and 76-100%) which evenly divide the ~60 min blastoderm stage (Keränen et al., 2006). All enhancer reporters are in pBOY and integrated at attP2 (Groth et al., 2004; Hare et al., 2008) (Appendix C Table S1). The *eve* locus bacterial artificial chromosome (BAC) *lacZ* reporter was a gift from Miki Fujioka (Thomas Jefferson University, PA). It begins 6.4 kb upstream of the *eve* transcriptional start site (TSS) and ends 11.3 kb downstream of the *eve* TSS. The *eve* coding sequence has been replaced by *lacZ* and the adjacent gene, *TER94*, has been fused to GFP. It is effectively a reporter for the whole *eve* locus. *hb* ventral misexpression was performed as described in Clyde et al., 2003 using two copies of the transgene on chromosome 2.

Building the coarsely aligned *sna::hb* gene expression atlas.

We determined the genotype of the *sna::hb* embryos by examining the *eve* or *fushi-tarazu (ftz)* mRNA patterns. Embryos were aligned morphologically to create a coarsely registered gene expression atlas (Fowlkes et al., 2008). Data is available at depace.med.harvard.edu.

Logistic modeling of enhancer gene regulatory functions

The logistic modeling framework was developed and described in detail previously (Ilsley et al., 2013). All modeling was performed in MatLab (MathWorks, Natick, MA) using the DIP image toolbox (diplib.org) and the PointCloudToolBox (bdtnp.lbl.gov). Ilsley et al. used protein data for Hb and Gt, whereas we used Hb protein and *gt* mRNA data. For genes where we used mRNA data, the mRNA and protein patterns are known to be correlated (Fowlkes et al., 2008; Pisarev et al., 2009). For the enhancer *lacZ* reporters, we thresholded cells to be ON or OFF by creating a histogram of the expression data (50 bins), identifying the bin with the most counts and adding one standard deviation. Our ON set included all cells expressing the reporter, and our OFF set includes all other cells. All regulators are maintained as continuous values.

To threshold the endogenous WT *eve* pattern into ON and OFF cells we used 0.2 for all time points (Ilsley et al., 2013). To threshold the endogenous *eve* patterns in the *bcd* RNAi atlas, we used the lowest threshold that would separate the stripes: 0.1, 0.15, 0.15, 0.2, and 0.21 for T=2 through T=6 respectively. To compare the modeling of the reporter and the endogenous patterns, the ON set included all cells in the endogenous *eve* stripes 3 and 7 and the OFF set included all other cells. This OFF set is different from Ilsley et al., but this change does not have a large effect on the model prediction AUC scores in *bcd* RNAi embryos (Appendix C Table S1).

Sensitivity analysis

For the sensitivity analysis (Appendix C Figure S5), for each TF, we scaled the concentration of the *bcd* RNAi atlas *in silico* and recomputed the model AUC scores.

Binding site predictions

For the *Kr* binding site analysis in Appendix C Figure S7, we predicted binding sites using PATSER (stormo.wustl.edu) with a position weight matrix derived from bacterial 1-hybrid data (Noyes et al., 2008). Binding sites were visualized using InSite (cs.utah.edu/~miriah/projects).

Quantifying concordance between reporters and endogenous patterns

For each embryo, we used the pointcloud toolbox in Matlab to find pattern boundaries by creating 16 anterior-posterior line traces and finding the inflection point of each. Finding the boundary by using half the maximum value of the stripe peak identifies a very similar boundary to the inflection point. To find the peaks of the endogenous and reporter stripes, we took one line trace along the lateral part of the embryo using the pointcloud toolbox and found the local maxima.

Acknowledgements:

We thank Miki Fujioka for sharing the *eve* locus BAC reporter flies ahead of publication, Tara Lydiard-Martin for making the enhancer reporter fly lines, and Steve Small for the *sna::hb* flies; Garth Ilsley for developing the initial models and very helpful discussions; John Reinitz for the Hb antibody; Steve Small, Becky Ward, Garth Ilsley, Peter Combs, Alistair Boettiger and members of the DePace lab for extensive comments on the manuscript. This work was supported by the Harvard Herchel Smith Graduate Student Fellowship (MVS), Jane Coffin Childs Memorial Fund for Medical Research (ZBW), NIH K99HD073191 (ZBW), and NIH U01 GM103804-01A1 (AHD).

References

Andrioli, L. P. M., Vasisht, V., Theodosopoulou, E., Oberstein, A. L. and Small, S. (2002). Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. *Development (Cambridge, England)* **129**, 4931-4940.

Arnosti, D. N., Barolo, S., Levine, M. S. and Small, S. (1996). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development (Cambridge, England)* **122**, 205-214.

Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* **34**, 135-141.

Bieler, J., Pozzorini, C. and Naef, F. (2011). Whole-Embryo Modeling of Early Segmentation in Drosophila Identifies Robust and Fragile Expression Domains. *Biophysj* **101**, 287-296.

Boettiger, A. N. and Levine, M. S. (2009). Synchronous and Stochastic Patterns of Gene Activation in the Drosophila Embryo. *Science* **325**, 471-473.

Clyde, D. E., Corado, M. S. G., Wu, X., Pare, A., Papatsenko, D. and Small, S. (2003). A self-organizing system of repressor gradients establishes segmental complexity in Drosophila. *Nature* **426**, 849-853.

Crocker, J. and Stern, D. L. (2013). TALE-mediated modulation of transcriptional enhancers in vivo. *Nat Methods* **10**, 762-767.

Deng, Z., Cao, P., Wan, M. M. and Sui, G. (2010). Yin Yang 1: a multifaceted protein beyond a transcription factor. *Transcription* **1**, 81-84.

Di Stefano, B., Sardina, J. L., van Oevelen, C., Collombet, S., Kallin, E. M., Vicent, G. P., Lu, J., Thieffry, D., Beato, M. and Graf, T. (2014). C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506**, 235-239.

Dunipace, L., Ozdemir, A. and Stathopoulos, A. (2011). Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. *Development* **138**, 4075-4084.

Dunipace, L., Saunders, A., Ashe, H. L. and Stathopoulos, A. (2013). Autoregulatory feedback controls sequential action of cis-regulatory modules at the brinker locus. *Dev Cell* **26**, 536-543.

Fowlkes, C. C., Eckenrode, K. B., Bragdon, M. D., Meyer, M., Wunderlich, Z., Simirenko, L., Luengo Hendriks, C. L., Keranen, S. V., Henriquez, C., Knowles, D. W. et al. (2011). A conserved developmental patterning network produces quantitatively different output in multiple species of Drosophila. *PLoS Genet* **7**, e1002346.

Fowlkes, C. C., Hendriks, C. L. L., Keränen, S. V. E., Weber, G. H., Rübel, O., Huang, M.-Y., Chatoor, S., DePace, A. H., Simirenko, L., Henriquez, C. et al. (2008). A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. *Cell* **133**, 364-374.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F. and Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490-493.

Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T. and 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 (Cambridge, England)* **126**, 2527-2538.

Goto, T., Macdonald, P. and Maniatis, T. (1989). Early and late periodic patterns of< i> even skipped</i> expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413-422.

Groth, A. C., Fish, M., Nusse, R. and Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775-1782.

Gunawardena, J. (2014). Models in biology:'accurate descriptions of our pathetic thinking›. *BMC Biology*

Harding, K., Hoey, T., Warrior, R. and Levine, M. S. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. *The EMBO journal* **8**, 1205-1212.

Hare, E. E., Peterson, B. K., Iyer, V. N., Meier, R. and Eisen, M. B. (2008). Sepsid even-skipped Enhancers Are Functionally Conserved in Drosophila Despite Lack of Sequence Conservation. *PLoS Genetics* **4**, e1000106.

Ilsley, G. R., Fisher, J., Apweiler, R., DePace, A. H. and Luscombe, N. M. (2013). Cellular resolution models for even skipped regulation in the entire Drosophila embryo. *Elife* **2**, e00522.

Janssens, H., Hou, S., Jaeger, J., Kim, A.-R., Myasnikova, E., Sharp, D. and Reinitz, J. (2006). Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. *Nature genetics* **38**, 1159-1165.

Keränen, S. V. E., Fowlkes, C. C., Luengo Hendriks, C. L., Sudar, D., Knowles, D. W., Malik, J. and Biggin, M. D. (2006). Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution II: dynamics. *Genome Biology* **7**, R124.

Kim, A.-R., Martinez, C., Ionides, J., Ramos, A. F., Ludwig, M. Z., Ogawa, N., Sharp, D. H. and Reinitz, J. (2013). Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic. *PLoS Genetics* **9**, e1003243.

Kim, Y., Coppey, M., Grossman, R., Ajuria, L., Jiménez, G., Paroush, Z. and Shvartsman, S. Y. (2010). MAPK Substrate Competition Integrates Patterning Signals in the< i> Drosophila Embryo. *Current Biology* **20**, 446-451.

Levine, M., Cattoglio, C. and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13-25.

Li, X.-Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D. A., Iyer, V. N., Hechmer, A., Simirenko, L., Stapleton, M., Hendriks, C. L. L. et al. (2008). Transcription Factors Bind Thousands of Active and Inactive Regions in the Drosophila Blastoderm. *PLoS biology* **6**, e27.

Luengo Hendriks, C. L., Keränen, S. V. E., Fowlkes, C. C., Simirenko, L., Weber, G. H., DePace, A. H., Henriquez, C., Kaszuba, D. W., Hamann, B., Eisen, M. B. et al. (2006). Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. *Genome Biology* **7**, R123.

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences of the United States of America* **104 Suppl 1**, 8597-8604.
Maeda, R. K. and Karch, F. (2011). Gene expression in time and space: additive vs hierarchical organization of cis-regulatory regions. *Curr Opin Genet Dev* **21**, 187-193.
Noyes, M. B., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M. H. and Wolfe, S. A. (2008). A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. *Nucleic Acids Res* **36**, 2547-2560.
Papatsenko, D. and Levine, M. S. (2008). Dual regulation by the Hunchback gradient in the Drosophila embryo. *Proc Natl Acad Sci U S A* **105**, 2901-2906.
Perry, M. W., Boettiger, A. N., Bothma, J. P. and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. *Curr Biol* **20**, 1562-1567.
Peterson, B. K., Hare, E. E., Iyer, V. N., Storage, S., Conner, L., Papaj, D. R., Kurashima, R., Jang, E. and Eisen, M. B. (2009). Big genomes facilitate the comparative identification of regulatory elements. *PLoS ONE* **4**, e4688.
Pisarev, A., Poustelnikova, E., Samsonova, M. and Reinitz, J. (2009). FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic acids research* **37**, D560-D566.
Ren, X., Sun, J., Housden, B. E., Hu, Y., Roesel, C., Lin, S., Liu, L. P., Yang, Z., Mao, D., Sun, L. et al. (2013). Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. *Proc Natl Acad Sci U S A* **110**, 19012-19017.
Sauer, F. and Jäckle, H. (1993). Dimerization and the control of transcription by Krüppel. *Nature* **364**, 454-457.
Schulz, C. and Tautz, D. (1994). Autonomous concentration-dependent activation and repression of Kruppel by hunchback in the Drosophila embryo. *Development* **120**, 3043-3049.
Shore, D. and Nasmyth, K. (1987). Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell* **51**, 721-732.
Simpson-Brose, M., Treisman, J. and Desplan, C. (1994). Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. *Cell* **78**, 855-865.
Small, S., Blair, A. and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. *EMBO J* **11**, 4047-4057.
Small, S., Blair, A. and Levine, M. S. (1996). Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. *Developmental biology* **175**, 314-324.
Small, S., Kraut, R., Hoey, T., Warrior, R. and Levine, M. S. (1991). Transcriptional regulation of a pair-rule stripe in Drosophila. *Genes & Development* **5**, 827-839.
Staller, M. V., Fowlkes, C. C., Bragdon, M. D., Wunderlich, Z. and DePace, A. H. (2014). A gene expression atlas of a *bicoid*-depleted *Drosophila* embryo reveals early canalization of cell fate
. *In Review. Available from BioRXiv*
Staller, M. V., Yan, D., Randklev, S., Bragdon, M. D., Wunderlich, Z. B., Tao, R., Perkins, L. A., DePace, A. H. and Perrimon, N. (2013). Depleting gene activities

in early Drosophila embryos with the «maternal-Gal4-shRNA» system. *Genetics* **193**, 51-61.

Stanojevic, D., Small, S. and Levine, M. S. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science* **254**, 1385-1387.

Struffi, P., Corado, M., Kaplan, L., Yu, D., Rushlow, C. and Small, S. (2011). Combinatorial activation and concentration-dependent repression of the Drosophila even skipped stripe 3+7 enhancer. *Development (Cambridge, England)* **138**, 4291-4299.

Struhl, G., Johnston, P. and Lawrence, P. A. (1992). Control of Drosophila body pattern by the hunchback morphogen gradient. *Cell* **69**, 237-249.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science* **240**, 1285-1293.

Venken, K. J. and Bellen, H. J. (2012). Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and PhiC31 integrase. *Methods Mol Biol* **859**, 203-228.

Wunderlich, Z., Bragdon, M. D., Eckenrode, K. B., Lydiard-Martin, T., Pearl-Waserman, S. and DePace, A. H. (2012). Dissecting sources of quantitative gene expression pattern divergence between Drosophila species. *Mol Syst Biol* **8**, 604.

Wunderlich, Z. and DePace, A. H. (2011). Modeling transcriptional networks in Drosophila development at multiple scales. *Curr Opin Genet Dev* **21**, 711-718.

Yan, R., Small, S., Desplan, C., Dearolf, C. R. and Darnell, J. E. (1996). Identification of a Stat gene that functions in Drosophila development. *Cell* **84**, 421-430.

Zuo, P., Stanojevic, D., Colgan, J., Han, K., Levine, M. S. and Manley, J. L. (1991). Activation and repression of transcription by the gap proteins hunchback and Kruppel in cultured Drosophila cells. *Genes & Development* **5**, 254-264.

Chapter 5: Improving shRNA depletion of early zygotic genes

Improving shRNA knockdown in the embryo

We have been working to improve shRNA depletion to target genes that remain refractory to knockdown. In the embryo, we can very effectively knock down maternally deposited genes using shRNAs and some genes essential after gastrulation, but we cannot yet effectively deplete early zygotic genes. These early zygotic genes include most of the segmentation genes we normally study in the lab. Working with mutant alleles for these genes is labor intensive because the stocks are sickly and the alleles vary in quality. shRNA depletion would enable collection of large pools of affected embryos for building gene expression atlases and biochemical analysis.

There are two strategies for knocking down early zygotic genes: zygotic expression of shRNAs (called F1 phenotypes in Chapter 2) or maternal loading of shRNAs (called F2 phenotypes in Chapter 2). In practice, maternally loaded shRNAs are not stable enough to efficiently deplete zygotic genes (Chapter 2, Table 1). The two exceptions are efficient knockdown of *dpp* and partial knockdown of *hb* ,described in detail in Chapter 2. By crossing shRNA expressing mother to heterozygous mutant fathers, we showed that sometimes we could generate a phenotype, indicating the hairpins work in principle, but not enough hairpin is being delivered. Some of the heterozygote experiments suggested that combining maternal loading and zygotic expression may be the most effective strategy. Improving knockdown will require delivering more hairpin at the right time.

There are two strategies for delivering more hairpin into the embryo: activating zygotic shRNAs earlier or increasing stability of maternally loaded shRNAs (Arrows in Figure 5.1). In the 2013 paper, we attempted to increase the stability of the shRNAs. We hypothesized that shRNA half-life is encoded in the sequence of the miRNA precursor. We identified long lived miRNAs (Votruba, 2009) and used the precursor backbones to rebuild some shRNA constructs (Chapter 2). We did not measure any significant improvement. These data were published as a negative result in the 2013 paper.

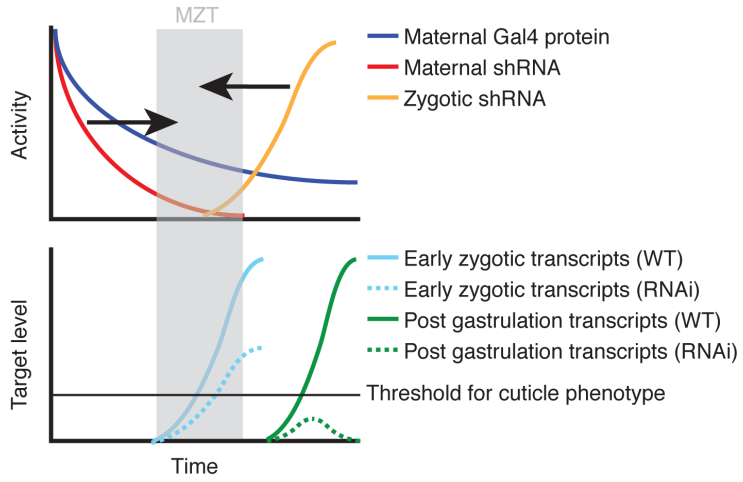


Figure 5.1 Two strategies for improving shRNA depletion of early zygotic genes. The first strategy is to increase the stability of maternally loaded shRNAs (left black arrow) and the seconds strategy is to activate zygotic shRNAs earlier (right black arrow).

Earlier activation of zygotic shRNAs

We have used two tactics to activate the zygotic shRNAs earlier: maternally loaded Gal4-VP16 and modifying the promoter of the *UAS-shRNA* construct to 'mark' it as an early zygotic gene.

Characterizing additional maternal drivers

We hypothesized that maternally loaded Gal4-VP16 could improve RNAi efficacy. Others have shown that Gal4 is not a great activator in the early embryo, but Gal4-VP16 is very potent (Brand chapter in drosophila protocols). We ordered two *mat-tub-Gal4-VP16* lines from Bloomington: B7062 w[*]; P(w[+mC]=matalpha4-GAL-VP16)V2H and B7063 w[*]; P(w[+mC]=matalpha4-GAL-VP16)V37. We tested each line alone and also created a line that combines both drivers (DePace Line 482, MVS 6CAD1AA). To assess the potency of these new drivers, we crossed them with the *UAS-shRNA-bcd* construct. While our standard drivers yielded 100% lethality, both of these new drivers yielded escapees. Based on the escapee rate and phenotype strength (see below), we determined relative driver potency to be: B7062 < B7063 < MTD-Gal4 < 482. The presence of escapees in this treatment is consistent with the variable phenotypes observed below.

Early activation of zygotic shRNAs with GAl4-VP16 (F1 phenotypes)

We first tried maternally loading Gal4-VP16 to activate paternally supplied UAS-shRNA constructs (F1 phenotypes). Crossing these drivers lines to *UAS-shRNA* constructs (Valium 20 and Valium22) against early zygotic genes did not yield increased lethality. We conclude these drivers do not create strong new F1 phenotypes with existing UAS-shRNA constructs.

Maternal loading of shRNAs with GAl4-VP16 (F2 phenotypes)

We next tested if the Gal4-VP16 could produce more severe phenotypes when maternally loaded (F2 phenotypes). We I used hb RNAi as a sensitized test case. The *UAS-shRNA-hb* (DePace Line 344; GL01321) gives the same phenotype using *MTD-Gal4* and *mat-tub-gal4* at 25C: T2 and T3 are missing (Chapter 2 Figure 4). Importantly, the posterior A8 segment is intact, meaning there is enough posterior zygotic *hb* to correctly specify this segment. When we cross *MTD-Gal4/UAS-shRNA-hb* mothers to hb[12]/+ males, we see a second class of embryos (presumably the *hb* heterozygous) that are missing A1 and additional head structures but the posterior is still instact. Together these results show that anterior patterning is more sensitive to *hb* levels than posterior patterning. They also indicate we can detect improvements in shRNA depletion using the cuticle phenotype.

As a baseline control, we repeated *hb* knockdown at 29C with old and new drivers. shRNA depletion is more effective at higher temperature, especially for constructs in the Valium20 backbone which uses an *HSP70* promoter (Ni et al., 2011a). To our surprise, we found that the original *UAS-shRNA-hb* construct form TRiP (HMS01183) yielded a phenotype at using the *mat-tub-Gal4* driver: 65% lethality at 25C and 75% lethality at 29C. This construct was 100% viable at 25C in conjunction with the *MTD-Gal4* driver. The phenotype was similar to GL01321 at 25C (Chapter 2), where two thoracic segments are missing. At 29C, *mat-tub-Gal4>> UAS-shRNA-hb* (GL01321) flies laid embryos with a stronger, more variable phenotype than they did at 25C.

The preliminary experiments at 29C suggested that the Gal4-VP16 did improve knockdown. We crossed *MTD-Gal4,* 1x Gal4-VP16 (B7063; chromosome 3) and 2x Gal4-VP16 (482) to *UAS-shRNA-hb* (GL01321, Valium22 backbone) and collected F2 embryos. Adam Carte quantified the strength of knockdown by counting segments on individual cuticles (Figure 5.2). The phenotypes are very variable, but by examining the average phenotype there is a clear trend: 2x Gal4-VP16 (482) is stronger than *MTD-Gal4* which is in turn stronger than B7063 (1x Gal4-VP16). We were very excited to see that some of these embryos had defects in the posterior. Since posterior *hb* is entirely zygotic, posterior defects mean that we really are knocking down more zygotic transcripts.

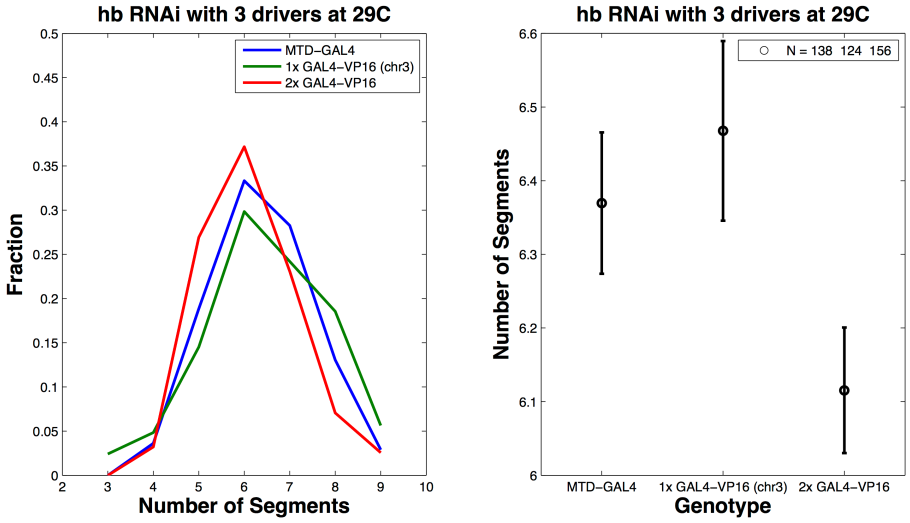


Figure 5.2 The 2xGal4-VP16 driver yields a stronger hb phenotype at 29C.
Left the histogram of cuticle segment counts (ventral denticle bands). The average number of segments in each population. Counts in legend. GL01321 was used.

To further characterize the Gal4-VP16 drivers, we activated UAS-shRNA-GFP constructs as a control. I noticed that the 2xGal4-VP16 driver stock was sickly and

Norbert Perrimon told me that too much Gal4 can mess up an embryo. Embryos laid by *2xGal4-VP16>>UAS-shRNA-GFP* females have substantially reduced viability at 29C (never quantified, but it was obvious by eye). Cuticle preparations revealed a very diverse array of messed up embryos, some resembling anterior-posterior, terminal or dorsal-ventral phenotypes. At this time, the apparent improvement in shRNA knockdown observed by using the *2xGal4-VP16* driver and collecting embryos at 29C is convolved with the detrimental effects of Gal4-VP16 on development and temperature stress.

In an effort to reduce the deleterious effects of "just having too much" Gal4-VP16 in the embryo, we examined phenotypes at 25C, a less stressful environment. We simultaneously tested for the influence of *UAS-shRNA-hb* copy number in the embryo. Crossing *1xGal4-VP16>>UAS-shRNA-hb* females with WT males yields embryos with 0 or 1 copies of the *UAS-shRNA-hb* construct, while *1xGal4-VP16>>UAS-shRNA-hb* females with *UAS-shRNA-hb* males yields embryos with 1 or 2 copies of the *UAS-shRNA-hb* construct. As a control we collected embryos from *mat-tub-Gal4>>UAS-shRNA-hb* flies. In this experiment we noticed that <6% of these embryos had a stronger than normal hb RNAi phenotype, containing only 8 segments, a subpopulation we may have perviously missed. Even with this group, embryos collected from *mat-tub-Gal4>>UAS-shRNA-hb* have a very uniform phenotype compared to other RNAi phenotypes. We were surprised to see that having fewer copies of the transgene yielded a stronger phenotype. These results are unexpected and merit repetition and comparison to GFP RNAi embryos.

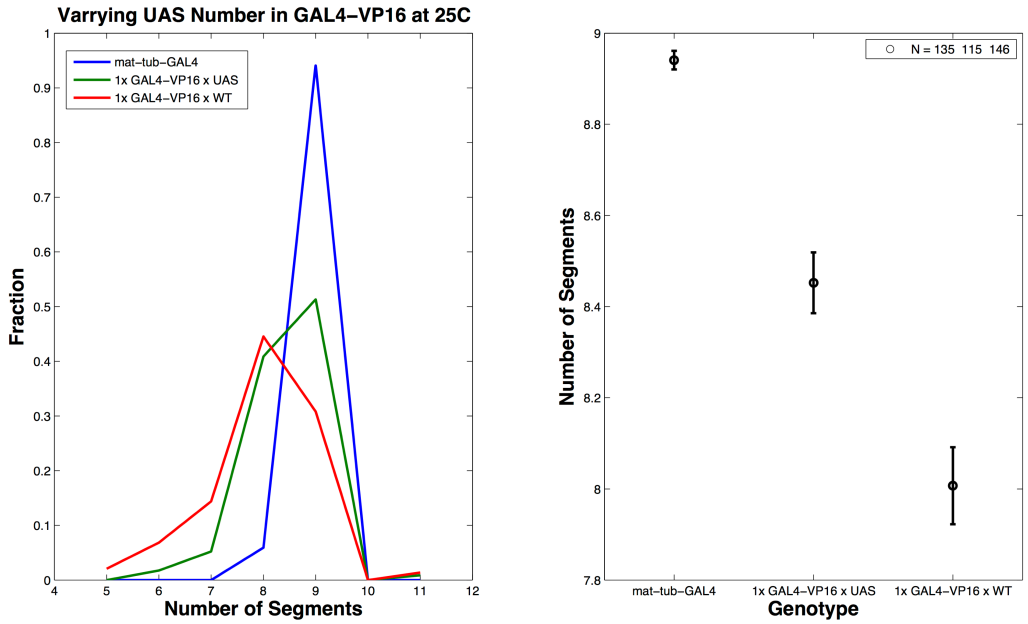


Figure 5.3 The number of UAS-shRNA-hb constructs in the embryo has an unexpected effect on phenotype at 25C.
The green population of embryos has 1 or 2 transgenes with the red population has 0 or 1 transgene. Number of segments was determined by counting ventral denticle bands. GL01321 was used.

In conclusion, we have made some progress increasing the severity of the *hb* RNAi phenotype using higher temperature and Gal4-VP16. The improvement comes at the cost of increased phenotypic variability and potential detrimental effects of Gal4-VP16. Alternatively, the shRNA against GFP may have off target effects, so we propose testing a second GFP shRNA.

Changing the *UAS-shRNA* promoter

Our second strategy for improving shRNA depletion has been to add *Zelda* binding sites to the promoter of the *UAS-shRNA* construct to 'mark' it as an early zygotic gene. *Zelda* binding appears to mark early expressing enhancers and promoters and has been called a pioneer factor (Harrison et al., 2011). *Zelda* motifs predict early

zygotic TF binding and are enriched in the regulatory DNA of the earliest expressed genes (De Renzis et al., 2007; Zeitlinger et al., 2007; Xu et al., 2014). We hypothesized that adding Zelda sites would cause earlier zygotic activation of shRNAs. We added 6 zelda sites to the promoter of the Valium20 shRNA backbone, which is more effective in somatic tissue (Figure 5.4). Valium20 works in virtually all tissues, including the female germline, while Valium22 is optimized for the female germline and shows very little efficacy in other tissues (Ni et al., 2011b). We hypothesized these extra sites would keep this promoter more open and prime it for earlier, stronger activation.

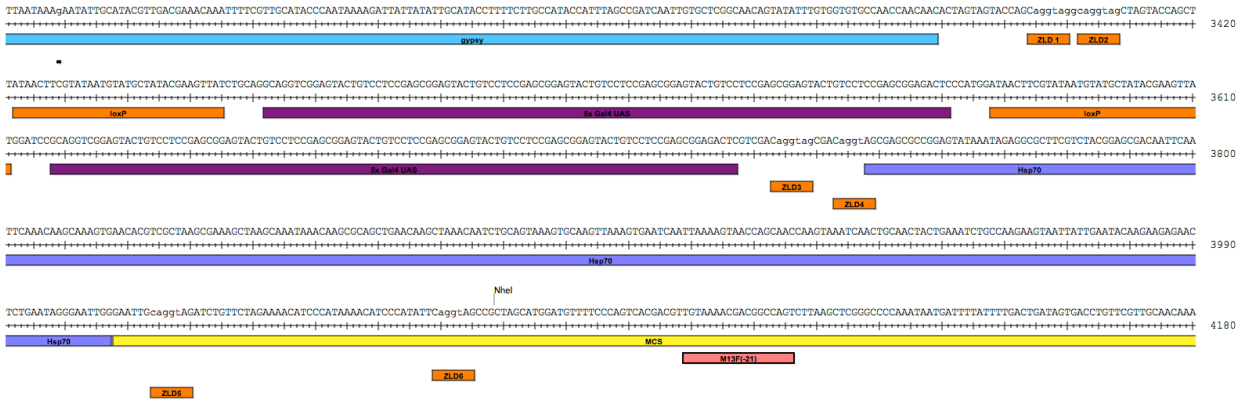


Figure 5.4 We inserted 6 *Zelda* sites into the promoter region of the Valium20 shRNA delivery vector.

So far, there is no measurable improvement in shRNA depletion from adding *Zelda* sites to the promoter of *Valium20.* When we crossed flies carrying *UAS-shRNA +6xZld-gt* or *UAS-shRNA+6xZld-Kr* males to *MTD-Gal4* or *mat-tub-Gal4* females, we had high viability. These promoter mutation do not create new F1 phenotypes for these two very early expressing genes. We next examined embryos laid by *maternal-Gal4>>UAS-shRNA+6xZld* mothers for F2 phenotypes and saw no significant

improvement. We tested if combining the *2xGal4-VP16* drivers with the *Zelda* promoter variants could improve knockdown. There was no F1 lethality at 20C. At 29C, we saw reduced viability and highly variable phenotypes. For *gt* these phenotypes did not resemble published images of mutant embryos. For Kr, some individuals resembled weak Kr mutant alleles. However, we do not yet have a good system for "subtracting: the distribution of phenotypes laid by *2xGal4-VP16>>UAS-shRNA-GFP* females, so we have not confirmed these effects are specific. It is most likely at this time, that adding *Zelda* sites does not measurably improve shRNA depletion.

There remain a few options for improving shRNA knock down in the early embryo. We could insert into Valium20 a different early acting promoter, such as the *eve* basal promoter, or the *engrailed* promoter(Ali-Murthy et al., 2013). It is also possible to encode multiple hairpins into the same construct (Haley et al., 2008; Haley et al., 2010). Building constructs with multiple hairpins against the same gene may improve efficacy.

shRNAs can simultaneously deplete two targets

The maternal Gal4 system can be used to knock down two genes at once. I crudely combined the *UAS-shRNA-hb* on chromosome 2 (attP40) with the *UAS-shRNA-bcd* on chromosome 3 (attP2) and found embryos with a perfect bicaudal (two-tailed) phenotype (Figure 5.5). In principle it should work for more genes. Future efforts may find it useful to put multiple shRNAs into the same transgene, simplifying the genetics (Haley et al., 2008; Haley et al., 2010).

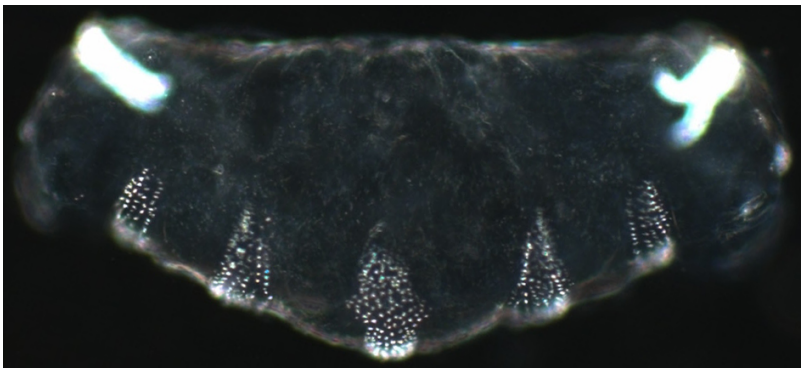


Figure 5.5: Embryos laid by mat-tub-Gal4>>UAS-shRNA-bcd, UAS-shRNA-hb females have the expected phenotype. The embryos are completely symmetric, with 2 well formed sets of tail structures and 4-5 segments.

References:

Ali-Murthy, Z., Lott, S. E., Eisen, M. B. and Kornberg, T. B. (2013). An Essential Role for Zygotic Expression in the Pre-Cellular Drosophila Embryo. *PLoS Genetics* **9**, e1003428.

De Renzis, S., Elemento, O., Tavazoie, S. and Wieschaus, E. F. (2007). Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. *PLoS biology* **5**, e117.

Haley, B., Foys, B. and Levine, M. S. (2010). Vectors and parameters that enhance the efficacy of RNAi-mediated gene disruption in transgenic Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11435.

Haley, B., Hendrix, D., Trang, V. and Levine, M. S. (2008). A simplified miRNA-based gene silencing method for Drosophila melanogaster. *Developmental biology* **321**, 482-490.

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. and Eisen, M. B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genetics* **7**, e1002266.

Ni, J.-Q., Zhou, R., Czech, B., Liu, L.-P., Holderbaum, L., Yang-Zhou, D., Shim, H.-S., Tao, R., Handler, D., Karpowicz, P. et al. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nature Methods* **8**, 405-407.

Ni, J. Q., Zhou, R., Czech, B., Liu, L. P., Holderbaum, L., Yang-Zhou, D., Shim, H. S., Tao, R., Handler, D., Karpowicz, P. et al. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat Methods* **8**, 405-407.

van Roessel, P. and Brand, A. D. (2000). GAL4-mediated Ectopic Gene Expression in *Drosophila*. In *Drosophila Protocols* (ed. W. Sullivan, M. Ashburner and R. S. Hawly), pp. 439-447. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Votruba, S. (2009). MICRORNAS IN THE DROSOPHILA EGG AND EARLY EMBRYO. 1-75.

Xu, Z., Chen, H., Ling, J., Yu, D., Struffi, P. and Small, S. (2014). Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. *Genes & Development* **28**, 608-621.

Zeitlinger, J., Zinzen, R. P., Stark, A., Kellis, M., Zhang, H., Young, R. A. and Levine, M. S. (2007). Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. *Genes & Development* **21**, 385-390.

Chapter 6: Discussion

Overview

In this thesis I described our efforts to make *trans* perturbations in embryos, characterize how the embryo canalizes cell fate in response to perturbation, and use these perturbations to validate computational models. Our continuing efforts to improve shRNA depletion in the embryo were described in Chapter 5. In this chapter, I discuss some ongoing work and propose new follow up experiments. I propose how to further investigate canalization of cell fate following maternal perturbations to the embryo. Next, I include an extended description of proposed work to investigate the mechanisms and implications of Hb bifunctional regulation of *eve* stripe 7, some of which is already under investigation by others in the lab.

Canalization

What features of network architecture contribute to canalization of cell fate?

We would like to identify the features of the gene regulatory network architecture (connections) that lead to canalization. For *bcd* RNAi, the current explanation in the literature is that there are no new cell types because without Bcd, Cad is translated throughout the embryo and the head defaults to posterior (Lawrence, 1992). The persistence of maternal Hb in the anterior complicates this cartoon explanation. Many people have invoked the abundant cross repression in the network as responsible for the canalization we observe. Theoretical work has shown that mutual repression can sharpen boundaries (Sokolowski et al., 2012). Testing the role of cross repression in endogenous networks is difficult because while removing genes (nodes) or enhancers

(collections of edges) is straightforward, no technique exists to remove one edge from an endogenous network. Ben Vincent has piloted using RNAi against co-repressors to remove large subsets of edges, which is a very promising technique. It is becoming possible to use CRISPR-Cas9 genomic engineering to replace enhancers in the endogenous locus with engineered variants and this technique could be refined to remove specific edges from the network (Ren et al., 2013).

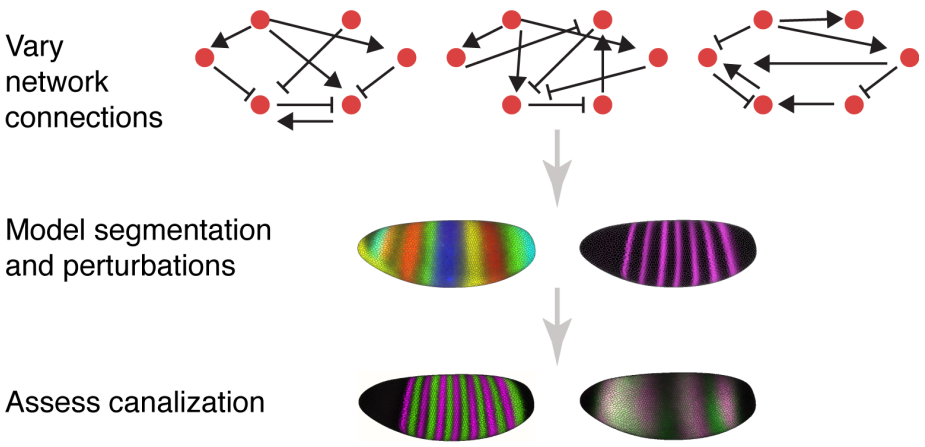


Figure 6.1 A computational screen for features of network architecture that support canalization.

An immediately available strategy for uncovering the features of network architecture responsible for canalization is to use computer simulations. We propose to simulate variant gene regulatory networks and random networks and determine which ones canalize cell fate and which ones do not (Figure 6.1). Indeed, simulations have been used to argue that canalization is merely a consequence of densely connected developmental networks (Siegal and Bergman, 2002). However, the underlying model in this work may not an appropriate abstraction of this network. Simulations will require a

model of segmentation to evaluate if each topology can canalize cell fate, and choosing the appropriate model will be tricky. Abstracted dynamical models are promising (Bieler et al., 2011; Papatsenko and Levine, 2011). In the long run, we could build synthetic networks to test the predictions of our modeling.

What regulatory DNA mediates eve/ftz canalization?

The exquisite robustness of embryogenesis results from canalization and compensation. We have shown that under *bcd* depletion, many cells initially express the pair-rule genes *eve* and *ftz*, but this fraction decreases over time. The initially overlapping anterior stripe resolves into mutually exclusive stripes. We interpret this "clean-up" as canalization of gene expression, but from another perspective it could be viewed as compensation for an error. In either interpretation, the perturbation revealed a backup patterning mechanism. The most likely mechanism of this canalization is the strong mutual repression of *eve* and *ftz* (Jiang et al., 1991; Manoukian and Krause, 1992; Fujioka et al., 1996; Saulier-Le Drean et al., 1998; Nasiadka and Krause, 1999; Schroeder et al., 2011). In WT it appears that the stripe specific enhancers specify position accurately enough that there is relatively little need for clean up by cross repression. In contrast, under maternal genetic perturbation, these strong cross repressive interactions appear to be important for building segments.

We propose to test the generality of pair-rule gene expression canalization. We saw *eve/ftz* canalization in *bcd* RNAi embryos (Chapter 3). In unpublished experiments, we also see canalization in *nanos (nos)* RNAi embryos (Figure 6.2A). Preliminary data for hb RNAi embryos suggests there is no increase in initial *eve/ftz* overlap and no clean

up, consistent with other evidence that this is a milder perturbation. In *bcd* RNAi, we also see canalization of the *hairy* and *runt* stripes, another set of pair-rule genes with strong cross repression (Figure 6.2B) (Ingham and Gergen, 1988; Carroll and Vavra, 1989). In *ctpb* RNAi embryos, *eve/ftz* canalization is delayed. This delay may indicate *ctbp* is a co-repressor for *eve*, *ftz* or *odd*. Other have reported similar data on *eve/ftz* canalization in *Kr* mutant embryos (Teytelman et al., 2013). We will examine *eve/ftz* stripe canalization in other mutants and maternal RNAi knockdowns to survey the generality of this phenomenon.

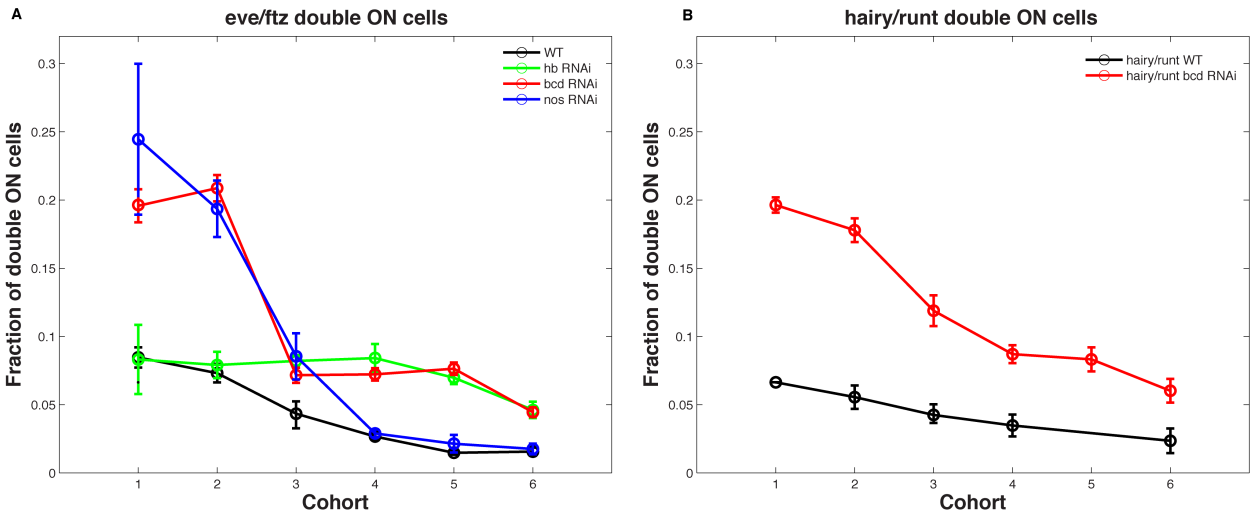


Figure 6.2 Canalization of pair-rule genes in multiple genotypes.

Uncovering the regulatory DNA mediating *eve/ftz* stripe refinement will help identify the specific regulators and gene regulatory network design principles responsible for canalization. We hypothesize that clean up is mediated by 7-stripe, pair-rule responsive enhancers in each of these loci. To test this claim, we could build BAC reporter constructs that have deleted the seven-stripe enhancers and examine if these

reporters canalize following maternal insult. If canalization is localized to these enhancers, then we will not see canalization. In this case, we would infer that canalization is dominated by pair-rule feedback and we can begin to test the roles of individual regulators and/or binding sites in the seven stripe enhancers. If canalization is not localized to these enhancers we will examine reporter constructs for the gap responsive enhancers. If these refine, then they are responsible for clean up. In this case changes in gap gene expression must control canalization. If we still cannot account for canalization, we will look at additional, unannotated regulatory DNA.

We went to great lengths to reduce *eve* and *ftz* variability when building our gene expression atlas, but we anticipate variability will be useful for future analyses. A striking example of variability is the tendency for the anterior *eve* stripe to bifurcate in bcd RNAi embryos. In 22% of embryos the endogenous anterior *eve* stripe splits into 2 stripes, while the *eve3+7* reporter never splits. There is abundance of spatial inhibition and there is evidence that if the anterior eve stripe is too wide it bifurcates into two stripes. While it is well established that the *eve* and *ftz* stripes are not established by a proper Turing Pattern mechanism (Kondo, 2002), a Turing-like mechanism may underlie the canalization of compartment boundaries (proposed by Tom Hiscock in the Megason lab). We will examine the length scale of spatial correlation, variability and dynamics of these expression patterns to test this hypothesis.

How do maternal inputs give rise to the gap gene patterns?

One of the grand goals in developmental systems biology is to build models that can accurately predict the behavior of WT and perturbed embryos. In particular, we

want to predict the behavior of naturally occurring cis-regulatory variants. The way towards that goal is to build models that can explain existing data. In sea urchins, a binary model captures the majority of the behavior of the gene regulatory network (Peter et al., 2012). In Drosophila, gene circuit models have used ordinary differential equations with considerable success at explaining the WT embryo (Jaeger et al., 2004a; Jaeger et al., 2004b; Manu et al., 2009a; Manu et al., 2009b). Thermodynamic models have been used with mixed success to model regulatory DNA (Janssens et al., 2006; Kazemian et al., 2010; Samee and Sinha, 2013; Suleimenov et al., 2013). However, one model that was successful in 1D did not generalize well to 3D data (Hengenius et al., 2011). In addition, a recent effort to use the established gene circuit model to predict a zygotic mutant forced a major revision of the network architecture, suggesting these models are not yet complete (Kozlov et al., 2012). A similar effort to use the standing model to explain a cis-regulatory perturbation led to another revision (Kim et al., 2013). These modeling efforts remain severely limited by the available data.

Although the cartoon model of anterior-posterior patterning has been clear for two decades, there is still no computational model that can take only the maternal inputs and simulate the behavior of the segmentation network through gastrulation (Jaeger et al., 2007). The existence of many models than can fit the WT data suggests the data do not adequately constrain the possible space of alternative models. Collecting data from multiple, simultaneous maternal perturbations will constrain models, ruling out some and validating others. The next set of experiments would be to deplete all combinations of the anterior-posterior maternal determinants: *bcd, hb, nanos, torso or torso-like,* and *capicua* (Lawrence, 1992). Constitutively active *torso* mutants may also be useful

(Casanova and Struhl, 1993). We could either build coarse gene expression atlases, or use a recently developed antibody staining protocol that can stain all 4 gap proteins in single embryos (Dubuis et al., 2013b). Many options for modeling frameworks exist, but one that explicitly contains time information would be most appropriate, perhaps similar to published dynamical modeling approaches (Jaeger et al., 2004b; Bieler et al., 2011; Hengenius et al., 2011; Papatsenko and Levine, 2011). Building a complete model will test if known regulatory connections are sufficient to explain the self-organizing behavior of this network in WT and in mutants.

Mechanisms of Hunchback bifunctional regulation

What is the mechanism of Hunchback bifunctional regulation?

We would like to understand the DNA sequence features that control whether Hb activates or represses a target enhancer. We demonstrated that Hb bifunctionally regulates *eve* stripe 7 by activating the *eve2+7* enhancer and repressing the *eve3+7* enhancer. Since these enhancers are active in the same cells (stripe 7), local DNA sequence must determine the sign of Hb regulation. This feature rules out some mechanisms to control of Hb bifunctionality, such as cell wide changes in protein state. To uncover the mechanism of Hb bifunctional regulation we propose to find other examples, separate the activating and repressing activities of the protein, test the role of dimerization, test the function of the phosphorylation sites and examine orthologous sequences.

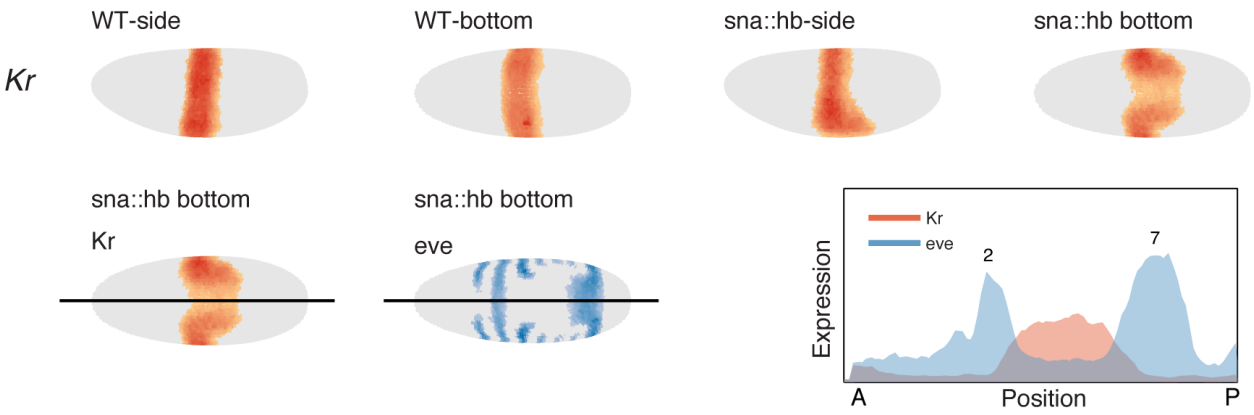


Figure 6.3 Hb appears to bifunctionally regulate Kr in sna::hb embryos.
Reproduced from Appendix C Figure 7

Identifying additional examples of Hb bifunctional regulation

To search for another example of Hb bifunctional regulation that might be mediated by separate enhancers, Ben Vincent and his summer intern, Jonathan White investigated the *Kr* locus. *Kr* is the most studied target of Hb bifunctional regulation (Struhl et al., 1992; Schulz and Tautz, 1994), but the regulatory DNA responsible has never been identified. In addition, the *Kr* locus contains two shadow enhancers, but each is comprised of different predicted transcription factor binding sites. Indeed, Zeba Wunderlich has shown that one enhancer is strongly activated by *dStat* while the other is not. In *sna::hb* embryos endogenous *Kr* expands, but it is repressed slightly in the region of highest Hb misexpression, once more hinting at bifunctional regulation (Figure 6.3 reproduced from Appendix C Figure S7). Inspired by our findings in the *eve* locus, Ben and Jonathan tested if the two *Kr* shadow enhancers responded differently to *sna::hb* misexpression. The results are less clear than for *eve* stripe 7. One of the shadow enhancers expands in *sna::hb* embryos, and is clearly activated (Figure 6 .4).

The other appears to be both activated and repressed: it bends and expands somewhat, but is also slightly repressed in the region of highest *hb* misexpression, strongly resembling the endogenous pattern. This example warrants further study and may provide a useful complement to *eve* stripe 7 regulation.

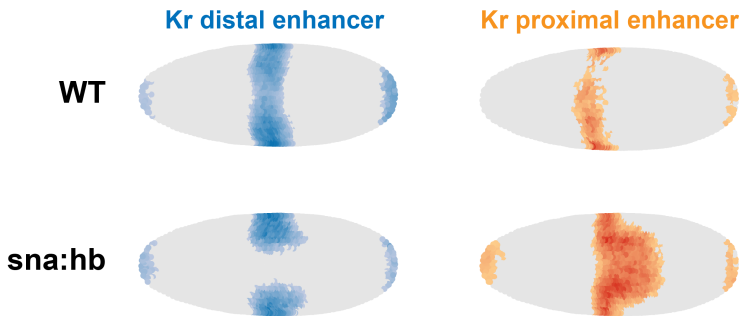


Figure: 6.4 The *Kr* shadow enhancers respond differently to *sna::hb* misexpression

Separating the mechanisms of Hb activation and repression

We hypothesize the activating and repressing activities of Hb are mediated by different regions of the protein. The Hb protein sequence contains 4 zinc fingers implicated in DNA binding, a coiled-coil domain, and two more c-terminal zinc fingers implicated in dimerization (Figure 6.5) (Tautz et al., 1987; McCarty et al., 2003). Outside the zinc fingers, the majority of the protein is predicted to be intrinsically disordered (Dosztanyi et al., 2010). The Perrimon lab has mapped ten phosphorylation sites by mass spec, eight of which are perfectly conserved across twelve sequenced *Drosophila* species (Richelle Sopko and Norbert Perrimon, personal communication). Several computational models have proposed that Hb bifunctionality is controlled by dimerization (Papatsenko and Levine, 2008; Bieler et al., 2011). If the actives are

separable, they may be controlled by dimerization, phosphorylation or by distinct activation/repression domains.

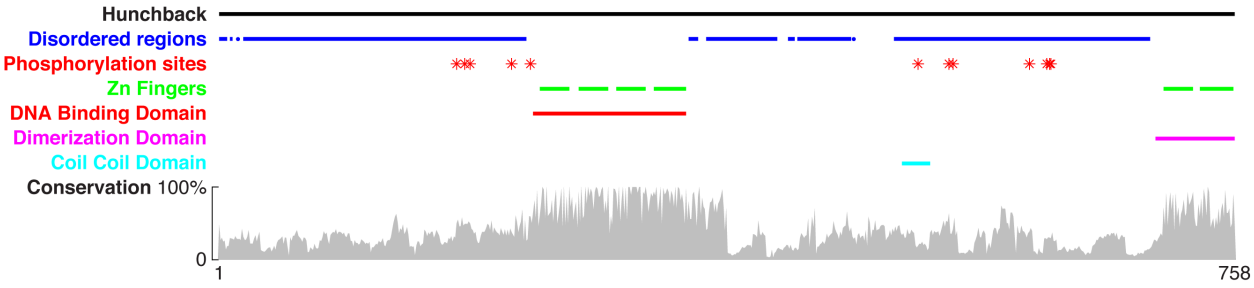


Figure 6.5 The Hb protein sequence annotated with domains and conservation

We will use the *eve* stripe 7 shadow enhancers to identify the protein elements responsible for activation and repression. We will use the *sna* promoter to mis-express truncated versions of the protein. By measuring the response of the *eve2+7* enhancer, we can identify defects in activation, i.e.. if we remove a part of the protein critical for activation, the *eve2+7* enhancer reporter will no longer expand. Alternatively, by measuring the response of the *eve3+7* enhancer reporter, we can identify defects in repression. Importantly, the dorsal region of the embryo is not perturbed and serves as a control.

To test for the role of dimerization we will remove the c-terminal zinc fingers, the coiled-coil domain, or both domains and measure the response of each reporter. If dimerization is essential for repression, as has been proposed by modeling studies, the *eve2+7* reporter should still expand (be activated). In contrast, the *eve3+7* reporter retreat will be less pronounced as purely monomeric Hb competes for binding sites and

may even behave like a dominant negative (Figure 6.6). If instead, activation is disrupted, the *eve3+7* response will remain unchanged and *eve2+7* will not expand.

We will further use this assay to investigate the role of the mapped phosphorylation sites by removing them or mutating them to mimic constitutive activation. To determine if bifunctionality is ancestral or derived we will examine distant orthologous Hb sequences. In the beetle Tribolium, a short germ-band insect, *hb* is the maternally deposited anterior determinant, analogous to *bcd* in Drosophila (Lynch and Desplan, 2003). An alternative way to separate activation and repression is to identify a coactivator or corepressor. Ben Vincent is using shRNA depletion in the female germline to test one candidate, *Mi-2,* which is known to be an important Hb corepressor after gastrulation (Kehle et al., 1998).

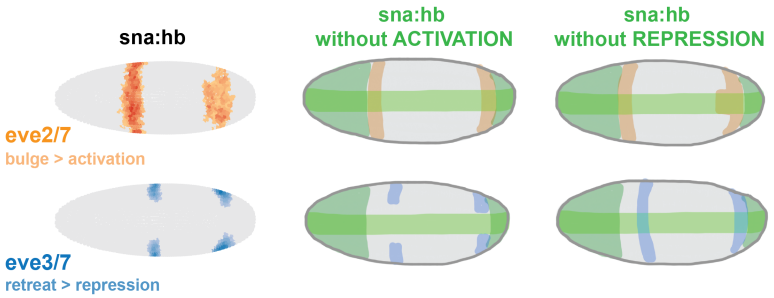


Figure 6.6 Predicted responses of the *eve2+7* and *eve3+7* enhancers when variant forms of Hb that behave as only activators or only repressors are expressed from the sna promoter. Figure credit: Angela DePace.

Separating the activating and repressing activities of Hb will allow us to do a targeted screen for additional targets that are differentially affected. We could perform *in situ* hybridizations agains many genes in *sna::hb-activation-only* and *sna::hb-repression-only* embryos to look for targets with differential responses. By identifying

additional examples we could test if 1) the Hb binding motif is subtly different in activated vs. repressed enhancers, and 2) test if a motif for another factor is consistently enriched in one class. Sequence signatures like these further rule out potential mechanisms of bifunctional regulation and help us predict targets and read genomes.

Building mechanistically accurate models of *eve2+7* and *eve3+7*

Improved computational models for the *eve2+7* and *eve3+7* enhancers will generate new hypotheses for how each enhancer computes and how the computations are combined at the promoter. Chapter 4 illustrates how we can test mechanistic predictions made by computational models. The next step is to use the new constraints offered by the data to build new computational models. These models will make new predictions which can use to iteratively prioritize the next round of experiments and test mechanistic insights.

Neither the linear model or the quadratic model can accurately predict the behavior of the *eve3+7* enhancer under perturbation. We have shown in Chapter 4 that the quadratic model is not an accurate predictor of the *eve3+7* enhancer under *sna:hb* perturbation. The linear model includes repression of *eve3+7* by *gt*, which is not directly supported by previous experimental work (Frasch and Levine, 1987; Small et al., 1996; Clyde et al., 2003; Struffi et al., 2011). Ben Vincent has performed additional experiments to show the linear model is not an accurate predictor of the *eve3+7* enhancer under *gt* perturbation. Using *gt* misexpression, he has shown *eve2+7* is sensitive to *gt*, but the *eve3+7* is not sensitive to *gt* (Figure 6.7). This finding implies that the linear model may also be a super position of the input/output functions of each

enhancer. Thus neither the quadratic model nor the linear model can capture the activity of the eve3+7 enhancer under perturbation.

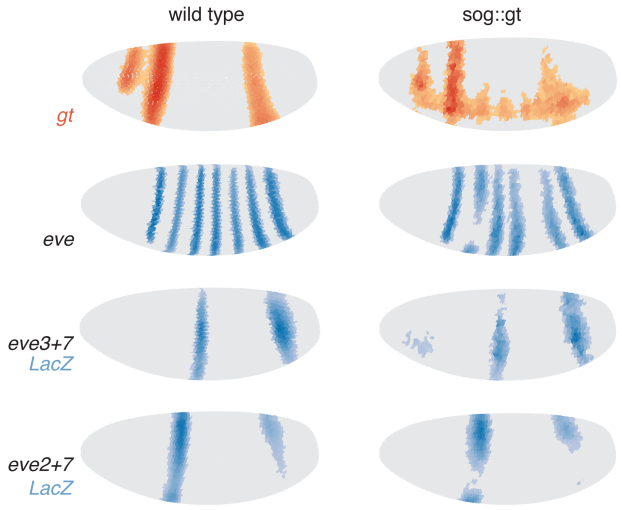


Figure 6.7: Misexpression reveals that *eve2+7* is sensitive to *gt* while *eve3+7* is not.

Building models of each enhancer would benefit from some additional data. We currently use Hb protein and *kni* mRNA data. Adding Kni protein data may help, because the Kni protein domain is slightly wider and posterior to the mRNA domain (according to the FlyEx data)(Pisarev et al., 2009). The current models may underestimate the ability of Kni to set the anterior boundary of stripe 7 alone, without *gt*. For *eve2+7,* the models developed by Ilsley et al. would be a good starting point, but we may need to add a posterior activator, such as Cad (Ilsley et al., 2013). We have not yet tested the *eve* stripe 2 models in the *sna::hb* gene expression atlas.

While logistic models do not include sequence, we can select model structures that assume different mechanisms, for example the hypothesis is that Cad coverts Hb

into an activator (Kim et al., 2013). Using mutations in enhancer reporter constructs, Ben Vincent has shown that Cad activates the *eve3+7* enhancer. If a Hb/Cad cross term can fit the *eve2+7* data more accurately, this improvement would support Hb/Cad activator synergy. We could then test this prediction with enhancer sequence mutations. In this way the models will formalize hypotheses for how Hb activates and represses different enhancers and lead to testable predictions.

Mechanistically accurate computational models of the *eve2+7* and *eve3+7* enhancers would further enable study of how the activities of the two enhancers are integrated at the promoter. We now believe that the quadratic model is an accurate predictor of the endogenous *eve* stripes 3 and 7 in WT and under perturbation because it is a super-position of the real input/output functions controlling each enhancer. By developing models for each enhancer input/output function, we can explore how two are combined and integrated at the promoter. Qualitatively, the data are consistent with simple addition, but more perturbations will be necessary to test this idea. In the long term, explaining why the quadratic model is such an accurate predictor of the endogenous pattern will inform how promoters integrate information from two enhancers that are active at the same time.

Implications of Hunchback bifunctional regulation for the embryo

What is the function of the incoherent input motif in the embryo?

We call the regulatory architecture of Hb activating and repressing distinct enhancers the *incoherent input motif* (Figure 6.8). This network motif resembles the incoherent feed forward motif, but the absence of the intermediate factor will have

important ramifications for circuit dynamics (Alon, 2007). Combining experimental and theoretical studies has previously been a successful way to approach and study network motifs (Goentoro and Kirschner, 2009; Goentoro and Kirschner, 2009; Ma et al., 2009). We propose to investigate the behavior of the incoherent input motif using toy models. We hypothesize that this motif will confer robustness to some kinds of stress at the expense of sensitivity to others.

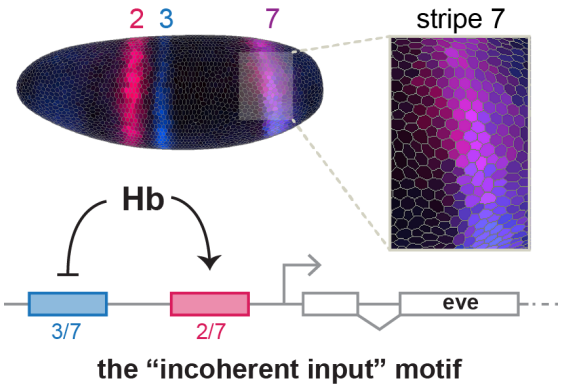


Figure 6.8 The *incoherent input motif.* Hb regulates eve streip 7 by repressing the *eve3+7* enhancer and activating the *eve2+7* enhancers

Do *eve* stripe 7 shadow enhancers contribute to robustness of embryonic patterning?

The the *eve2+7* and *eve3+7* enhancers provide a unique opportunity to pinpoint how bifunctional regulation contributes to embryonic properties like precision and robustness. In other loci, shadow enhancers confer robustness to environmental or genetic stress (Boettiger and Levine, 2009; Frankel et al., 2010; Perry et al., 2010; Dunipace et al., 2011). In this locus, we can first determine if these shadows confer robustness and then investigate the specific role of Hb bifunctional regulation.

We will look for whether the *eve* stripe 7 shadow enhancers increase precision and robustness. The peak position of *eve* stripe 7 in WT is the most variable stripe (Dubuis et al., 2013a). Using large reporter constructs, we will remove each enhancer and measure the precision of stripe 7 position (Figure 6.9). Removing *eve2+7* will have strong effects on *eve* stripe 2 and removing *eve3+7* will likely have strong effects on *eve* stripe 3, but for *eve* stripe 7, the enhancers may compensate. We expect stripe 7 will only disappear when we remove both enhancers. We can next expose these reporters to environmental stress (high temperature) or genetic stress (heterozygotes) and look for changes in position and precision of stripe 7.

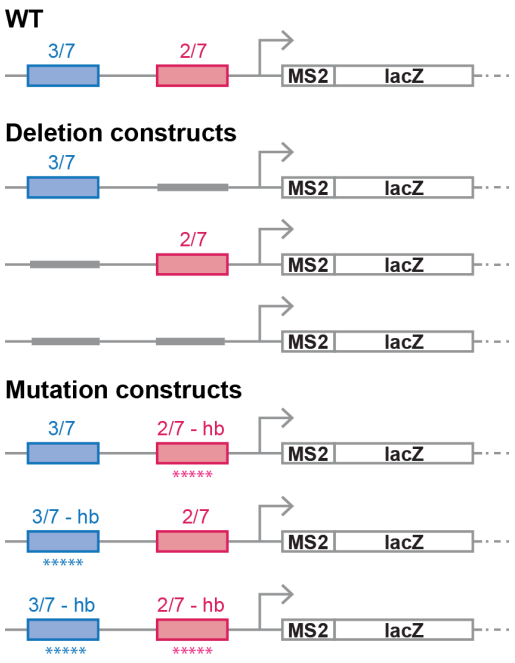


Figure 6.9 Reporter constructs for isolating the role of Hb in controlling the precision and robustness of eve stripe 7. The first set will delete either or both enhancers to establish a baseline. Comparison the second set will isolate the role of Hb bifunctional regulation.

Next, we will isolate the role of Hb bifunctional regulation in establishing the precision and robustness of *eve* stripe 7 (Figure 6.9). In reporter constructs, we will remove all the predicted Hb binding sites in *eve3+7*, *eve2+7*, and the entire region. We will again measure precision of *eve* stripe 7 in WT and under stress. The difference between these mutation constructs and the enhancer deletion constructs will isolate the contribution of Hb to precision and robustness. The key innovation over existing work is that we will be able to look at the role of bifunctional regulation, specifically the incoherent input motif, in controlling emergent embryonic properties.

Theoretical analysis: Can a bifunctional auto-regulatory transcription factor clean up its own gene expression pattern?

What happens when a bifunctional transcription factor regulates itself? This special instance of the incoherent input motif is called the *bifunctional auto-regulatory motif* (Figure 6.10). The *auto regulatory motif* can speed approaches to steady state (Alon, 2007). It is not clear if the bifunctional auto-regulatory motif actually exists in nature, but *Kr* is a known repressor that is also thought to activate its own expression (Jaeger, 2011). Kr protein has also been shown to be bifunctional in cell culture, but there is no direct evidence for bifunctionality in the embryo (Sauer and Jäckle, 1991; Sauer and Jäckle, 1993). Given the abundance of auto regulatory connections and the likelihood that bifunctional factors are under annotated, this motif may yet be discovered in other contexts. We focused on the potential role in embryonic patterning: does there exist an regime where the bifunctional auto-regulatory motif can sharpen its own pattern?

Given a few assumptions, it is possible for the bifunctional auto-regulatory motif to sharpen its own spatial pattern over time. Let us assume there is a transcriptional factor that dimerizes with standard Michaelis-Menton kinetics (Figure 6.10A). We further assume that as a monomer it is a repressor and as a dimer it is an activator. Now let us consider the situation that this factor regulates its own transcription. For simplicity, we assume monomers and dimers have the same DNA binding affinity and compete for a single binding site (Figure 6.10F). We further assume separation of time scales: dimerization is fast compared to DNA binding which is in turn fast compared to

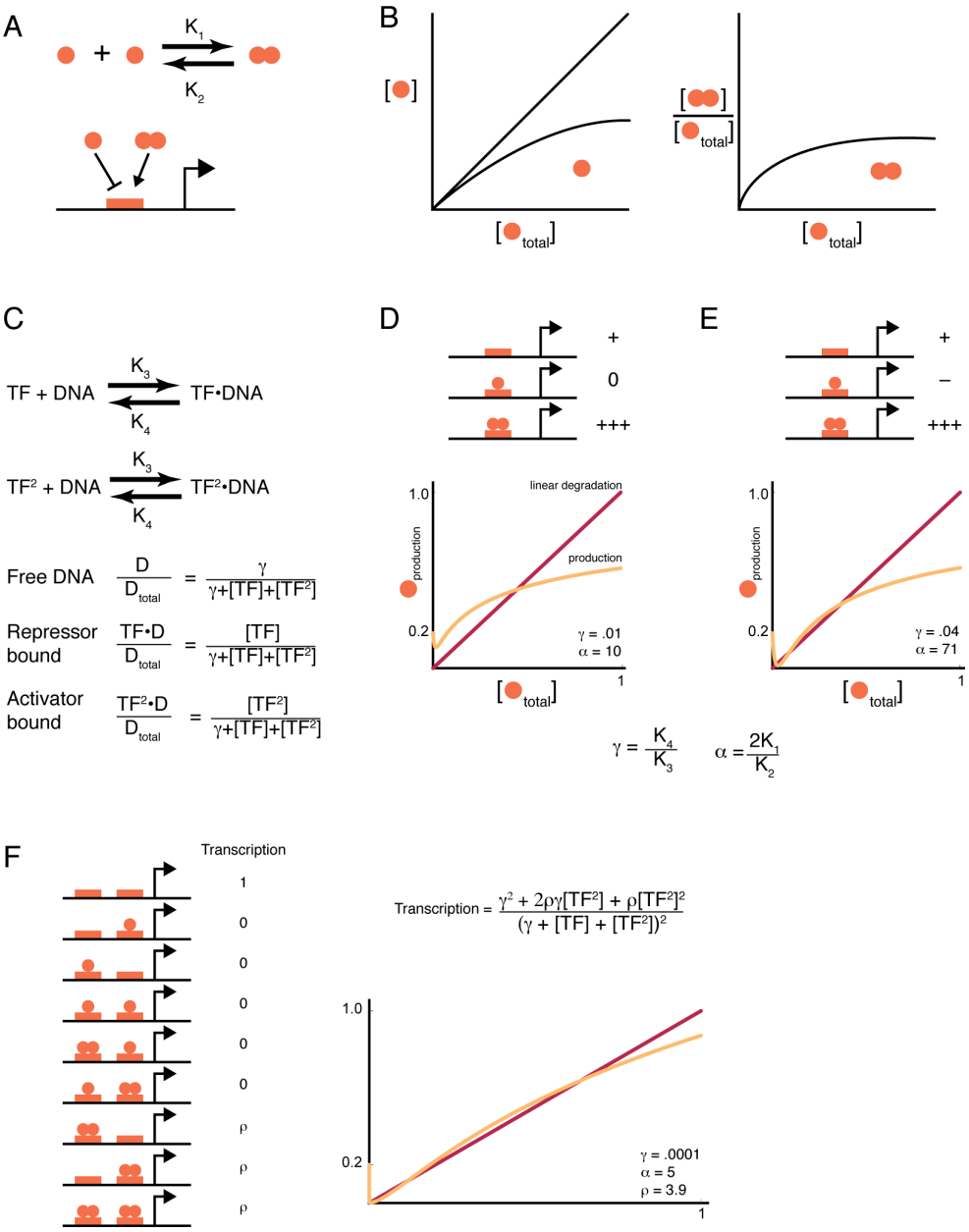


Figure 6.10 The bifunctional auto-regulatory transcription factor can sharpen its own spatial pattern

A) We assume monomers repress and the dimers activate transcription. B) Left, the amount of free monomer. Right, the fraction of total TF in dimers. C) We assume monomers and dimers compete for a single DNA binding site with the same affinity. D) If the repressor brings transcription to zero, this motif has only one stable state, ON, and it cannot sharpen its own pattern. E) If the repressor bound state causes negative transcription, there are are 3 steady states: stable Off, unstable transition and stable ON. In this regime the system exhibits bistability allowing the motif to sharpen a pattern. F) If there are 2 binding sites and the repressor is dominant, the system can have 3 steady states and exhibit bistability.

transcription and translation. Intuitively, there are two stable states: at high concentrations, the factor will mostly dimerize and activate transcription, leading to positive feedback that keeps the levels high; in contrast at low factor concentrations, repressing monomers will be more common, and negative feedback keeps levels low (Figure 6.11). Bistability is possible, but in a limited parameter regime. If we assume that basal transcription is zero (off) and that the repressor binding also causes zero transcription (off) then the system cannot have two stable states (Figure 6.10D). Only when the repressor causes negative transcription, can this simple system have two stable states, as shown by the production and degradation curves crossing three times (Figure 6.10E). This analysis implies that without negative transcription, a bifunctional auto-regulatory transcription factor cannot sharpen its own pattern using a single binding site.

If we consider two independent binding sites, we can invoke one possible mechanism for negative transcription, and sharpening can occur (Figure 6.10F). The analog of negative transcription we assume is dominate repression: that if either site is bound by a monomer, transcription is repressed, even if an activating dimer is also bound. Under this assumption there is a tiny parameter regime where the system is bistable. This analysis suggests this motif will require multiple binding sites. Together this theoretical analysis suggests 1) There is a small regime where an auto-regulatory transcription factor can sharpen its own pattern is space and time, 2) This regime requires basal transcription (additional activators) that the repressor must inhibit and 3) multiple binding sites are required (Figure 6.11). Going forward, we should look for the bifunctional auto-regulatory motif as we investigate new networks. If we can confirm its

existence, then we can examine its role in generating precision and robustness. Alternatively, we could attempt to build a synthetic implantation of this network motif.

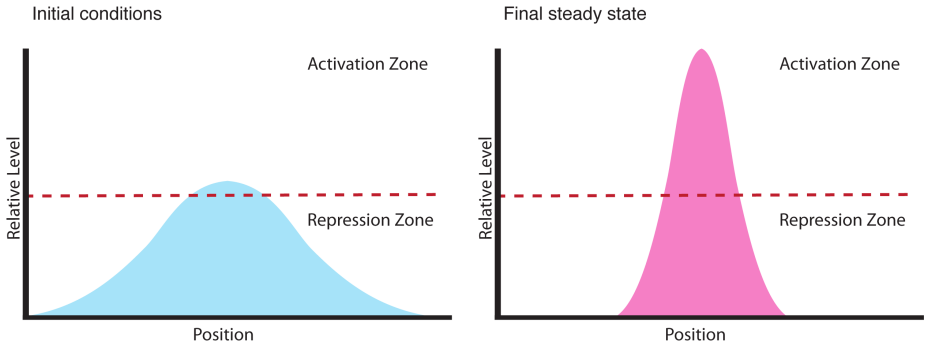


Figure 6.11 When there are two stable steady states, the bifunctional auto regulatory motif can sharpen its own spatial pattern over time.
Note the motif relies on asymmetric input from other regulators.

Speculative Hypothesis: shadow enhancers prevent the segmentation network from specifying a narrow *eve* stripe 7

We hypothesize that by using different regulators, shadow enhancers can prevent the creation of a narrow stripe 7. There may be an evolutionary advantage for the system to be biased towards creating a wider rather than narrower *eve* stripe 7: if a stripe is specified with too may cells (too wide), apoptosis increases to compensate, but if a stripe is too narrow, the entire segment defined by that stripe can be lost later in development (Hughes and Krause, 2001). If *eve* stripe 7 is too wide the worst that can happen is that the adjacent segment specified by *ftz* stripe 6 is lost. The body segment specified by *eve* stripe 7 is essential, but the segment patterned by the *ftz* stripe 6 is

likely dispensable for viability (Fujioka et al., 2002). Although this hypothesis is hard to prove, we propose the system may be biased towards producing a wider stripe 7.

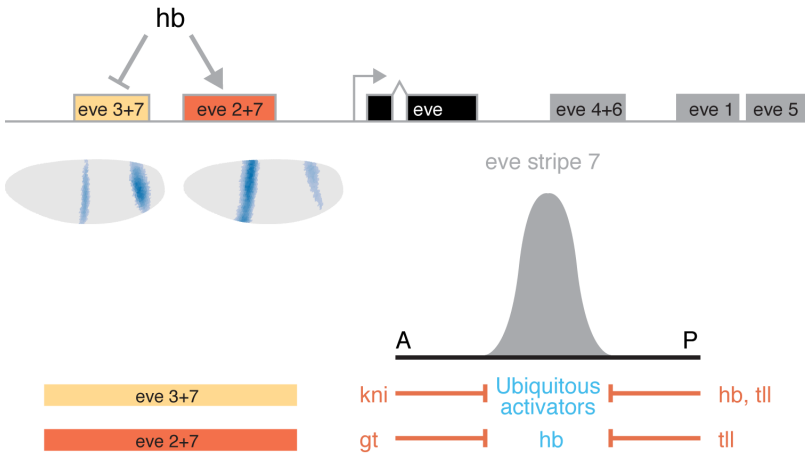


Figure 6.12 Model for Hb bifunctional regulation of eve stripe 7
The two shadow enhancers used different regulators to set boundaries. Based on figure from Ben Vincent

We propose that the anterior boundary of *eve3+7* is set by *kni* and the anterior boundary of *eve2+7* is set by *gt.* Ben Vincent used misexpression to show *gt* represses *eve* stripe 2, but not stripes 3 or 7, similar to previously reported data (Figure 6.7) (Andrioli et al., 2002; Ribeiro et al., 2010). The reporters tell a different story: both stripes from the *eve2+7* reporter are repressed, while the the *eve3+7* reporter pattern is unchanged. From these data and extensive evidence that Kni directly represses *eve3+7* (Struffi et al., 2011), we propose that anterior boundary of *eve3+7* is set by *kni* and the anterior boundary of *eve2+7* is set by *gt,* as proposed by other modeling efforts (Figure 6.12) (Janssens et al., 2006).

The *sna::hb* and *sog::gt* misexpression experiments suggest that if at least one of the enhancers is ON, then the endogenous pattern will also be ON. In *sna::hb, eve2+7*

expands and the endogenous stripe expands. In *sog:gt,* stripe 7 in *eve2+7* is repressed, but *since eve3+7* is active, the endogenous stripe 7 is not repressed. While it is clear that within an enhancer repression is dominant, at the locus level, activation appears to dominate. Simple addition may be a good approximation for how the promoter integrates information from both enhancers. For the remainder of this section, we will assume that if either enhancers is active, then the endogenous gene will be ON.

If a different repressor sets the boundary of each enhancer and if one active enhancer is enough to activate the endogenous pattern, the system could guard against the creation of a narrow stripe. The anterior stripe 7 boundary on *eve3+7* is set by low concentrations of Kni, making fluctuations from gene expression noise more significant. If Kni fluctuates downward, then *eve3+7* will be wide; if Kni fluctuates upward, then *eve3+7* will be narrow. In addition, Gt is on the X chromosome and early, non-canonical dosage compensation while very good, is not perfect, making Gt levels a potentially variable input (Lott et al., 2011). Indeed, the anterior boundary of *eve* stripe 2, also set by Gt, is quite variable from embryo to embryo (Manu et al., 2013). If Gt fluctuates downward, then *eve2+7* will be wide; if Gt fluctuates upward, then *eve2+7* will be narrow. Only in cases when Gt and Kni fluctuate upward would the stripe be narrow (1/4 of cases). Since Gt weakly represses Kni (Jaeger, 2011), this scenario might be rare and the system would be biased towards creating a wider stripe.

Testing this hypothesis would require several steps. First we would need to more carefully show that *gt* and not *kni* is setting the boundary of *eve2+7.* We predict that *eve2+7* would have a distinct boundary in a *kni* mutant (*eve3+7* does not) (Struffi et al., 2011)). Second, we can try to force a simultaneous increase in Gt and Kni levels using

duplications lines and look for changes in *eve* stripe 7 expression or tail defects in the cuticle. Third, we could delete each enhancer from the endogenous locus and examine *eve* stripe 7 expression and the posterior region of larval cuticles. When both enhancers are missing, we expect defects in A8 and the tail. If either enhancer is sufficient for eve stripe 7, we expect A8 and the tail to be intact in larvae missing one or the other enhancer raised under normal lab conditions. It is possible that defects will arise only when these larvae are stressed genetically or environmentally.

Despite the storytelling appeal of the evolutionary arguments presented, metazoan evolution is dominated by genetic drift (Lynch, 2007). The existence of the *eve* stripe 7 shadow enhancers may merely be a product of neutral genetic drift. There may be no purifying selection against an enhancer expressing where another enhancer is already active. If deleting shadow enhancers does not cause defects in eve stripe 7 expression, we may conclude drift is responsible for this network motif.

Does the incoherent input motif increase information transfer through the network?

It may be possible that using two enhancers to respond to Hb in opposite ways may enable the network build an error correcting code. The gap genes all regulate each other and regulate *eve* stripe 7. If fluctuations in maternal inputs dominate extrinsic noise, these fluctuations could propagate through the network in a consistent way. Over evolutionary time scales, the network could “learn” the cross-correlation matrix of gap gene fluctuations and evolved to compensate for errors. Optimal information transfer is achieved when the cumulative probability density of system outputs matches the

cumulative probability density of the inputs (Laughlin, 1981). The incoherent input motif may enable stripe 7 expression to better match input space of WT fluctuations. For example, increases in Hb likely cause decreases in kni, potentially shifting *eve3+7* anteriorly, but the increased *eve2+7* (from Hb activation) may compensate.

Closing remarks

The Drosophila blastoderm embryo is an exciting system for building and testing models of how enhancer compute cell type specific gene expression patterns. We can measure expression in every cell of intact animals and combine data from many embryos to create gene expression atlases. These atlases are will suited for modeling input/output functions. We have developed shRNA techniques for efficiently perturbing embryonic gene expression. Building a gene expression atlas of one maternal perturbation revealed very strong and early canalization of cell fate in individual cells. We used the indirect effects of this perturbation to test models for how one enhancer was computing a specific pattern, revealing that the canonical annotation of the locus was incomplete. As the technology for making measurements and perturbations improves, we will be able to refine models of enhancer computation to include sequence and predict the effect of naturally occurring regulatory mutations. We can look forward to a reading developmental programs directly from genome sequence.

References

Alon, U. (2007). Introduction to Systems Biology: And the Design Principles of Biological Networks. Chapman & Hall/CRC. Boca Raton.

Andrioli, L. P. M., Vasisht, V., Theodosopoulou, E., Oberstein, A. L. and Small, S. (2002). Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. *Development (Cambridge, England)* **129**, 4931-4940.

Berg, H. C. and Purcell, E. M. (1977). Physics of chemoreception. *Biophys J* **20**, 193-219.

Bialek, W. and Setayeshgar, S. (2005). Physical limits to biochemical signaling. *Proc Natl Acad Sci U S A* **102**, 10040-10045.

Bieler, J., Pozzorini, C. and Naef, F. (2011). Whole-Embryo Modeling of Early Segmentation in Drosophila Identifies Robust and Fragile Expression Domains. *Biophysj* **101**, 287-296.

Boettiger, A. N. and Levine, M. S. (2009). Synchronous and Stochastic Patterns of Gene Activation in the Drosophila Embryo. *Science* **325**, 471-473.

Carroll, S. B. and Vavra, S. H. (1989). The zygotic control of Drosophila pair-rule gene expression. II. Spatial repression by gap and pair-rule gene products. *Development (Cambridge, England)* **107**, 673-683.

Casanova, J. and Struhl, G. (1993). The torso receptor localizes as well as transduces the spatial signal specifying terminal body pattern in Drosophila. *Nature* **362**, 152-155.

Clyde, D. E., Corado, M. S. G., Wu, X., Pare, A., Papatsenko, D. and Small, S. (2003). A self-organizing system of repressor gradients establishes segmental complexity in Drosophila. *Nature* **426**, 849-853.

Dosztanyi, Z., Meszaros, B. and Simon, I. (2010). Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. *Brief Bioinform* **11**, 225-243.

Dubuis, J. O., Samanta, R. and Gregor, T. (2013). Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol Syst Biol* **9**, 639.

Dubuis, J. O., Samanta, R. and Gregor, T. (2013). Accurate measurements of dynamics and reproducibility in small genetic networks. *Molecular Systems Biology* **9**, 639.

Dubuis, J. O., Tkačik, G., Wieschaus, E. F., Gregor, T. and Bialek, W. (2013). Positional information, in bits. *Proceedings of the National Academy of Sciences* **110**, 16301-16308.

Dunipace, L., Ozdemir, A. and Stathopoulos, A. (2011). Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. *Development* **138**, 4075-4084.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F. and Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490-493.

Frasch, M. and Levine, M. (1987). Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in Drosophila. *Genes & development* **1**, 981-995.

Fujioka, M., Miskiewicz, P., Raj, L., Gulledge, A. A., Weir, M. and Goto, T. (1996). Drosophila Paired regulates late even-skipped expression through a composite

binding site for the paired domain and the homeodomain. *Development (Cambridge, England)* **122**, 2697-2707.

Fujioka, M., Yusibova, G. L., Patel, N. H., Brown, S. J. and Jaynes, J. B. (2002). The repressor activity of Even-skipped is highly conserved, and is sufficient to activate engrailed and to regulate both the spacing and stability of parasegment boundaries. *Development (Cambridge, England)* **129**, 4411-4421.

Goentoro, L. and Kirschner, M. W. (2009). Evidence that fold-change, and not absolute level, of beta-catenin dictates Wnt signaling. *Mol Cell* **36**, 872-884.

He, X., Samee, M. A. H., Blatti, C. and Sinha, S. (2010). Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Computational Biology* **6**,

Hengenius, J. B., Gribskov, M., Rundell, A. E., Fowlkes, C. C. and Umulis, D. M. (2011). Analysis of Gap Gene Regulation in a 3D Organism-Scale Model of the Drosophila melanogaster Embryo. *PLoS ONE* **6**, e26797.

Hughes, S. C. and Krause, H. M. (2001). Establishment and maintenance of parasegmental compartments. *Development* **128**, 1109-1118.

Ilsley, G. R., Fisher, J., Apweiler, R., DePace, A. H. and Luscombe, N. M. (2013). Cellular resolution models for even skipped regulation in the entire Drosophila embryo. *eLife* **2**, e00522.

Ingham, P. and Gergen, P. (1988). Interactions between the pair-rule genes runt, hairy, even-skipped and fushi tarazu and the establishment of periodic pattern in the Drosophila embryo. *Development* **104**, 51-60.

Ingolia, N. T. (2004). Topology and robustness in the Drosophila segment polarity network. *PLoS biology* **2**, e123.

Jaeger, J., Sharp, D. H. and Reinitz, J. (2007). Known maternal gradients are not sufficient for the establishment of gap domains in Drosophila melanogaster. *Mech Dev* **124**, 108-128.

Jaeger, J. (2011). The gap gene network. *Cellular and Molecular Life Sciences* **68**, 243-274.

Jaeger, J., Blagov, M., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Surkova, S., Vanario-Alonso, C. E., Samsonova, M., Sharp, D. H. et al. (2004). Dynamical analysis of regulatory interactions in the gap gene system of Drosophila melanogaster. *Genetics* **167**, 1721-1737.

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Vanario-Alonso, C. E., Samsonova, M. et al. (2004). Dynamic control of positional information in the early Drosophila embryo. *Nature* **430**, 368-371.

Janssens, H., Hou, S., Jaeger, J., Kim, A.-R., Myasnikova, E., Sharp, D. and Reinitz, J. (2006). Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. *Nature genetics* **38**, 1159-1165.

Jiang, J., Hoey, T. and Levine, M. (1991). Autoregulation of a segmentation gene in Drosophila: combinatorial interaction of the even-skipped homeo box protein with a distal enhancer element. *Genes Dev* **5**, 265-277.

Kazemian, M., Blatti, C., Richards, A., McCutchan, M., Wakabayashi-Ito, N., Hammonds, A. S., Celniker, S. E., Kumar, S., Wolfe, S. A., Brodsky, M. H. et al.

(2010). Quantitative analysis of the Drosophila segmentation regulatory network using pattern generating potentials. *PLoS biology* **8**,

Kehle, J., Beuchle, D., Treuheit, S., Christen, B., Kennison, J. A., Bienz, M. and Müller, J. (1998). dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science* **282**, 1897-1900.

Kim, A.-R., Martinez, C., Ionides, J., Ramos, A. F., Ludwig, M. Z., Ogawa, N., Sharp, D. H. and Reinitz, J. (2013). Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic. *PLoS Genetics* **9**, e1003243.

Kondo, S. (2002). The reaction-diffusion system: a mechanism for autonomous pattern formation in the animal skin. *Genes to cells : devoted to molecular & cellular mechanisms* **7**, 535-541.

Kozlov, K., Surkova, S., Myasnikova, E., Reinitz, J. and Samsonova, M. (2012). Modeling of gap gene expression in Drosophila Kruppel mutants. *PLoS Computational Biology* **8**, e1002635.

Laughlin, S. B. (1981). A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch* **36**, 51.

Lawrence, P. A. (1992). The making of a fly: the genetics of animal design.

Lott, S. E., Villalta, J. E., Schroth, G. P., Luo, S., Tonkin, L. A. and Eisen, M. B. (2011). Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq. *PLoS Biol* **9**, e1000590.

Lynch, J. and Desplan, C. (2003). 'De-evolution' of Drosophila toward a more generic mode of axis patterning. *The International Journal of Developmental Biology* **47**, 497-503.

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences of the United States of America* **104 Suppl 1**, 8597-8604.

Ma, W., Trusina, A., El-Samad, H., Lim, W. A. and Tang, C. (2009). Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760-773.

Ma, W., Lai, L., Ouyang, Q. and Tang, C. (2006). Robustness and modular design of the Drosophila segment polarity network. *Molecular Systems Biology* **2**, 70.

Manoukian, A. S. and Krause, H. M. (1992). Concentration-dependent activities of the even-skipped protein in Drosophila embryos. *Genes Dev* **6**, 1740-1751.

Manu, Ludwig, M. Z. and Kreitman, M. (2013). Sex-specific pattern formation during early Drosophila development. *Genetics* **194**, 163-173.

Manu, Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M. et al. (2009). Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. *PLoS Computational Biology* **5**, e1000303.

Manu, Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M. et al. (2009). Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. *PLoS biology* **7**, e1000049.

McCarty, A. S., Kleiger, G., Eisenberg, D. and Smale, S. T. (2003). Selective dimerization of a C2H2 zinc finger subfamily. *Molecular Cell* **11**, 459-470.

Nasiadka, A. and Krause, H. M. (1999). Kinetic analysis of segmentation gene interactions in Drosophila embryos. *Development* **126**, 1515-1526.
Papatsenko, D. and Levine, M. S. (2008). Dual regulation by the Hunchback gradient in the Drosophila embryo. *Proc Natl Acad Sci U S A* **105**, 2901-2906.
Papatsenko, D. and Levine, M. S. (2011). The Drosophila gap gene network is composed of two parallel toggle switches. *PLoS ONE* **6**, e21145.
Perry, M. W., Boettiger, A. N., Bothma, J. P. and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. *Curr Biol* **20**, 1562-1567.
Peter, I. S., Faure, E. and Davidson, E. H. (2012). Predictive computation of genomic logic processing functions in embryonic development. *Proceedings of the National Academy of Sciences* **109**, 16434-16442.
Pisarev, A., Poustelnikova, E., Samsonova, M. and Reinitz, J. (2009). FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic acids research* **37**, D560-D566.
Ren, X., Sun, J., Housden, B. E., Hu, Y., Roesel, C., Lin, S., Liu, L. P., Yang, Z., Mao, D., Sun, L. et al. (2013). Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. *Proc Natl Acad Sci U S A* **110**, 19012-19017.
Ribeiro, T. C., Ventrice, G., Machado-Lima, A. and Andrioli, L. P. (2010). Investigating giant (Gt) repression in the formation of partially overlapping pair-rule stripes. *Developmental dynamics : an official publication of the American Association of Anatomists* **239**, 2989-2999.
Samee, A. H. and Sinha, S. (2013). Evaluating thermodynamic models of enhancer activity on cellular resolution gene expression data. *Methods* **62**, 79-90.
Sánchez, L., Chaouiya, C. and Thieffry, D. (2008). Segmenting the fly embryo: logical analysis of the role of the segment polarity cross-regulatory module. *The International Journal of Developmental Biology* **52**, 1059-1075.
Sánchez, L. and Thieffry, D. (2003). Segmenting the fly embryo: a logical analysis of the pair-rule cross-regulatory module. *Journal of theoretical biology* **224**, 517-537.
Sauer, F. and Jäckle, H. (1991). Concentration-dependent transcriptional activation or repression by Krüppel from a single binding site. *Nature* **353**, 563-566.
Sauer, F. and Jäckle, H. (1993). Dimerization and the control of transcription by Krüppel. *Nature* **364**, 454-457.
Saulier-Le Drean, B., Nasiadka, A., Dong, J. and Krause, H. M. (1998). Dynamic changes in the functions of Odd-skipped during early Drosophila embryogenesis. *Development* **125**, 4851-4861.
Schroeder, M. D., Greer, C. and Gaul, U. (2011). How to make stripes: deciphering the transition from non-periodic to periodic patterns in Drosophila segmentation. *Development (Cambridge, England)* **138**, 3067-3078.
Schulz, C. and Tautz, D. (1994). Autonomous concentration-dependent activation and repression of Kruppel by hunchback in the Drosophila embryo. *Development* **120**, 3043-3049.
Segal, E., Raveh-Sadka, T., Schroeder, M. D., Unnerstall, U. and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature* **451**, 535-540.

Sharp, D. H. and Reinitz, J. (1998). Prediction of mutant expression patterns using gene circuits. *Bio Systems* **47**, 79-90.

Siegal, M. L. and Bergman, A. (2002). Waddington's canalization revisited: developmental stability and evolution. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10528-10532.

Small, S., Blair, A. and Levine, M. S. (1996). Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. *Developmental biology* **175**, 314-324.

Sokolowski, T. R., Erdmann, T. and Ten Wolde, P. R. (2012). Mutual repression enhances the steepness and precision of gene expression boundaries. *PLoS Computational Biology* **8**, e1002654.

Struffi, P., Corado, M., Kaplan, L., Yu, D., Rushlow, C. and Small, S. (2011). Combinatorial activation and concentration-dependent repression of the Drosophila even skipped stripe 3+7 enhancer. *Development (Cambridge, England)* **138**, 4291-4299.

Struhl, G., Johnston, P. and Lawrence, P. A. (1992). Control of Drosophila body pattern by the hunchback morphogen gradient. *Cell* **69**, 237-249.

Suleimenov, Y., Ay, A., Samee, M. A. H., Dresch, J. M., Sinha, S. and Arnosti, D. N. (2013). Global parameter estimation for thermodynamic models of transcriptional regulation. *Methods* **62**, 99-108.

Tautz, D., Lehmann, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jäckle, H. (1987). Finger protein of novel structure encoded by hunchback, a second member of the gap class of Drosophila segmentation genes. *Nature* **327**, 383-389.

Teytelman, L., Thurtle, D. M., Rine, J. and van Oudenaarden, A. (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proceedings of the National Academy of Sciences* **110**, 18602-18607.

von Dassow, G., Meir, E., Munro, E. M. and Odell, G. M. (2000). The segment polarity network is a robust developmental module. *Nature* **406**, 188-192.

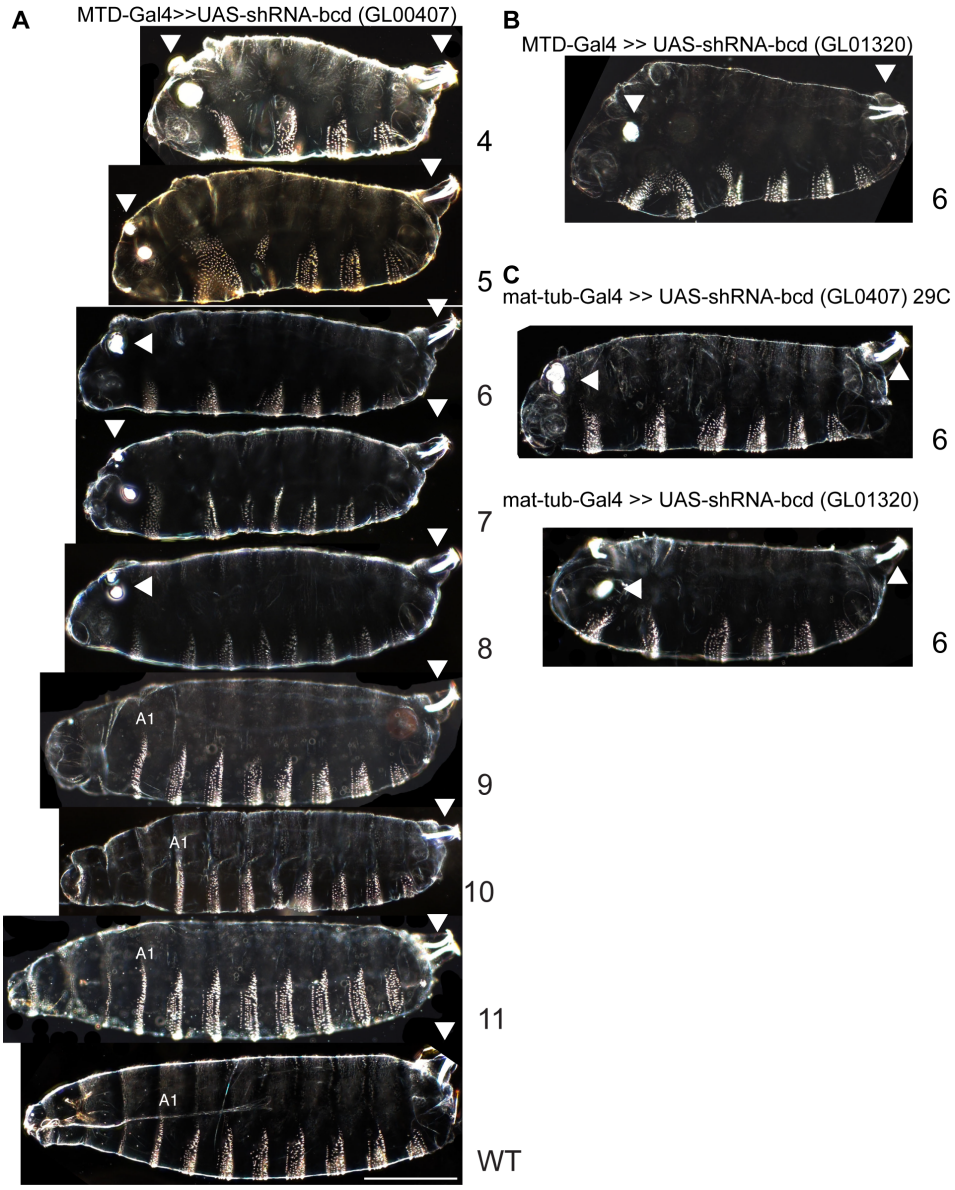
Von Dassow, G. and Odell, G. M. (2002). Design and constraints of the Drosophila segment polarity module: robust spatial patterning emerges from intertwined cell state switches. *The Journal of experimental zoology* **294**, 179-215.

Appendix A: Supplementary Materials for Chapter 2

Appendix A Table S1: List of oligos for cloning shRNAs in miRNA backbones. Complementary oligos were annealed and cloned into VALIUM20 as described in the methods.

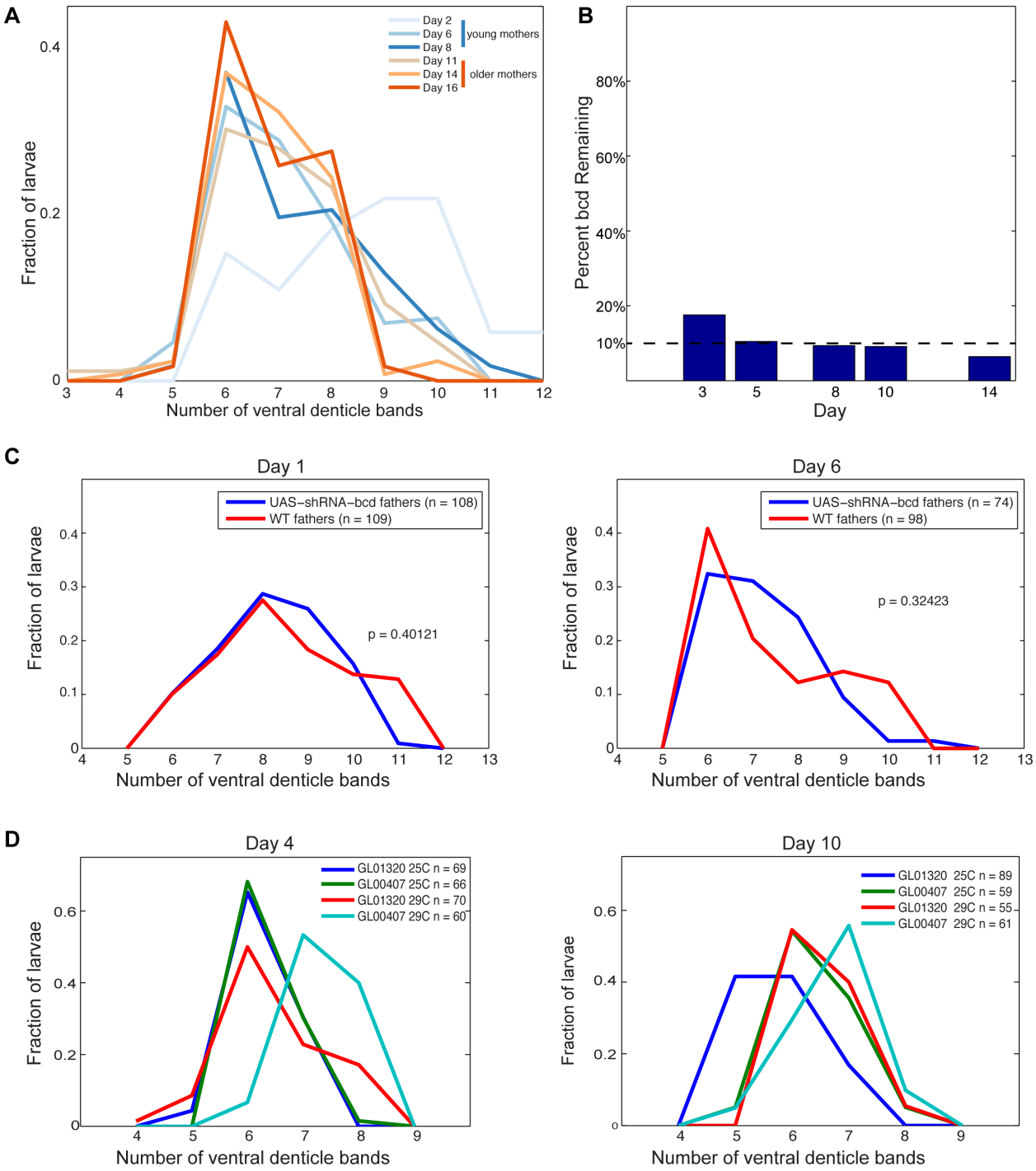
mir-275 backbone		
Gene Name	**Top Oligo**	**Bottom Oligo**
ovarian tumor (otu)	ctagctgtaaagtctcctaccttgGCAGAACAACACTGATCAACActggttttttatatacagTGTTGATCAGTGTTGTTCTGCcgtggtggcagacatatatg	aattcatatatgtctgccaccacgGCAGAACAACACTGATCAACActgtatataaaaaaccagTGTTGATCAGTGTTGTTCTGCcaaggtaggagactttacag
bicoid (bcd)	ctagctgtaaagtctcctaccttgAACGGGAGCGATAAACTACAActggttttttatatacagTTGTAGTTTATCGCTCCCGTTcgtggtggcagacatatatg	aattcatatatgtctgccaccacgAACGGGAGCGATAAACTACAActgtatataaaaaaccagTTGTAGTTTATCGCTCCCGTTcaaggtaggagactttacag
giant (gt)	ctagctgtaaagtctcctaccttgCAGCTAGCTATTAATGTTTAActggttttttatatacagTTAAACATTAATAGCTAGCTGcgtggtggcagacatatatg	aattcatatatgtctgccaccacgCAGCTAGCTATTAATGTTTAActgtatataaaaaaccagTTAAACATTAATAGCTAGCTGcaaggtaggagactttacag
miR-92a backbone		
Gene Name	**Top Oligo**	**Bottom Oligo**
Notch (N)	ctagcaatatgaatttcccgGCGGCGGTTAACAATACCGAAttttgcatttcgaataaaTTCGGTATTGTTAACCGCCGCtgggcggtttgtaataaacag	aattctgtttattacaaaccgcccaGCGGCGGTTAACAATACCGAAtttattcgaaatgcaaaaTTCGGTATTGTTAACCGCCGCcgggaaattcatattg
bicoid (bcd)	ctagcaatatgaatttcccgAACGGGAGCGATAAACTACAAttttgcatttcgaataaaTTGTAGTTTATCGCTCCCGTTtgggcggtttgtaataaacag	aattctgtttattacaaaccgcccaAACGGGAGCGATAAACTACAAtttattcgaaatgcaaaaTTGTAGTTTATCGCTCCCGTTcgggaaattcatattg
Kruppel (Kr)	ctagcaatatgaatttcccgTTGTTGCTGCTTCAAATATAAttttgcatttcgaataaaTTATATTTGAAGCAGCAACAAtggcggtttgtaataaacag	aattctgtttattacaaaccgcccaTTGTTGCTGCTTCAAATATAAtttattcgaaatgcaaaaTTATATTTGAAGCAGCAACAAcgggaaattcatattg
wingless (wg)	ctagcaatatgaatttcccgACGAATAGATTTCAAGAAGAAttttgcatttcgaataaaTTCTTCTTGAAATCTATTCGTtggcggtttgtaataaacag	aattctgtttattacaaaccgcccaACGAATAGATTTCAAGAAGAAtttattcgaaatgcaaaaTTCTTCTTGAAATCTATTCGTcgggaaattcatattg
armadillo (arm)	ctagcaatatgaatttcccgTACGATTGCTGTTCAACGAAAttttgcatttcgaataaaTTTCGTTGAACAGCAATCGTAtgggcggtttgtaataaacag	aattctgtttattacaaaccgcccaTACGATTGCTGTTCAACGAAAtttattcgaaatgcaaaaTTTCGTTGAACAGCAATCGTAcgggaaattcatattg

Appendix B: Supplementary Materials for Chapter 3



Appendix B Figure S1: The phenotype spectrum of *bcd* RNAi cuticles resembles an allelic series

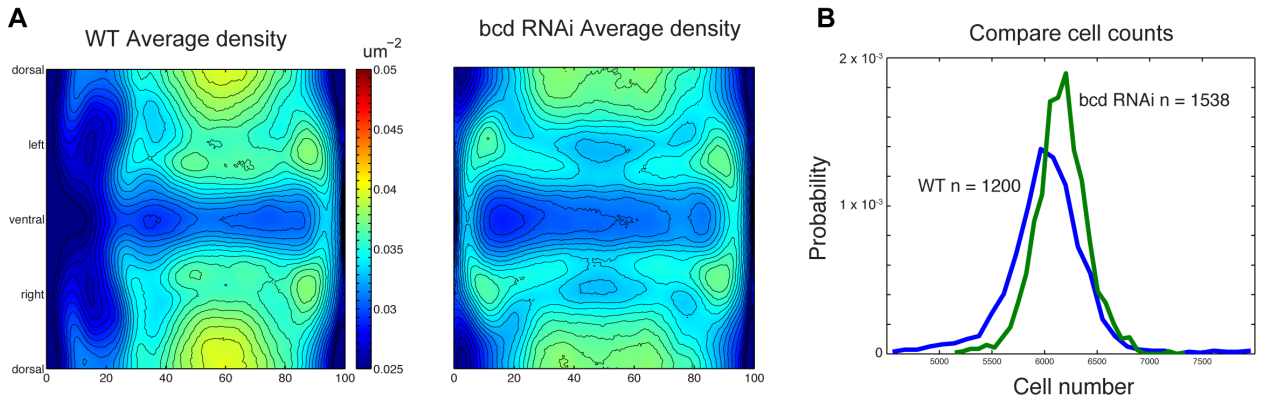
(A) Unhatched cuticles of embryos laid by *MTD-Gal4/UAS-shRNA-bcd* females display variable numbers of denticle bands. A representative individual from each class is shown. The filzkörper, a tail structure, is indicated with arrow heads. Cuticles with weak *bcd* phenotypes have 9 or more bands and some thoracic segments, but no head structures. Cuticles with strong *bcd* phenotypes have only abdominal segments and a duplicated filzkörper. Scale bar 200 microns. (B) An additional *UAS-shRNA-bcd* (GL01320) line gives a similar phenotype. (C) The maternal driver *mat-tub-Gal4* gives a similar phenotype with both *UAS-shRNA-bcd* lines. To test viability, we arrayed 200 embryos on an agar plate and counted the number hatched after 48 hrs at 25C. For *bcd* RNAi embryos, 0/200 and 0/200 embryos hatched; for embryos laid by *MTD-Gal4/UAS-shRNA-GFP* females 222/239 and 178/200 of embryos hatched (TRiP Toolbox stock 182) (Neumuller et al., 2012).



Appendix B Figure S2: Mother age contributes strongly to the strength and variability of the *bcd* RNAi phenotype while temperature, *UAS-shRNA-bcd* line, maternal driver, zygotic *UAS-shRNA-bcd* construct copy number, and paternal genotype do not.

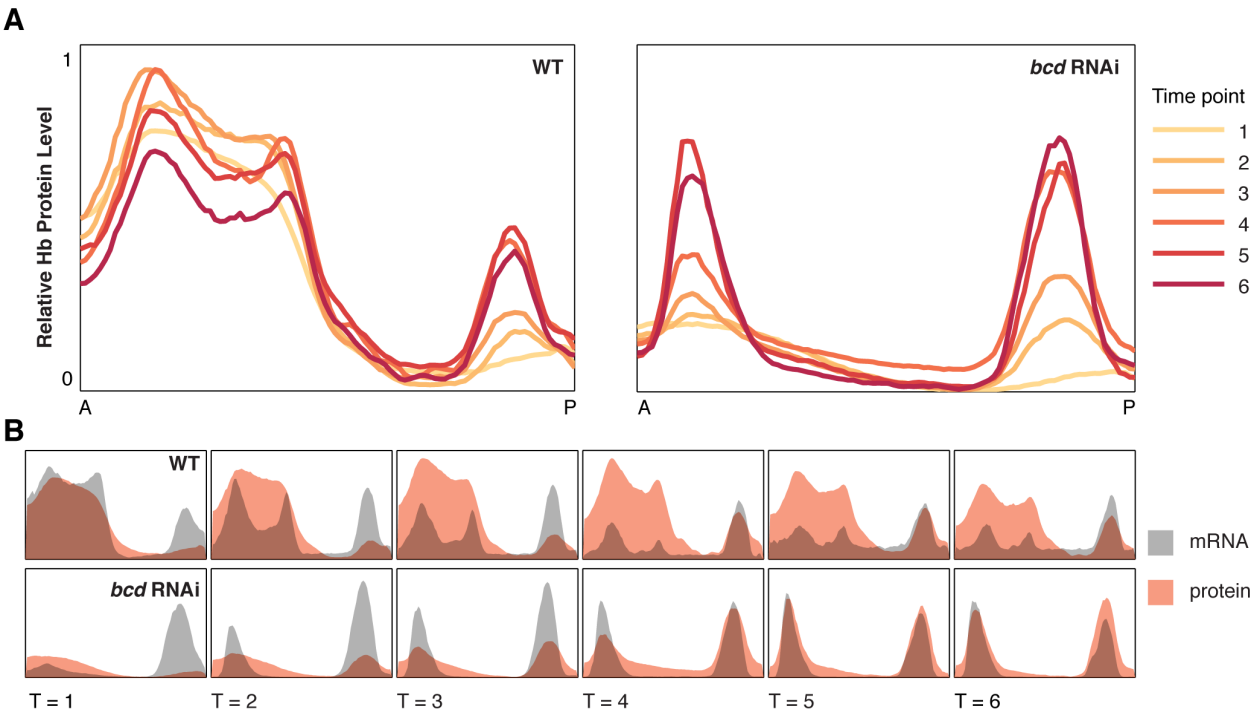
(A) The severity of *bcd* phenotype increases as the flies ages. We allowed *MTD-Gal4/ UAS-shRNA-bcd* flies to eclose for 48 hrs and counted denticle bands every few days. Old mothers were ≥11 days old and embryos were collected for the gene expression

Appendix B Figure S2 (Continued)
atlas. Counts: Day 2 n = 137, Day 6 n = 173, Day 8 n = 224, Day 11 n = 86, Day 14 n = 127, Day 16 n = 58. (B) qPCR indicates >90% knockdown in embryos laid by older mothers. Embryos were collected for 2 hrs. As a reference for WT *bcd* levels we averaged deltaCT counts from Days 3, 5, 8 and 12 of *MTD-Gal4/UAS-shRNA-GFP* embryos. We then subtracted this average number from the *bcd* RNAi ΔCT counts for Days 3, 5, 8, 10, and 14, and converted ΔΔCT to percent of WT levels.
(C) Paternal genotype does not meaningfully influence the *bcd* RNAi phenotype. *MTD-Gal4/UAS-shRNA-bcd* virgin females were crossed to males homozygous for *UAS-shRNA-bcd* or with males homozygous for a enhancer lacZ reporter (WT). Progeny from the first cross will have 1-2 copies of the *UAS-shRNA-bcd* construct (blue), while progeny from the second cross will have 0-1 copies of the *UAS-shRNA-bcd* construct (red). We could not detect a difference between the number of ventral denticle bands visible in each populations (p values from Kolmogorov-Smirnov test). Note that the effect of mother age is much greater than paternal genotype (compare Day 1 and Day 6 samples). Although it is possible that shRNAs against *bcd* are zygotically expressed, they do not meaningfully contribute to the phenotype, consistent with the purely maternal effect of *bcd.* These cross also showed there was no detectable paternal effect.
(D) Two *UAS-shRNA-bcd* lines yield similar phenotypes. We tested two shRNA lines (GL00407 and GL01320) at 25^{o}C and 29^{o}C, taking samples on Day 4 and Day 10. For these crosses we used the *mat-tub-Gal4* maternal driver; this driver tends to give a more consistent phenotype than *MTD-Gal4*. ANOVA of temperature, UAS-line, and day reveal that each has a significant but small effect, less than one segment in all cases. For the *mat-tub-Gal4* driver, the distribution of phenotypes laid by young mothers approaches the steady state distribution seen for old mothers with *MTD-Gal4* more quickly. This result is consistent with more uniform phenotypes for the *mat-tub-Gal4* driver with shRNAs against other genes (Staller et al., 2013). The distributions of embryos laid by old mothers of both genotypes are comparable. For future work depleting other maternal effect genes, we recommend the *mat-tub-Gal4* driver.

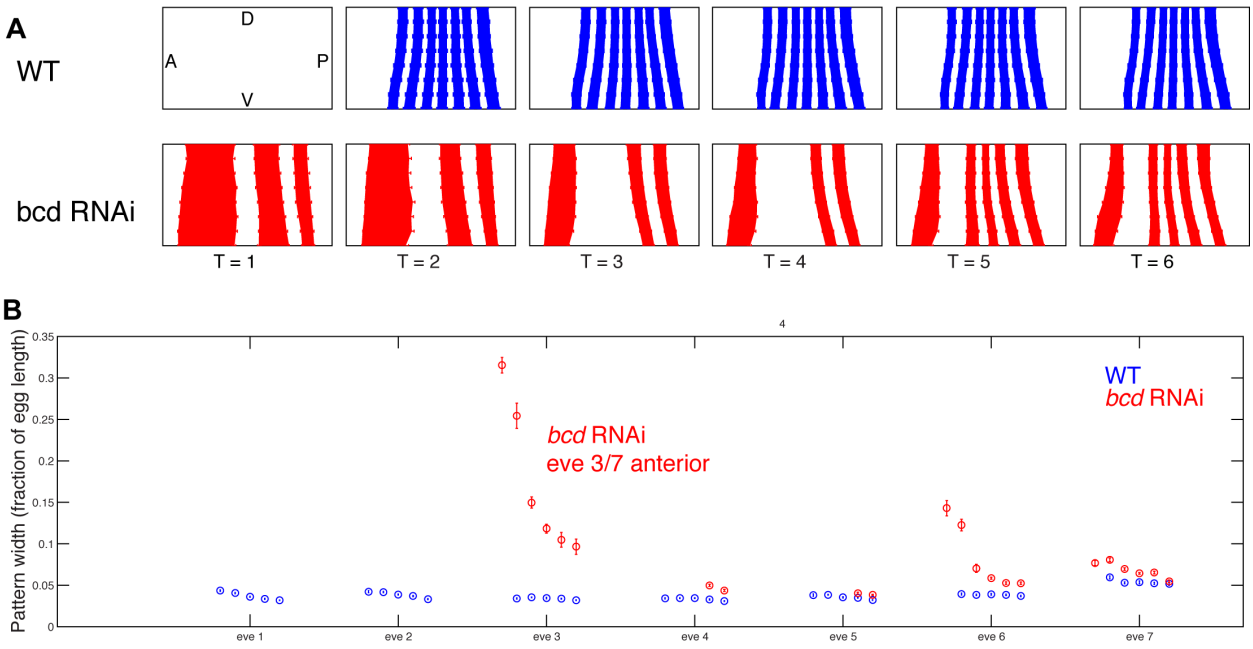


Appendix B Figure S3: *bcd* RNAi embryos have more cells and altered cell density patterns.

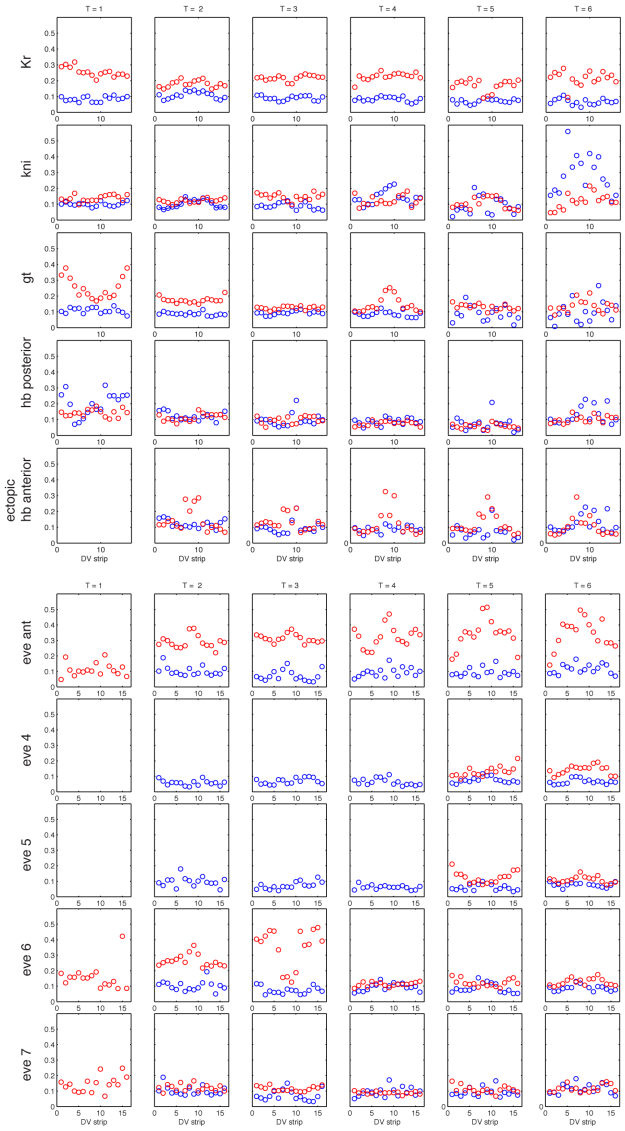
(A) Average cell density maps of WT and *bcd* RNAi embryos. While the physical shape of the embryos remains asymmetric, the posterior density pattern is duplicated in the anterior of *bcd* RNAi embryos, like some mRNA patterns. Embryos from stage 5:51-100% (time points 5 and 6 in the gene expression atlases) are shown. Note these images sometimes do not load in Preview and are best viewed in Adobe Acrobat.
(B) Histogram of cell counts in *bcd* RNAi embryos and WT (transgenic) embryos.



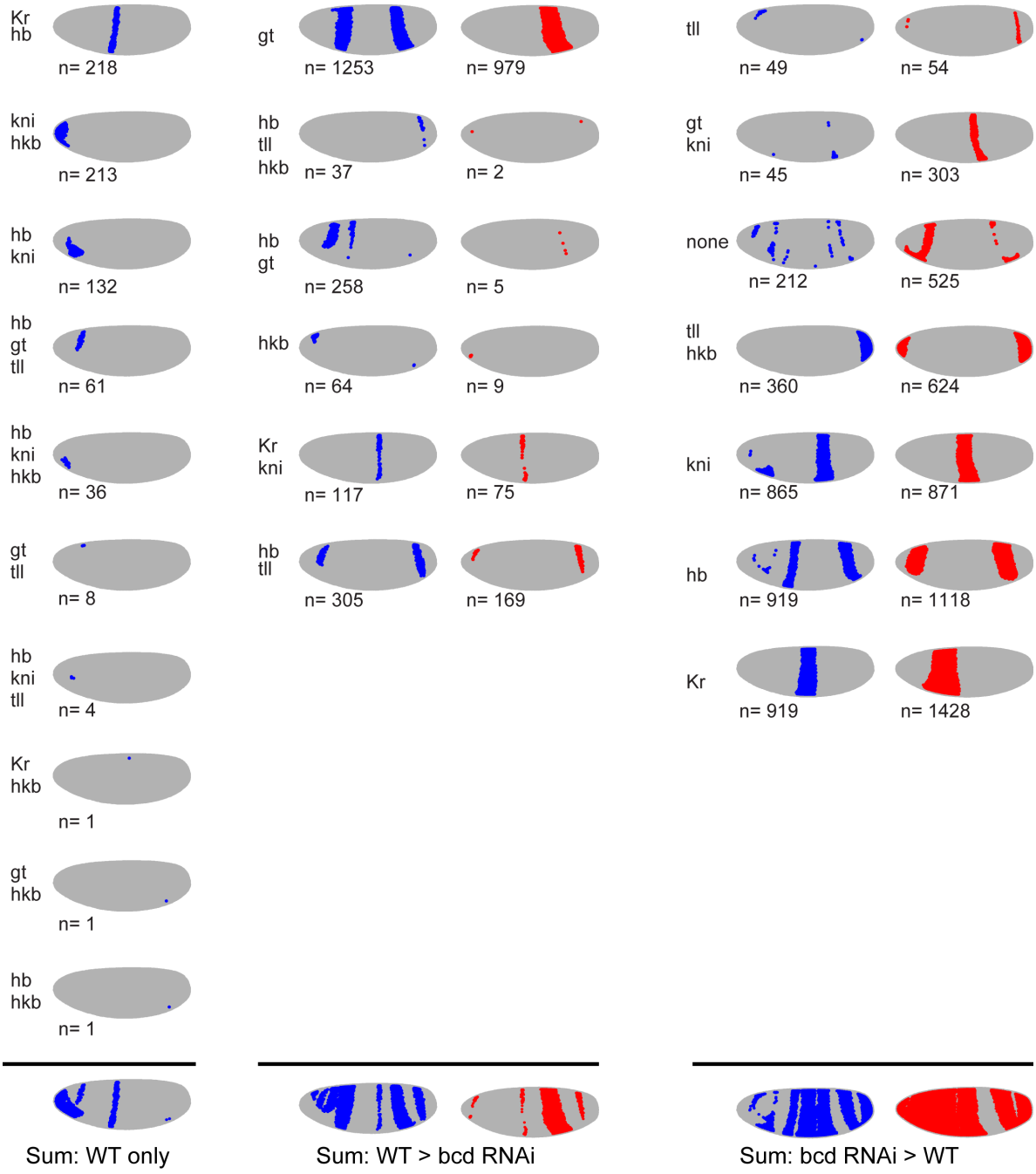
Appendix B Figure S4: The *bcd* RNAi gene expression atlas perturbs *hb* mRNA and protein levels.
(A) Hb protein expression pattern changes over stage 5 in both WT and *bcd* RNAi. In WT both maternal *hb* mRNA and *bcd* activated zygotic mRNA contribute to the anterior pattern, while in *bcd* RNAi, only maternal mRNA contributes to the early, broad anterior pattern (Tautz, 1988). Note each atlas is normalized separately, so absolute levels are not comparable between atlases. Relative levels change extensively.
(B) In both WT and *bcd* RNAi, *hb* mRNA (gray) and protein (red) patterns are different.



Appendix B Figure S5: The boundaries of the *eve* stripes move over stage 5 in *bcd* RNAi embryos.
(A) Boundary positions calculated using inflection points of individual embryos. Stripe 4 and 5 appear in a handful of embryos is cohort 4, but not frequently enough to reliably quantify boundary position. In each plot, anterior is left, dorsal is top.
(B) The widths of the *eve* stripes contract in *bcd* RNAi embryos. At T = 5 *eve* stripes 4-7 are approximately 1.7, .6, 1.4, and 1.3 cell widths wider in *bcd* RNAi than WT. At T = 6 eve stripes 4-7 are approximately 1.3, .6, 1.5, and .3 cell widths wider in *bcd* RNAi than WT. Data calculated from one DV strip along the left side of the embryo.



Appendix B Figure S6: The coefficient of variation of most the gap and *eve* stripe mRNA pattern widths are similar between WT (blue) and *bcd* RNAi (red). The exceptions are *Kr* and the anterior *eve* stripe, which are more variable in *bcd* RNAi embryos. The ventral region of the ectopic anterior hb pattern (DV strips 7-11) is very faint in *bcd* RNAi embryos, and our analysis script struggles to reliably find a boundary, so this analysis likely overstates the variability in this region. Pattern widths calculated with the inflection point, but using the half maximum led to very similar measurements. The ectopic anterior *hb* pattern in *bcd* RNAi is compared to the *hb* posterior pattern in WT.



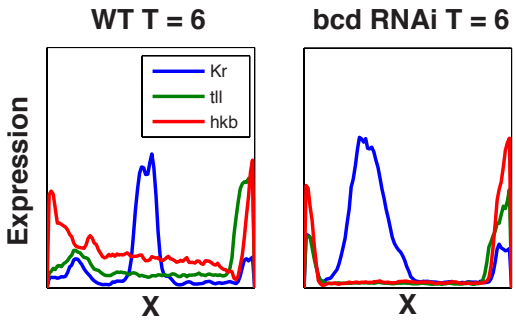
Appendix B Figure S7: The locations of each transcription factor combination at T=3 in WT (blue) and *bcd* RNAi (red).
The “sum” of each category is shown at the bottom.

A

Number of Gap Gene Combinations Unique to bcd RNAi (fine atlas)

Threshold	T = 1	T = 2	T = 3	T = 4	T = 5	T = 6
0.5	2	0	1	0	0	1
0.6	0	0	0	0	0	1
0.7	1	0	0	0	0	1
0.8	1	0	0	0	0	2
0.9	1	0	0	1	0	3
1	0	0	0	1	0	3
1.1	0	0	0	1	0	1
1.2	0	1	0	0	0	1
1.3	0	0	0	0	0	2
1.4	1	0	0	1	1	2
1.5	1	2	0	2	1	2

C



B **Identies and Counts of Gap Gene Combinations Unique to bcd RNAi—fine atlas**

	T = 1		T = 2		T = 3		T = 4		T = 5		T = 6	
Threshold	Genes	Count	Genes	Count	Genes	Count	Genes	Count	Genes	Count	Genes	Count
0.5	hb, gt, tll, hkb	13			none	90					hb, tll, hkb	5
	tll	10										
0.6											hb, tll, hkb	5
0.7	hkb	1									hb, tll, hkb	1
0.8	hkb	10									hb, Kr, tll, hkb	4
											hb, tll, hkb	2
0.9	hkb	9					hb, tll, hkb	10	hb, hkb	1	hb, tll, hkb	2
											hb, Kr, tll	1
											hb, Kr, tll, hkb	1
1							hb, tll, hkb	7			Kr, hkb	22
											hb, Kr, tll	1
											hb, Kr, tll, hkb	1
1.1							hb, tll, hkb	2			Kr, hkb	22
1.2			hb, hkb	1							Kr, hkb	34
1.3											Kr, kni	116
											Kr, hkb	39
1.4	hb, tll, hkb	216					gt, kni	98	Kr, kni	18	Kr, kni	90
											Kr, hkb	42
1.5	hb, tll, hkb	181	Kr, kni	13			gt, kni	60	Kr, kni	5	Kr, kni	65
			hb, tll, hkb	1			Kr, kni	16			Kr, hkb	41

Appendix B Figure S8: Changing the ON/OFF threshold does not meaningfully change the conclusions of the combination analysis.
(A) For a range of thresholds and most time points, all combinations present in bcd RNAi also present in WT. See Appendix B Figure S9B for schematic of how we vary thresholds. **(B)** When a combination was detected as unique to bcd RNAi, this was generally because it was not detected in WT for that threshold and time point. At high thresholds, overlap between adjacent patterns (Kr and kni or kni and gt) were not detected in WT. At T=4, the adjacent hb, hkb and tll patterns do not overlap enough to be detected in WT at most thresholds, but this combination was found at T=3, so it is not a true new combination. The new combinations found at T=6 arise either because the WT hkb data is low quality or because the anterior duplicated tll domain is smaller than the posterior domain.
(C) Line traces of Kr, hkb, and tll at T=6 in WT and bcd RNAi. Anterior-posterior position is on the x-axis and expression level is on the y-axis. The high levels of background in the WT hkb pattern may confound the combination analysis. The duplicated tll pattern in the anterior has weaker expression than in the posterior at T=6, which may explain the apparent emergence of the Kr, hkb combination. See also Appendix B Figure S10C.

Number of Gap Gene Combination Unique to bcd RNAi—hb protein

Threshold	T = 1	T = 2	T = 3	T = 4	T = 5	T = 6
0.5	2	0	0	2	1	3
0.6	2	0	0	1	0	4
0.7	2	0	0	1	0	4
0.8	2	0	0	0	1	4
0.9	2	0	0	0	2	3
1	1	0	0	0	1	1
1.1	0	0	0	0	1	1
1.2	1	1	1	1	2	3
1.3	2	1	1	0	2	5
1.4	1	1	1	3	3	4
1.5	1	2	1	4	3	4

Identies and Counts of Gap Gene Combinations Unique to bcd RNAi—hb protein

	T = 1		T = 2		T = 3		T = 4		T = 5		T = 6	
Threshold	Combo	Count	Combo	Count	Combo	Count	Combo	Count	Combo	Count	Combo	Count
0.5	gt, tll	51					Kr	878	Kr	922	Kr	932
	hbP, gt, tll, hkb	13					gt, tll	1			tll, hkb	45
											tll	1
0.6	gt, tll	88					gt, tll	1			Kr	1070
	hbP, Kr, kni	218									tll, hkb	87
											tll	3
											Kr, tll	1
0.7	gt, tll	69					gt, tll	1			Kr	1196
	hbP, Kr, kni	122									tll	8
											Kr, hkb	2
											Kr, tll	1
0.8	gt, tll	27							tll	2	Kr	1310
	hbP, Kr, kni	59									tll	14
											Kr, hkb	10
											Kr, tll	6
0.9	gt, tll	5							tll	10	tll	30
	hbP, Kr, kni	19							hkb	1	Kr, hkb	18
											Kr, tll	13
1	hbP, Kr, kni	2							hkb	2	Kr, hkb	22
1.1									hkb	4	Kr, hkb	22
1.2	hbP, tll, hkb	216	hbP, tll, hkb	377	hbP, tll, hkb	313	hkb	1	gt, kni	57	Kr, hkb	34
									hkb	6	gt, kni	29
											hkb	1
1.3	hbP, tll, hkb	211	hbP, tll, hkb	337	hbP, tll, hkb	239			gt, kni	33	Kr, kni	116
	hkb	2							hkb	7	Kr, hkb	39
											hkb	2
											gt, kni	13
											hbP, tll hkb	1
1.4	hbP, tll, hkb	197	hbP, tll, hkb	303	hbP, tll, hkb	186	gt, kni	98	gt, kni	15	Kr, kni	90
							hbP, tll, hkb	17	Kr, kni	18	Kr, hkb	42
									hkb	9	hkb	7
											gt, kni	6
1.5	hbP, tll, hkb	189	hbP, tll, hkb	269	hbP, tll, hkb	139	gt, kni	60	gt, kni	4	Kr, kni	65
			Kr, kni	13			Kr, kni	16	Kr, kni	5	Kr, hkb	41
							hbP, tll, hkb	6	hkb	15	hkb	15
							hkb	2			gt, kni	2

Appendix B Fig S9: Substituting Hb protein for *hb* mRNA does not meaningfully change the conclusions of the combination analysis

(A) For a range of thresholds and most time points, all combinations present in *bcd* RNAi also present in WT.

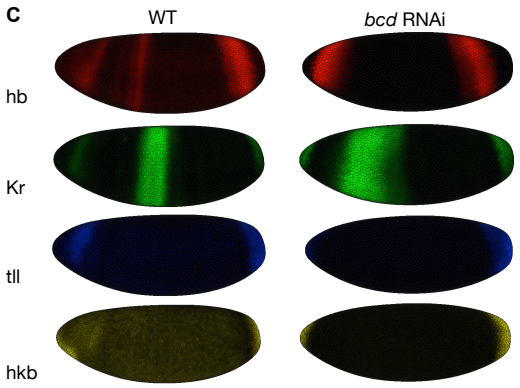
(B) Compare to Appendix B Figure S8B. In practice, the Hb protein (hbP) data is more difficult to partition into ON and OFF cells. Our method for finding ON cells is to make a histogram of the expression data, find the peak of the OFF cells, and add one s.d. For Hb protein, we add 0.5 s.d. instead. Accordingly, in this table the threshold of 0.9 (or 1.1) means we used 0.9 (or 1.1) s.d. for the 5 mRNAs and 0.45 (or 0.55) for Hb protein. When using the Hb protein data, the analysis is more sensitive to changes in threshold. For example, at T=3 for a threshold of 1.2 (0.6 for Hb protein), the posterior Hb protein domain in WT is no longer detected, leading to the false detection of the *hbP, tll, hkb* combination in *bcd* RNAi.

A

Number of Gap Gene Combinations Unique to bcd RNAi

Threshold	T = 1	T = 2	T = 3	T = 4	T = 5	T = 6
0.5	2	0	1	1	0	1
0.6	0	0	0	0	0	1
0.7	1	0	0	0	0	2
0.8	1	0	0	0	0	3
0.9	1	0	0	1	1	2
1	0	0	0	1	2	3
1.1	0	0	0	1	0	2
1.2	0	1	0	1	1	2
1.3	0	0	0	1	0	2
1.4	1	0	0	1	1	2
1.5	1	2	0	2	1	2

C

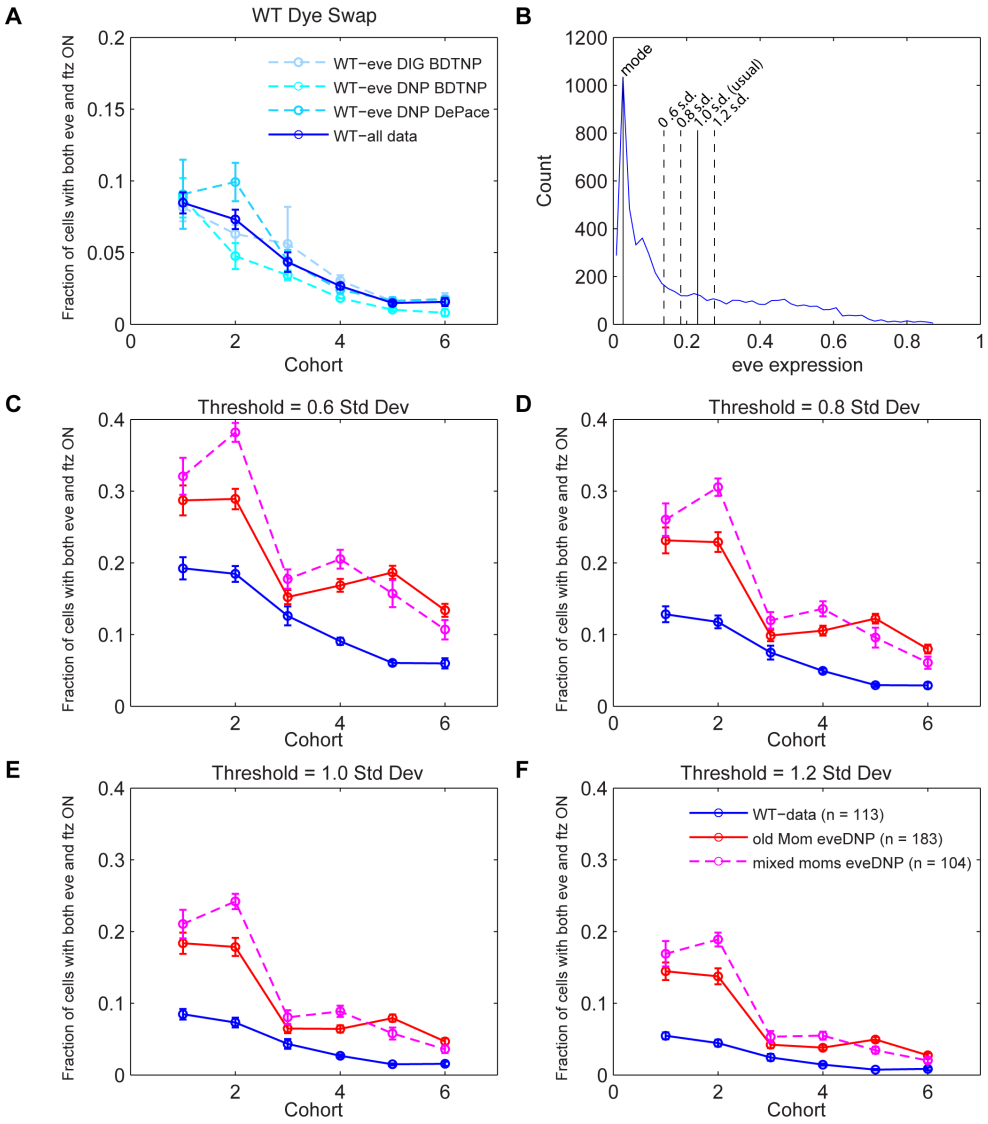


B **Identies and Counts of Gap Gene Combinations Unique to bcd RNAi course atlas**

	T = 1		T = 2		T = 3		T = 4		T = 5		T = 6	
Threshold	Genes	Count	Genes	Count	Genes	Count	Genes	Count	Genes	Count	Genes	Count
0.5	hb, gt, tll, hkb	13			none	90	**Kr, kni, tll**	**2**			hb, tll, hkb	20
	tll	10										
0.6											hb, tll, hkb	7
0.7	hkb	1									hb, tll, hkb	4
											hb, Kr, tll	**1**
0.8	hkb	10									hb, Kr, tll, hkb	3
											hb, tll, hkb	3
											hb, Kr, tll	**2**
0.9	hkb	9					hb, tll, hkb	23	hb, hkb	2	hb, tll, hkb	3
											hb, Kr, tll, hkb	1
1							hb, tll, hkb	14	**Kr, tll, hkb**	**2**	Kr, hkb	30
									hb, hkb	**1**	hb, Kr, tll	1
											hb, tll, hkb	**1**
1.1							hb, tll, hkb	9			Kr, hkb	33
											hb, tll, hkb	**1**
1.2			hb, hkb	1			**hb, tll, hkb**	**5**	**hb, tll, hkb**	**1**	Kr, hkb	41
											hb, tll, hkb	**1**
1.3							**hb, tll, hkb**	**1**			Kr, kni	85
											Kr, hkb	45
1.4	hb, tll, hkb	216					gt, kni	116	Kr, kni	20	Kr, kni	63
											Kr, hkb	47
1.5	hb, tll, hkb	181	Kr, kni	13			gt, kni	85	Kr, kni	3	Kr, kni	37
			hb, tll, hkb	1			Kr, kni	6			Kr, hkb	48

Appendix B Figure S10: Repeating the combination analysis with a coarsely aligned atlas suggests the fine scale alignment using the fiduciary marker does not confound our conclusions.

(A) For a range of thresholds and most time points, all combinations present in *bcd* RNAi also present in WT. **(B)** Compare to Appendix B Figure S6B. No new combinations arose for T=1-3. As expected, the abundances of many combinations changed subtly. New combinations are in bold. Most occurred with the fine atlas at other times or thresholds. e.g. *Kr,tll,hkb* in T=5 can be found in WT at T=6. **(C)** The mRNA patterns of *hb, Kr, tll,* and *hkb* at T=6 in WT and *bcd* RNAi. The combinations that were detected as unique to bcd RNAi in this time point are at the boundaries of these pattern in the termini. The may reflect subtle changes in the dynamics of terminal expression patterns, but are more likely artifacts of the differences in absolute levels between genotypes (which are not captured because each atlas is normalized separately) or because the T=6 *hkb* data has high background in WT, which causes this analysis to call it as ON in fewer cells.



Appendix B Figure S11: Thresholds, stain hapten, and mother age do meaningfully influence the fraction of cells expressing both *eve* and *ftz*. **(A)** The hapten (DIG or DNP) of the *eve* mRNA probe does not bias this analysis. *ftz* is always in the other channel. **(B)** A schematic shows the effects of varying the ON/OFF threshold for either *eve* or *ftz*. To find the threshold we create a histogram of the expression level of each gene separately and identify the peak of the OFF cell population (mode). For our normal threshold, we add one standard deviation. We varied the threshold to be 0.6, 0.8, 1 and 1.2 standard deviations. **(C-F)** Regardless of the ON/OFF threshold used, the fraction of cells expressing both *eve* and *ftz* (double ON cells) decreases over time. The mixed mom *eve* DNP embryos (magenta) were collected from cages that had a wide range of mother ages, before we started collecting from aged cages. We included these data because they have a very high quality stain. These data indicate that mother ages does not have a meaningful effect on this analysis. In Figure 4.6 the red and magenta data are combined. In all plots, error bars are the s.e.m.

Appendix B Table S1: Sequences of the enhancer lacZ reporter constructs included in the bcd RNAi atlas.

Table S1: Enhancer list

Construct	Enhancer sequence	Reference	Primer sequence
eve1	AGGCCTAATCACTTCCCTGAAATGCATAATTGTGCCGCGGCTTTTGATACGCTCCTGGCGGAGAGGGAGATGAGGAAAGGATGCACGGGAACCGC AGCCAAGTGGCAGTCGAGATTGGCAAATCCGCCAGCGGACAATGCCCAGAGAATGGGCAACAAGTAGCGGCGAATTAGCAATCCTATCATGCTTTT ATGGCCGGCCAACTCTTGCCCGCGCATCTCAGTTCATCCGAAGCGGGACCAGGTCCAGGTTCAAGTCGAGGTCCAGTACCCCTGCTATCCCGTCA ACCCCTTTAGGGCGATAATCCTTCTAAATGTTTGCATTAATTTCGAGGCGTGGACGGATTAGGGCGTGCTGGCTGGGCGGAACCCGCAGCAGAAAC	Fujioka et al., 1999	AGGCCTAATCA CTTCCCTG
	CGCCGAGGACACTGCACCGACTGACCTGCAGCCTACAGATCTCTGATCTTCGATCTCTAATCCTTTCGCATTTGCAACTGACTTCTGCACTGGGTCC GCCCCTAATCCTTCCGCCGAGAAGGCGGCAGAGTCGCGAGGTACTGGCCCGGGGTAATGGGATTATCTGCGATTACCCCAGATGATCCGCAGAAA GTCAATCTGGTTCAGGGGCTAATTGTCAGCGAAGTCAACTAAATCCAATCCTTTCGCGCCCCCTTCTGTTTATTTGTTTGTTTTCGTTTGTTTGAGA ATTTCTGGCAATTAAGTTGCCCGTTTTGATGCGCGGGGGCGGGTGCATCAAATCCTTTCGGCATACCTGTCCTGCACAAATGCTGAATTCCGCATC CCATGGATACCCAGATATTCAGATATCCCAAGGC		GCCTTGGGATAT CTGAATATCTGG
eve46	AGGATCCCTGGGCTCTGGGCTCTGGACTATCCGCCGACCCTCCATATCCATGATTTACAATTCTCGTTTTTTTCGCGTTATTTTTTTAGGGGCTTTAA TGACCGTCGTAAAGCCGCAGGAGGACCAGGACCAGGACTCTGCTCACATTTCGCGCACTGATTCTAAAAAATGAAATCATTTTTTCTTGAATTTCAC GGCGCGCCTCGAGCAGGACTCTTTGTTCTCGGCCAGGCAATTGTCCTTTTTTGCGCTCAGCTCTCAGTTTTTTCGTCCAGCGGGCATTACCTACAC GGCGTTTTATGGCGGAGATGATATTCGCCTGGGATCGGTTCCGTTTTTTAGGCCATAAAAATTAGGGCGGCATAAAAAAACTGCATTGGAATTCTAGT	Fujioka et al. 1999	AGGATCCCTGG GCTCTG
	TCTAGTTTCAAGTTTTTAGGTTTCCAGGTTTCTGCCAGCCCGCCTAGATTCGCATTTCGCGGAATTCGGAAGCGGAACAGAATGCCAGAATGGTCAG AATCCTGGCTGACCTTGCCTTTTGGCCAGGGGCCGTAAAAAAATTGACTCGCTGCGGTGCGCGGAATATTTTTTAAATCTGACTTTCCAACAATCTC TGATCTGGGTTCGAATCGTAAAAAAAAAGCAGAACAAAAAGCGGGCATTTTCGTCGGCAAATGATCTGTTAATGGGCCGGGCTAAAAAACTAAGTCA CAAAGTCACAAGGTTGTCCGGTAAATTGACCCGGTTAAGAATGTCTGTCTGTACCGAGAAGGATGCAGGACATTCAGCACTTCAAAGCTCCCACCG CTCGAAGGATTCCCCCGAAGATTCAC		GTGAATCTTCGG GGGAATCC
eve5	ATATCCCAAGGCCGCAAAGTCAACAAGTCGGCAGCAAATTTCCCTTTGTCCGGCGATGTGTTTTTTTTTTAGCCATAACTCGCTGCATTGTTTGGGC CAAGTTTTTCTTCTGCCAAATTGCGGAGATGATGCGGGGATTATGCGCTGATTGCGTGCAATTATGGACATCCTGCGAGGCCCCGAGGAACTTCCT GCTAAATCCTTTCATCCGCCTACAGAACCCCTTTGTGTCCCGTTCGCCGGGAGTCCTTGACGGGTCCTTCGACTATTCGCTTACAGCAGCTTGCGT	Fujioka et al. 1999	ATATCCCAAGG CCGCAAAG
	AAAATTTCATAACCCTACGAGCGGCTCTTCCGCGGAATCCCTGGCATTATCCTTTTTACCTCTTGCCAATCCGTTGGCTAAAAAACGGCTTCGACTTC CGCGTAACTGCTGGACAACAAAGACAAAAAACGGCGAAAGGACGGCGATTTCCAGGTAGCATTGCGAATTCCGTCAAACTAAAGGACCGGTTATAT AACGGGTTTATATGGCCAGAATCTCTGCATCTCCACGACCGCCAGAAGCTGCGTAAAACTGCAGGCTCTGTTTTGATTTCTGCAACTTCAGTTAATT GCCCGGGATGGCCAGCAATTGCCGGCAATTATAAAACAGCGCAGATGTGACTCAGCTTCCATATCTAACTCTATATCTCATGCCGAAAATCTAGGGT GGGGAGCGGAGGGGCGGGGTGCGTGGGTGACTTGCCTGCCAGGGAAAGGGGGCGGGGGTTCAGCGGGTGATAAATGTGCGTGATTTGGAATGA ATGCGCATCGATTAAAACCGCAGGGCAATCAATTT		AAATTGATTGCC CTGCGGT
eve3+7	GGATCCTCGAAATCGAGAGCGACCTCGCTGCATTAGAAAACTAGATCAGTTTTTTGTTTTGGCCGACCGATTTTTGTGCCCGGTGCTCTCTTTACGG TTTATGGCCGCGTTCCCATTTCCCAGCTTCTTTGTTCCGGGCTCAGAAATCTGTATGGAATTATGGTATATGCAGATTTTTATGGGTCCCGGCGATC CGGTTCGCGGAACGGGAGTGTCCTGCCGCGAGAGGTCCTCGCCGGCGATCCTTGTCGCCCGTATTAGGAAAGTAGATCACGTTTTTTGTTCCCAT	Small et al., 1996	GGATCCTCGAA ATCGAGAGC
	TGTGCGCTTTTTCGCTGCGCTAGTTTTTTCCCCGAACCCAGCGAACTGCTCTAATTTTTTAATTCTTCACGGCTTTTCATTGGGCTCCTGGAAAAA CGCGGACAAGGTTATAACGCTCTACTTACCTGCAATTGTGGCCATAACTCGCACTGCTCTCGTTTTTAAGATCCGTTTGTTTGTGTTTGTTTGTCCGC GATGGCATTCACGTTTTTACGAGCTC		GAGCTCGTAAA AACGTGAATGC
eve2	GGTTACCCGGTACTGCATAACAATGGAACCCGAACCGTAACTGGGACAGATCGAAAAGCTGGCCTGGTTTCTCGCTGTGTGTGCCGTGTTAATCCG TTTGCCATCAGCGAGATTATTAGTCAATTGCAGTTGCAGCGTTTCGCTTTCGTCCTCGTTTCACTTTCGAGTTAGACTTTATTGCAGCATCTTGAACA ATCGTCGCAGTTTGGTAACACGCTGTGCCATACTTTCATTTAGACGGAATCGAGGGACCCTGGACTATAATCGCACAACGAGACCGGGTTGCGAAG	Small et al., 1991	GGTTACCCGGT ACTGCATAAC
	TCAGGGCATTCCGCCGATCTAGCCATCGCCATCTTCTGCGGGCGTTTGTTTGTTTGTTTGCTGGGATTAGCCAAGGGCTTGACTTGGAATCCAATC CCGATCCCTAGCCCGATCCCAATCCCAATCCCTTGTCCTTTTCATTAGAAAGTCATAAAAACACATAATAATGATGTCGAAGGGATTAGGGG G		CCCCTAATCCC TTCGACATC
eve46mini	TCGAGCAGGACTCTTTGTTCTCGGCCAGGCAATTGTCCTTTTTTGCGCTCAGCTCTCAGTTTTTTCGTCCAGCGGGCATTACCTACACGGCGTTTTA TGGCGGAGATGATATTCGCCTGGGATCGGTTCCGTTTTTTAGGCCATAAAAATTAGGCGGCATAAAAAAACTGCATTGGAATTCTAGTTCTAGTTTCA AGTTTTTAGGTTTCCAGGTTTCTGCCAGCCCGCCTAGATTCGCATTTCGCGGAATTCGGAAGCGGAACAGAATGCCAGAATGGTCAGAATCCTGGC TGACCTTGCCTTTGGCCAGGGGCCGTAAAAAAATTGACTCGCTGCGGTGCGCGGAATATTTTTTAAATCTGACTTTCCAACAATCTCTGATCTGGG TT	Fujioka et al., 1999	TCGAGCAGGAC TCTTTGTTCTC
			AACCCAGATCA GAGATTGTTGG
hb posterior	TAGCACGAAAAACCGAAGGATTAAAAAAGGAAACTAGAGCAGAGGTCCCGGGGCAGGGCGAATAGTTGCTCTAATTTTCATTGTCCGCCTTAATGGTTA CGCCGTAAAATTGGCTATGCGGCCAAACAATAGTGCGAAGGACGACGGCAGGACGCGCAGGACAATCGTCTGGTGGATTTCCAGTCGACACGCCACG AGATTTTATGAAGGCAACTCGCTTTGCATGTTATTCCATAGATTTCGCTTCGGTCCCGGTTTGTTTTGGTCAGGTAAGACCTTCGATTAACAATGAAAGTA GCTGGAAAATCGCGAGAAACTTCGAAAGACACACAAAGATACAATATCTATGAGTCTAATGGTCATTAGAGCGGTGCGCTCTACATACAATTGTACCAGC CGTCTTGTTTGAAGCCTAAAAAACGTCGCAAAAAACACACTTCCGCGTAAGACATCCCATTTCTGTGGTCCGATCGTAAAATATTTAGTTTTTTATGACCA ACGGTGCGGGCAGGTAGCTGGCTGCCGTTTTTTGTGCGCGACCTCAACCCTTTCACCCATTAAGAAAAAATCGCATCCTGTGAGTGTCCTTGCCCGTTT CCCTCGAAACGGCCCACAATTTGTGTGCTTTGCGTTTTCTCCTCTCTTTTTGTTTCCACCTAATGTCGGCGTCATTGTCTTCTTTATGACGCCTCGGTTGT TTCTTTTTTATGGTGTCCTTTGTCCTTTGAGCCTCGTTGCACGGCCAAATCCCTACTTCCTCAACCCTTTGGCGGACGAGAAAGTTGCTAGGAGGAGAAC GGGTTAAGCGAAAACTCCATTGCACTTTTTACAAGCCGCGATCTTCTTGGAATTAGTTTTGGTCATTAGGCGAAAGGGTTAATTTCGATTTTGGCTCTCGG TGGGTTTACTGAGTGAATTCAATGGGCTAAGGCGAGTAAAGGGTTATACTGTTTTTACATTTTACTACTTGGAAAATACTGAAGAACTTGTAGGAAAAATTT CCAGCACTTTTAAAAGCCATATATAACTTTATGAATATGAACTTCAAATGTAAAAACCTGAAAGTGACATGTAGTTATTTTAAGGTCCTTGAAAATGATCATC GTCTAAAATTTCTTTTTTTTAAATAATTTTTAAAATATTTTTTTGATAGCATACGAAGTATTTAAAAATGTGAACAGATTAAACACATTAAATTTATAAAAGTAAA TACAACAGATTTAGCATAGAAATAAAAATCATTTTAATGTTCCGTCCATAAGTAACGGTCGTGGAAAATTCTTGAAAATCCCACAAATTATATTCGATCCCT TTGGCCGAACATTTGGTGCGATTACATTCGTAATTCGCTGGAAATTAAGGCCACTAAGTCGCCAGCGAAATGAATTCGGACATTGGGCATTGGACAAATG TAAAAAGGACTCTAGAGCCCCGACCATTGCAATGGTCCATTGTTGAGCGTCCGAAAGATCTGAAAACCAAACCCAAACCAAATCCCGAGCTTAGGCAAT CGGCATTGGGAAATAAGCGCCAAATATTCTACCCCCCACTCCAAAAACGAGCATT	Wunderlich et al., 2012	
gt posterior	AATTTATTACCAGAAACTTACCATCACTTCGAGATGAAAGTGCGGAGGAGAAGCGCTGATCTTTGCTTATCCCAAACTAGGGGTTTTTTAGGGGTAAA AAGCGGGAATACAGATTCTCAGAAACAAAATGGCAGTCCTAATATGTGATAAGTTGCTTTCAACTTACGTTTACGACAGGGGCCATTCGAAGCTAAG GATTCCAATATGCGACTGTTAACCCATTAAGACAAAGGGCCGCGAAAGGAGTTAACGGCAATTCCCCGAAACCAAGCACGAATCCAAATCCAAACC ACTCCTCACCCTTTTTTGGACGGGTGGGACGGGTCATATAAGGCAAAACCCCTTTCAGTCAATTAGTCGCTGATTTTCTTTGTCACCTAGCAGCGGA CCAATATAAAAATCGCAGCCACAATGGTCGGAGGAGAAGAACCCTTTTTTTTTTAAGGACCGCCGGTGTCCGAAATATCAGTTTATGGCTCCTTAAAA AACTGCAGCGGTTTTAGGGCCCGCGGACTCGGAATGAGGCCTTTTCGCACGAAGCGTCCATTATGCAATAAAACTTTGGATGTTTTTTAGGCAACA GCATAACAGCCTAATACAGCGCATACCGTGGGTGGGCAACCCGTTCGGCCATCAGGTAGATGATTCCGTAAAAACGCGCCCAAGGTCACAACTC AAAGGATTGCACAAATATTGCATCTTTGGTTCGGAGCTCATTATGGCGAAGGAACAGCAGTCCGAGTATATGAATATATGGTTTTATGGCCTGCTGT TTTTACGTTTTTTTTCTGCTGAATCTGGTTTTTACACCGAACTTGGCAACTCTTTGGCTTCGACTTCGGCCGCGACTTTTTATGGCATCTGCTCGGG ATCGATTGGGCTGGGCTATACGTATGTCCAGGATCTGGAAGGGCTCGGGTTCGGGCCCGGTAACCGCAGGTAAGGGATCGACTCAATGACGGCG ACGTGACGGTCCGAGTTTTAGTTTAATCCACCATTTTTACGGTTGGATTAGCGCACGGATTAGCGGATAAGTTCGCGGTTTTTTACTGATTACCATCG ATCGATCCCTCTTTTTATTTGCAATTCACCCAATCCCCCTGAATGGGCAGTTGTAAA	Wunderlich et al., 2013	
eve late seven stripe	-6.4kb (NdeI) to -4.8kb	Fujioka et al., 1996	-6.4kb (NdeI) to -4.8kb
BAC whole locus	Begins -6.4 kb upstream of *eve* transcription start site (TSS) and ends 11.3 kb downstream of *eve* TSS. The *eve* coding sequence has been replaced with *LacZ* and the neighboring *TER94* gene has been fused to GFP	Gift from M. Fujioka	

Appendix B Table S2: The numbers of individual embryos for each gene at each time point included in the bcd RNAi gene expression atlas

Table S2: Embryos per gene bcd RNAi atlas 20140807 v1.1

Genes	T=1	T=2	T=3	T=4	T=5	T=6	
D	9	19	26	13	5	13	
Kr	10	13	16	10	8	10	
hb posterior enhancer	6	8	20	9	6	3	
gt posterior enhancer	21	18	29	13	6	3	
eve3+7 enhancer	3	10	13	7	5	10	
eve4+6 enhancer	10	27	26	16	8	8	
eve5 enhancer	6	16	32	12	6	9	
eve4+6mini enhancer	3	8	12	11	0	3	
eveLocus lacZ	6	20	13	13	4	2	
cad	5	15	24	14	3	6	
eve	17	22	54	34	32	23	
fkh	5	14	16	9	4	7	
ftz	38	77	94	30	19	32	
gt	25	30	31	13	8	9	
h	8	11	11	9	14	15	
hb posterior enhancer	16	12	13	17	10	12	
hkb	16	11	20	14	6	8	
kni	9	15	14	8	6	8	
run	8	20	22	16	9	5	
tll	15	20	20	11	6	6	
Hb protein	7	10	14	7	6	9	
Average	11.6	18.9	24.8	13.6	8.1	9.6	
Sum	243	396	520	286	171	201	1817

Appendix B Table S3: The standard deviations of each gene in the bcd RNAi gene expression atlas. Data from the WT atlas is included for comparison (Fowlkes et al., 2008).

Table S3: Atlas Standard Deviations

Gene Name	WT s.d.	bcd RNAi s.d. v1.0	bcd RNAi s.d. v1.1
cad	0.165	0.066	0.067
eve	0.129	0.121	0.143
fkh	0.068	0.027	0.029
ftz	0.131	0.169	0.175
gt	0.108	0.072	0.073
hb	0.134	0.059	0.059
hkb	0.106	0.037	0.037
kni	0.099	0.058	0.061
Kr	0.066	0.081	0.079
tll	0.063	0.037	0.036
D	0.108	0.101	0.103
h	0.167		0.163
run		0.151	0.154
hbProtein	0.132	0.107	0.108
hb posterior enhancer		0.059	0.063
gt posterior enhancer		0.077	0.078
eve3+7 enhancer		0.074	0.075
eve4+6 enhancer		0.087	0.089
eve5 enhancer		0.125	0.131
eve4+6mini enhancer		0.074	0.077
eve locus reporter			0.104

Appendix B Table S4: ON/OFF thresholds used in the combination analysis in Figs 5, S8, S9.

Table S4A: WT thesholds

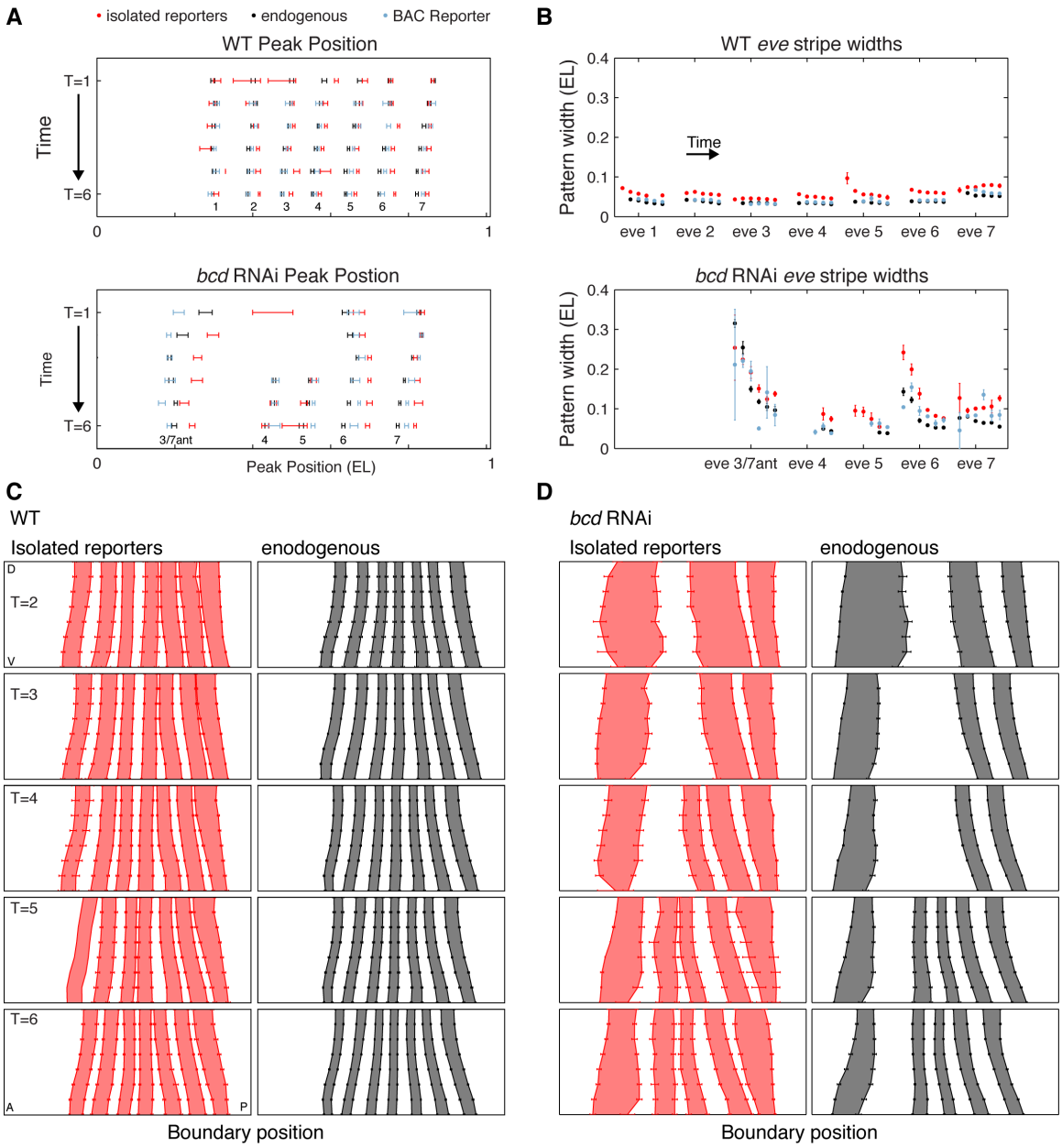
	T = 1	T = 2	T = 3	T = 4	T =5	T =6
gt	0.35	0.36	0.34	0.27	0.24	0.21
Kr	0.34	0.32	0.28	0.23	0.21	0.20
kni	0.32	0.30	0.31	0.34	0.35	0.32
tll	0.29	0.24	0.19	0.18	0.17	0.16
hkb	0.24	0.30	0.29	0.27	0.23	0.21
hb mRNA	0.41	0.28	0.23	0.23	0.28	0.25
hb Protein	0.21	0.20	0.22	0.22	0.22	0.16

Table S4B: bcd RNAi thesholds

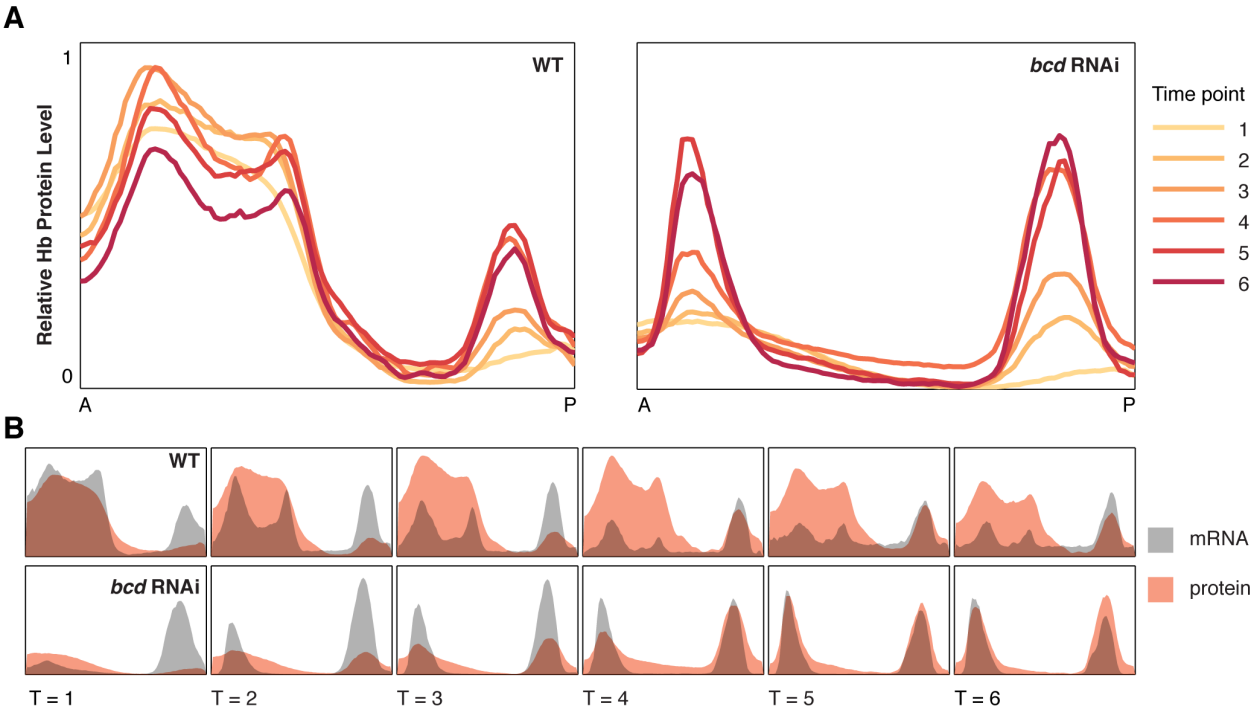
	T = 1	T = 2	T = 3	T = 4	T =5	T =6
gt	0.27	0.30	0.31	0.26	0.22	0.17
Kr	0.24	0.31	0.31	0.32	0.26	0.24
kni	0.20	0.23	0.27	0.29	0.25	0.20
tll	0.26	0.28	0.24	0.19	0.15	0.10
hkb	0.15	0.21	0.24	0.26	0.19	0.18
hb mRNA	0.21	0.26	0.28	0.25	0.22	0.23
hb Protein	0.08	0.13	0.14	0.24	0.27	0.35

Appendix C: Supplementary Materials for Chapter 4

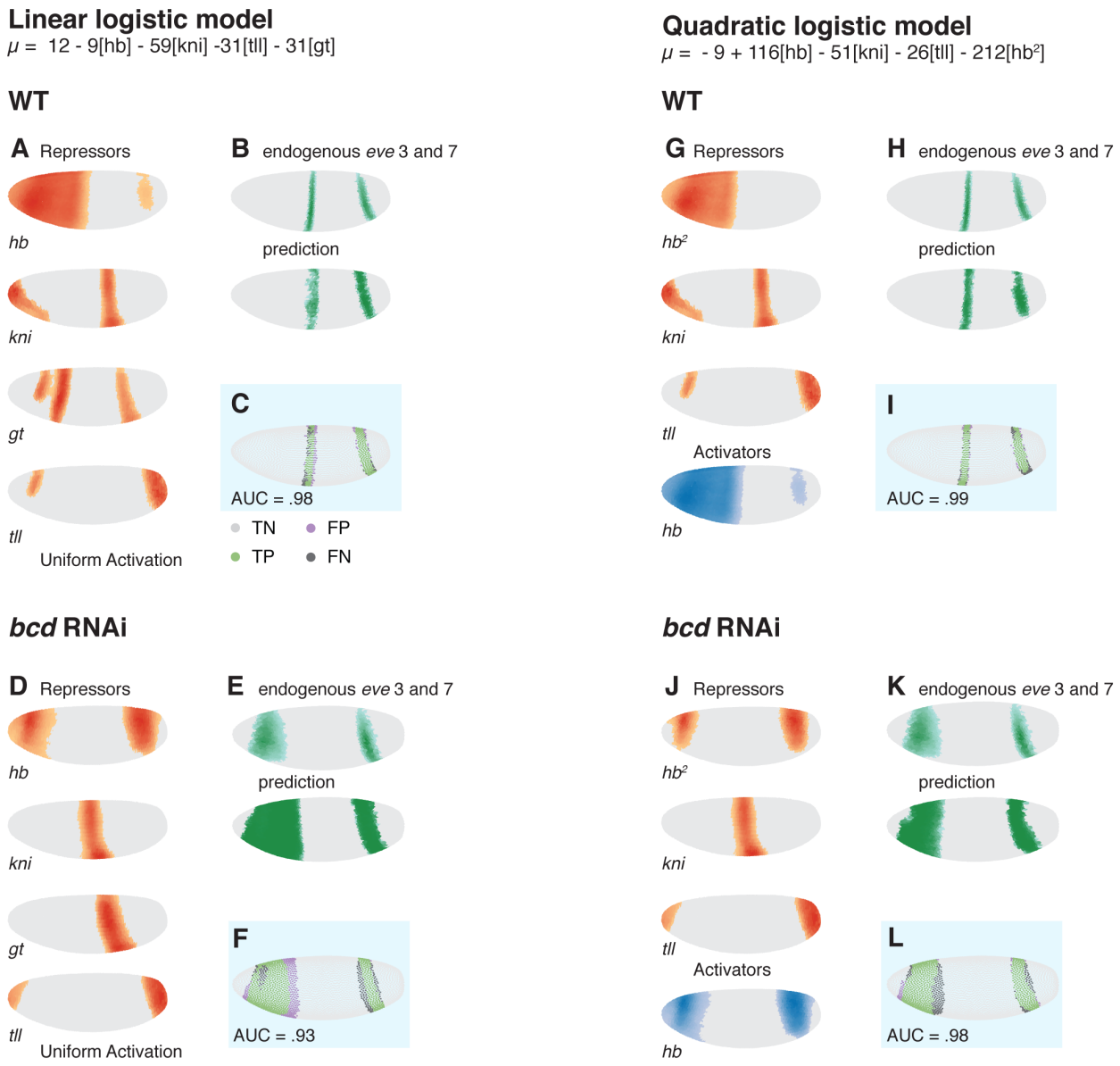
Appendix C Figure S1: The isolated *eve* enhancer *lacZ* reporters overlap the corresponding endogenous patterns with varying fidelity
Line traces of *lacZ* enhancer reporters (red) and the endogenous *eve* (gray) mRNA pattern in WT and *bcd* RNAi gene expression atlases. Anterior-posterior position (A-P) is plotted on the X-axis and expression level on the Y-axis for a lateral strip of the embryo.



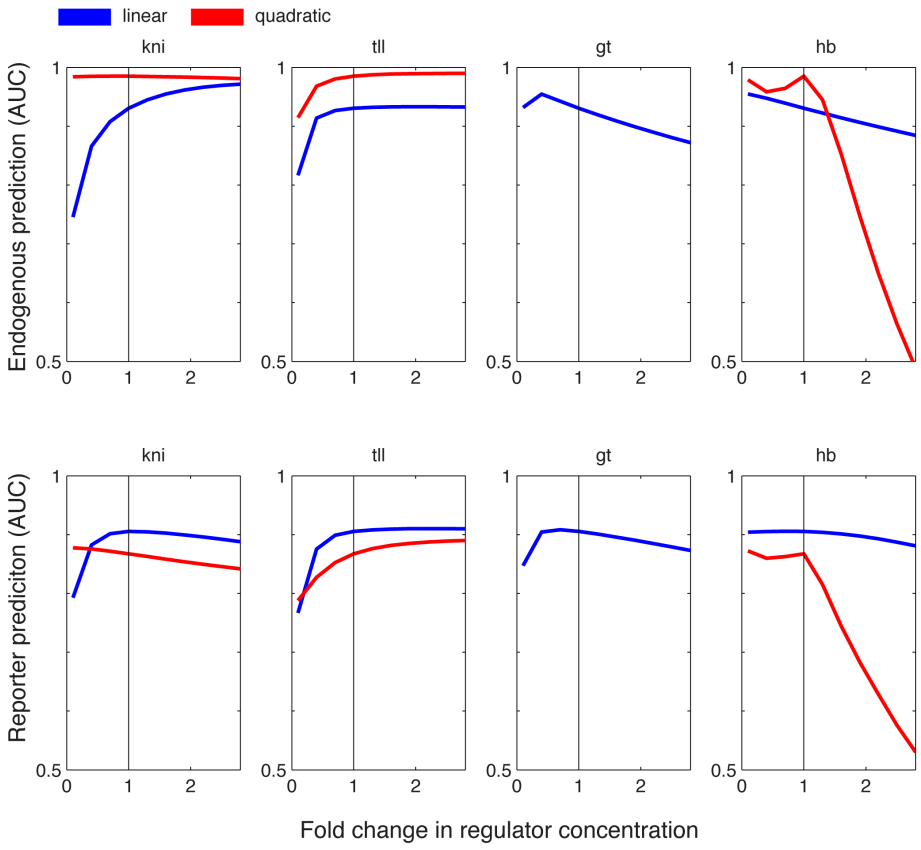
Appendix C Figure S2: The expression pattern driven by the whole locus BAC reporter is nearly identical to the endogenous pattern, while the patterns driven by isolated reporters differ
(A) The reporter peak positions (red) are slightly posterior to the endogenous *eve* peaks (black) and BAC reporter peaks (blue). Peak positions in WT and *bcd* RNAi are calculated from the lateral line traces shown in Figure 2–figure supplement 1. The anterior *eve3+7* pattern is faint and broad at T=1 and the peak is close to the middle of the embryo as seen in the lateral line trace in Figure 2–figure supplement 1. **(B)** Stripe driven by the isolated reporters (red) are wider than endogenous stripes (black) and BAC reporter (blue) in WT and *bcd* RNAi. Widths are calculated from the lateral line traces shown in Appendix C Figure S1. In WT, many of the error bars are smaller than the diameter of the point. **(C-D)** Boundary positions of the isolated reporters (dark blue) and endogenous stripes (light blue) in WT (C) and *bcd* RNAi (D). Note, the ventral most part of the *eve3+7* reporter anterior pattern is very faint in *bcd* RNAi embryos and the analysis software does not reliably detect the position of this boundary.



Appendix C Figure S3: The *bcd* RNAi gene expression atlas perturbs *hb* mRNA and protein levels.
(A) Hb protein expression pattern changes over stage 5 in both WT and *bcd* RNAi. In WT both maternal and *bcd* activated zygotic mRNA contribute to the anterior pattern, while in *bcd* RNAi, only maternal mRNA contributes. Note each atlas is normalized separately, so absolute levels are not comparable between atlases. Relative levels change extensively.
(B) In both WT and *bcd* RNAi, *hb* mRNA (gray) and protein (red) patterns are different.

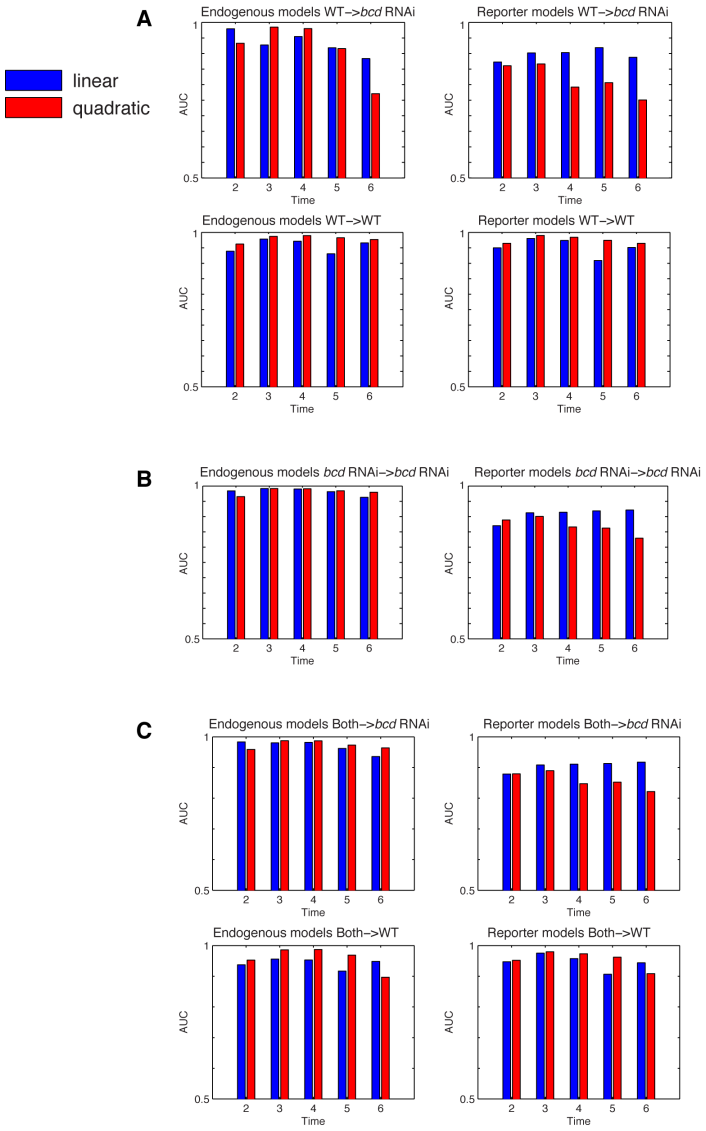


Appendix C Figure S4: Under perturbation of *bcd*, the expression patterns of endogenous *eve* stripes 3 and 7 are more accurately predicted by the quadratic model
(A) WT expression patterns of the regulators in the linear model. The expression level of each TF is shown for every cell. Cells with expression below an ON/OFF threshold (methods) are plotted in gray. For cells above this threshold, color intensity represents expression level. Repressors are red and activators are blue. **(B)** The expression pattern of the endogenous *eve* stripes 3 and 7 and the predictions of the linear model in WT. **(C)** Comparison of predictions to measurement in WT embryos. Green cells are true positives (TP), purple cells are false positives (FP), dark gray cells are false negatives (FN), and light gray cells are true negatives (TN). For visualization, the ON/OFF threshold is set to 80% sensitivity. The AUC metric quantifies performance over all thresholds. **(D)** The expression patterns of the regulators in the linear model in *bcd* RNAi embryos. **(E)** The expression pattern of the endogenous eve stripes 3 and 7 and the predictions of the linear model in *bcd* RNAi. **(F)** Comparison of linear model predictions to data in bcd RNAi. **(G-L)** Same as A-F, respectively, for the quadratic model.



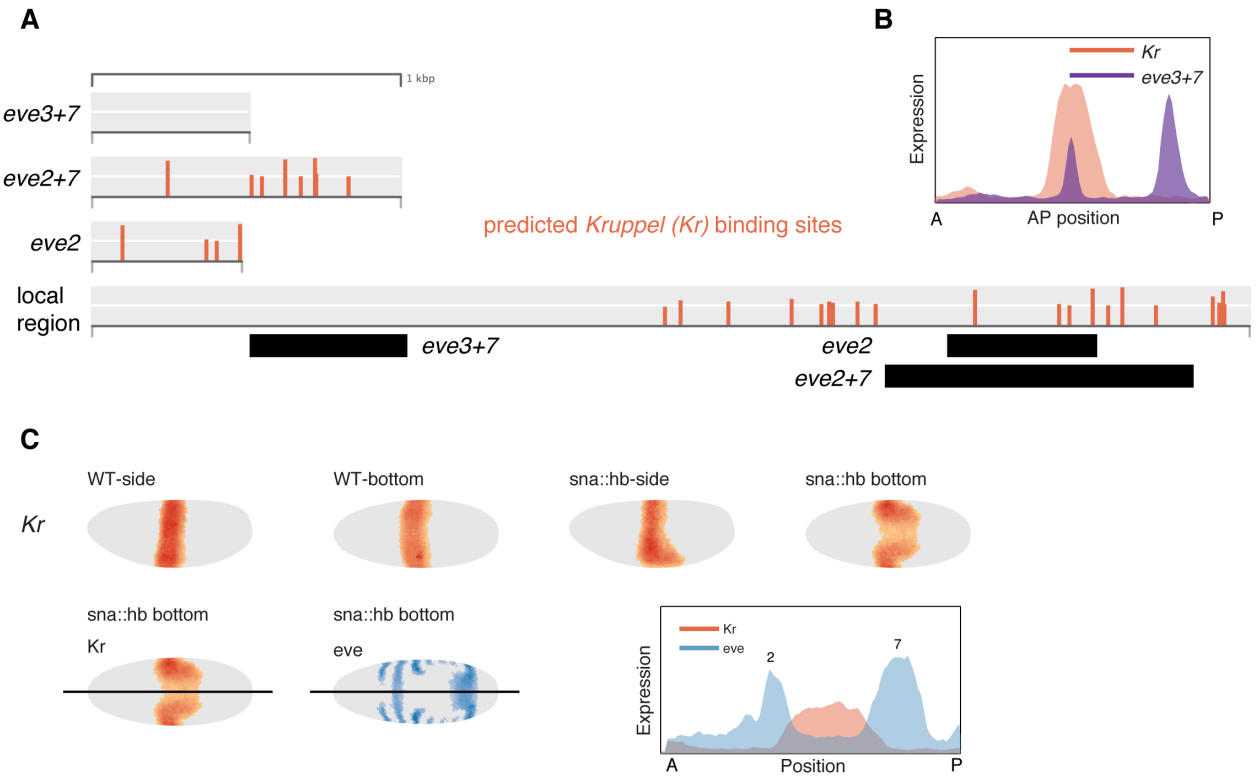
Appendix C Figure S5: Sensitivity analysis shows that scaling the relative level of a TF between atlases generally does not change the relative performance of the models

We varied the concentration of each TF separately in the *bcd* RNAi atlas and recalculated the AUC of the linear and quadratic models. This scaling simulates possible global changes in levels between genotypes. For the endogenous pattern, for all scalings of *kni* and *tll* , the quadratic model is more accurate than the linear model. For Hb, the quadratic model is more accurate than the linear model so long as maximal Hb levels in *bcd* RNAi are less than 1.38x maximal WT levels. Since *bcd* is a potent activator of Hb, Hb levels are very likely reduced in *bcd* RNAi embryos.
For the reporter pattern, all scalings of *hb* preserve relative model accuracy. The linear model is more accurate for a broad scaling of *kni* and *tll* levels.



Appendix C Figure S6: Fitting the linear and quadratic models on different datasets yielded similar results

(A) Fitting the models in WT at different time points and predicting the corresponding time points. The linear model always more accurately predicted the reporter *bcd* RNAi. Although both models are very accurate in WT, the quadratic model is more accurate. **(B)** Fitting the models in *bcd* RNAi and predicting *bcd* RNAi. The linear model more accurately predicted the reporter pattern. For the endogenous pattern, both models performed well. **(C)** Fitting the models on both the WT and *bcd* RNAi datasets led to similar results: the quadratic model more accurately predicted the endogenous pattern and the linear model more accurately predicted the reporter pattern in *bcd* RNAi.



Appendix C Figure S7: The expansion of the *Kr* expression pattern potentially explains the shape of the *eve2+7* expression pattern in *sna::hb* embryos
(A) The *eve2* enhancer is enriched for predicted *Kr* binding sites (red) while the *eve3+7* enhancer is depleted for *Kr* binding sites. We predicted binding sites using PATSER (stormo.wustl.edu) with a position weight matrix derived from bacterial 1-hybrid data (Noyes et al., 2008). **(B)** *Kr* expression overlaps stripe 3 of the *eve3+7* reporter mRNA in WT. *Kr* does not repress this pattern consistent with the absence of binding sites. **(C)** The distribution of *Kr* mRNA in WT and *sna::hb* misexpression embryos. The expanded ventral region of the *Kr* mRNA pattern appears to set the boundary of the expanded endogenous *eve* stripe 7 pattern.