Shadow Enhancers Provide Precision and Robustness for the Patterning of the Precellular Drosophila Embryo

By

Michael William Perry

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael Levine, Co-chair

Professor Nipam Patel, Co-chair

Professor Thomas Cline

Professor Leslea Hlusko

Fall 2011

Shadow Enhancers Provide Precision and Robustness for the Patterning of the Precellular Drosophila Embryo

Abstract

Shadow Enhancers Provide Precision and Robustness for the Patterning of the Precellular Drosophila Embryo

by

Michael William Perry

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Michael Levine, Co-chair

Professor Nipam Patel, Co-chair

Modification of *cis*-regulatory sequences is an important means by which developmental processes can evolve and generate changes in patterning and morphology. Understanding how these *cis*-regulatory enhancer sequences control gene expression is critical not only for understanding the mechanistic basis of development, but also for understanding how the systems are modified or evolve over time.

In this work I describe the discovery and characterization of many new enhancers that regulate genes for which a similar enhancer driving a similar expression pattern was previously known; these are "shadow" enhancers. The genes examined are critical developmental control genes responsible for the patterning and subdivision of the early *Drosophila* embryo. At first glance, many of these enhancers seem redundant, but are found to be evolutionarily conserved, suggesting a role in fitness and specific developmental function.

Sufficiently controlled tests of shadow enhancer function required the development and use of techniques such as bacterial artificial chromosome (BAC) recombineering and quantitative confocal imaging. Recombineering allowed the precise modification of large genomic regions and an unprecedented level of control in evaluating enhancer function when combined with the use of reporters and genetic rescue experiments. The ability to accurately quantify gene expression was aided by the development of a reporter gene containing an intron, allowing clear visualization of sites of

active transcription. These tools facilitated the evaluation of shadow enhancer function.

Evidence is presented which shows that in some cases one enhancer of the shadow enhancer pair can be removed without disrupting core function of the associated gene under normal developmental conditions. The system breaks down, however, under stress such as elevated temperatures during development, but only when one enhanceFr is missing. These results suggest that shadow enhancers represent a novel mechanism for the canalization of gene expression in varying environmental conditions and genetic backgrounds.

It is shown that in several cases shadow enhancers help ensure complete patterns of transcriptional activation across domains of gene expression. We present a mechanistic model in which increasing the number of enhancers for a given pattern helps ensure proper transcriptional state even allowing for stochastic interruptions of underlying molecular interactions such as enhancer/promoter looping or activator/repressor binding. It is suggested that shadow enhancers play a key role in ensuring robustness and reliability of gene expression patterns in development.

We propose a model whereby cryptic duplication events lead to the birth of shadow enhancers. Such enhancers can provide immediate value for fitness, for example, by ensuring robustness in response to environmental or genetic variations. The second enhancer might facilitate the evolution of novel patterns of gene expression, or become incorporated into the core developmental machinery to produce precise and rapid patterns of gene expression.

I dedicate this work to my parents, Sue and William Perry, for instilling in me a love of biology and providing the encouragement and advice necessary to make it through to the end.

Acknowledgements

First and foremost I thank my primary dissertation advisor, Mike Levine. I think everyone who comes through the Levine Lab experiences Mike in a slightly different way, but there are certainly some things in common. One of those is that he truly cares about helping students progress intellectually and in their approach to doing science. For me, that mostly meant learning to be more focused. Daily reminders and question(marks) were certainly one way to do that. "Where's the data?" will echo through my mind for some time to come. Frequent discussions about what is truly interesting helped keep me (mostly) on point. Another thing Mike cares a lot about a lot is producing a final product. I have truly come to appreciate the value of seeing things through to the end even when interests move on. I hope that lesson stays with me and helps me to do the same.

I thank my co-advisor, Nipam Patel. His easy accessibility, willingness to talk about interesting questions and problems, and hands on help and teaching style made him an invaluable resource. I learned much of what I know about practical developmental biology from him. I hope I become a fraction as good as Nipam at teaching these skills and interests to others. I am sure I will retain a strong interest in non-model systems, despite their difficulties.
I thank my committee members Tom Cline, especially for help with fly genetics, and Leslea Hlusko for her outside point of view (despite being my "inside" committee member!). I thank everyone in the Patel lab. In early days, I was helped especially by Ron Parchem, Paul Liu, Melinda Modrell, Mario Villa-vargas, Danielle Liubicich, Edward Jay Rehm, Cristina Grande, YaJen Ren and Francis Poulin. Crystal Chaw and other members of our evo-devo reading group were of great help in preparing for qualifying exams. Meredith Protas was a great source of kindness, positive encouragement, and ready advice then and throughout. I can't leave out Elaine Quan or Julia Serano, thank you Patel lab!

I can't say thank you enough to all of my fellow sufferers… I mean colleagues… in the Levine Lab. I thank Jacques Bothma for being an excellent partner in crime, and I hope we have the chance to work together again some day (How about tomorrow? I'm still waiting for that final analysis as soon as you get back from Bali!). Thanks to Jessica Cande for helping me get settled into the rhythm/grind of the Levine Lab and for great discussions and advice. Thanks to Alistair Boettiger for motivation and for many exciting discussions. We worked closely through several projects and without his particular skill

set and interests things would have turned out very differently. Thanks to all of the fly people, in particular: Vivek Chopra, Ben Haley, Dmitri Papatsenko, and Jess Piel for introducing me to the system. I thank Vivek for close help and collaboration in learning something of the right way to do fly crosses and for many discussions about genomics and ChIP, along with some excellent trips to Vik's Café. Thanks to the newcomers for keeping things lively, to Mounia Lagha, Valerie Hilgers, Emilia Esposito and Kasia Oktaba. To Valerie especially, being involved in the UTR extension project was a great side trip away from a focus on enhancers. To Mounia, I wish you many new phenotypes with your rescue experiments. To Emilia – thank you for bringing a bit of sunshine to the lab and to my everyday life. I thank especially Chihao (Kevin) Tsui, who spent a year or two as the Levine Lab tech/lab manager, a difficult and sometimes thankless job – I hope I was able to teach you something along the way, and your unfailing support and dedication helped me complete several projects. I thank Ryan Luu for being an excellent student and for his work and dedication at the bench. Thanks to the rest of the Levine Lab, the Ciona team - Eileen Wagner for putting up with all my molecular biology protocol questions and Alberto Stolfi for cloning advice and problem solving. Thanks to Phil Abitua, Wei-yang Shi, and for at least a time or two, Lionel Christiaen for sitting through my lab meetings and sharing the burden of keeping the boss occupied.

For outside help and support, I thank Russell Vance for initial recombineering advice and discussions. I thank Steve Small and Gary Strul for providing useful fly lines, and Christine Rushlow for some interesting discussions along the way.

I was supported by a National Science Foundation pre-doctoral fellowship, without which I wouldn't have been able to afford having my own apartment. The work was supported primarily by Mike Levine's NIH grants and the Center for Integrative Genomics.

Lastly, I thank my friends and family for supporting me in so many other ways. Thanks especially to my friends Phil Wolgin, Sam and Emily Redman, and Shaun Halper for always being there when I needed to escape – you guys have been true friends. Finally, I thank my parents, for always being supportive of my decision to go to graduate school despite understanding the difficulties and drawbacks of academic life. Thank you all.

Table of Contents

List of Figures

Chapter 1:

1.1 Dissertation Overview

Genes are turned on and off in the proper time and place during development through the action of *cis*-regulatory enhancers. Until recently, in nearly all documented cases each individual component of a pattern for a given gene was thought to be driven by a single and often separable enhancer. For a single pattern, there was a single enhancer. In this dissertation I present a growing body of evidence from my work and others that many developmental control genes in *Drosophila* have a second (and sometimes third) enhancer that drives expression which significantly overlaps the first in time and space. In many cases this expression is nearly indistinguishable from that driven by the other often previously characterized enhancer. This phenomenon was first noted in specific detail by Hong *et al.* 2008a, where these seemingly redundant enhancers were given the name "shadow" enhancers. A few similar cases have been noted in other systems but the specific function of having an additional enhancer had not been evaluated experimentally (e.g. Jeong *et al.* 2006; Werner *et al.* 2007; O'Meara *et al.* 2009; reviewed in Hobert 2010; Barolo 2011)

Despite their initial apparent redundancy, these enhancers are often evolutionarily conserved. In this chapter I introduce one of the two main developmental systems discussed – dorsal-ventral (DV) patterning in the *Drosophila* embryo – and review details of several recent case studies to provide examples of how enhancers can be modified to produce novel expression patterns. I also review evidence that in many cases enhancer position and function is deeply conserved. These sections draw heavily from my review published in a Cold Spring Harbor Symposium volume, Perry *et al.* 2009. I then describe my own efforts to trace the distribution of several pairs of "primary" and "shadow" enhancers across the *Drosophila* and look for clues to their origin. Despite these efforts we have yet to find more than a hint of how they may have arisen. Not all pairs of enhancers which seem nearly indistinguishable in *D. melanogaster* are as similar in more distantly related species of *Drosophila*, though in each case an enhancer still exists in the same relative genomic position. This high degree of conservation implies a role in fitness and specific function.

In the second chapter I describe experimental tests of shadow enhancer function. These experiments focus on *snail*, a gene critical for proper specification and patterning of the mesoderm. We present evidence that this gene is regulated by a distal shadow enhancer located in the intron of a neighboring gene. We test the function of this enhancer using a combination of bacterial artificial chromosome (BAC) recombineering to precisely replace one enhancer or the other, followed by either 1) quantitative confocal imaging and an analysis of the amount of gene activation across the *snail* domain in many embryos, or 2) genetic rescue experiments. Removal of the proximal primary enhancer does not significantly perturb *snail* function, including the repression of neurogenic genes and formation of the ventral furrow during gastrulation at normal temperatures. However, at elevated temperatures, there is sporadic loss of *snail* expression and coincident disruptions in gastrulation. Similar defects are observed at normal temperatures upon reductions in the levels of Dorsal, a key activator of *snail* expression. These results suggest that shadow enhancers represent a novel mechanism for the canalization of gene expression, ensuring a reproducible, reliable outcome despite extrinsic and intrinsic variability. Embryos in at least some conditions develop normally even when one enhancer is missing. It is therefore possible that selection on a subset of embryos from a population which experiences varying thermal environments and genetic backgrounds is responsible for the maintenance and even origin of shadow enhancers. This work was published in *Current Biology* in 2010 (Perry *et al.* 2010).

Shadow enhancers have so far been found for some 20-30% of DV patterning genes in *Drosophila*. In chapter three I provide evidence that shadows occur for perhaps every member of a second patterning system – the "gap gene" network that patterns the anterior-posterior (AP) axis and are critical for proper segmentation of the *Drosophila* embryo. I characterize a set of new enhancers and retest a few previously described (though mostly overlooked) enhancers. The understandable tradition in the field has been to dissect the first enhancer found for clues to regulatory logic, and not to look further once "the" enhancer for a pattern was indentified, which is probably why so many "shadows" were overlooked. It was only with the advent of techniques that provide genome-wide surveys of transcription factor binding such as chromatin immunoprecipitation-chip (ChIP) and more recently ChIP–seq that these extra enhancers came to light.

During a characterization of this set of enhancers it became clear that all shadow enhancers are not entirely equal. In all cases, by definition, each pair of enhancers drives significantly overlapping expression in time and space. In some cases, however, there are key differences in the patterns driven by enhancers in isolation. These cases involve "dominant" or "long-range" repressors, which are present at one enhancer but which are not always present at the other. The two enhancers together provide a function similar to that described in chapter two for *snail* – they help provide reliable, robust expression across the domain where they overlap. Taken separately, in cases where one enhancer lacks binding of a dominant repressor, it may show ectopically broad expression for some portion of its pattern. We document several such examples and provide evidence for the role of long-range repression in the refinement of the gene *hunchback* at the anterior tip, and for setting proper limits especially to the posterior boundary of the gene *knirps*. These interactions, where two enhancers are required both for complete activation of an overall pattern and also for setting proper limits on gene expression are different from previously described types of enhancer interactions. The work presented in chapter three was published in Perry *et al.*, 2011. A follow-up section details further BAC experiments that even more conclusively support the previous results and conclusions.

A final chapter (four) explores further details of *hunchback* regulation and a slightly different mechanism for the canalization of gene expression. While in chapter three, I describe finding a shadow enhancer for *hunchback* and how the two enhancers together are necessary for complete patterns of activation and a reliable, robust response to the Bicoid gradient. In chapter four, it becomes clear that the final Hunchback boundary depends on not just these two enhancers, but a third enhancer as well. This third enhancer produces a stripe of expression adjacent to the position of the Hunchback boundary, yet is not dependent on direct Bicoid input. The "central stripe" enhancer was dissected through mutagenesis experiments and surprisingly seems to be activated everywhere in the embryo and shaped into specific stripes by the combinatorial action of several repressors. Also a surprise was that hunchback itself helps shape this pattern through repression, and not auto-activation as had been previously proposed. Finally, this pattern is shown to be critical for proper positioning and sharpness of the final hunchback boundary. Embryos lacking this enhancer have a hunchback protein gradient with a significantly different shape and increased variability of expression, and this change has an effect on downstream targets such as even-skipped (eve) and engrailed (en). The final pattern with its sharp

on/off boundary is dependent on not just a single enhancer – but on three different, independent enhancers. While this third enhancer is not a shadow enhancer *per se*, it involves the earlier input of a proximal/shadow pair, is only partially distinct in time and space, and is also involved in the canalization and reliable output of the *hunchback* expression pattern despite variations in input. This chapter begins with an introduction to DV patterning.

1.2 Chapter Summary

The dorsal-ventral (DV) patterning of the early Drosophila embryo depends on Dorsal, a maternal sequence-specific transcription factor related to mammalian NF-kB. Dorsal controls DV patterning through the differential regulation of ~50 target genes in a concentration-dependent manner. Whole-genome methods, including ChIP-chip and ChIP-seq assays have identified ~100 Dorsal target enhancers, and over a third of these have been experimentally confirmed via transgenic embryo assays. Despite differences in DV patterning among divergent insects, a number of the Dorsal target enhancers are located in conserved positions relative to the associated transcription units. Thus, the evolution of novel patterns of gene expression might depend on the modification of old enhancers rather than the invention of new ones. As many as half of all Dorsal target genes appear to contain "shadow" enhancers: a second enhancer that directs the same or similar expression pattern as the primary enhancer. Preliminary studies suggest that shadow enhancers might help ensure resilience of gene expression in response to environmental and genetic perturbations. Finally, these shadow enhancers are often conserved over reasonably large evolutionary distances, suggesting a functional role in development and a distinct impact on fitness.

1.3 Introduction to DV Patterning

Dorsal-ventral patterning of the *Drosophila* embryo is controlled by Dorsal, a sequence-specific transcription factor related to mammalian NF-kB (Rushlow *et al.* 1989; Roth *et al.* 1989; Ip *et al.* 1991). The Dorsal protein is distributed in a broad nuclear gradient, with peak levels present in ventral nuclei and progressively lower levels in lateral and dorsal regions (see representation in Fig. 1.1; Rushlow *et al.* 1989; Steward 1989; Roth *et al.* 1989). This Dorsal nuclear gradient initiates dorsal-ventral patterning by regulating 50-60

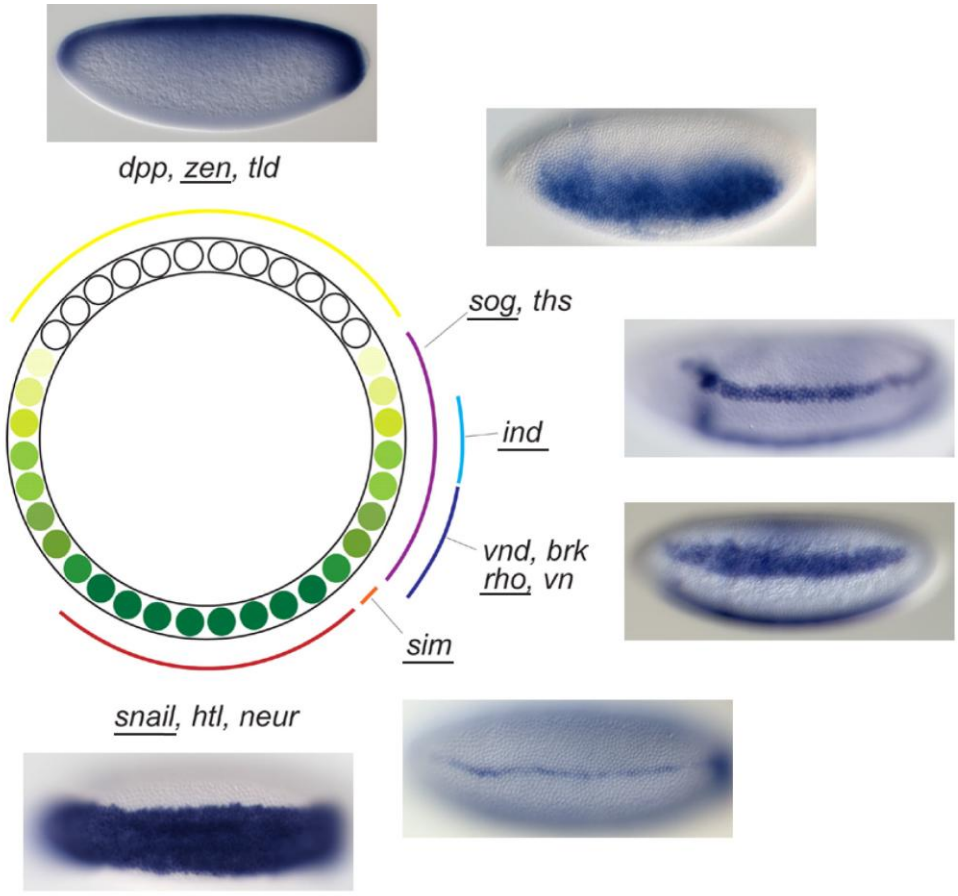


FIGURE 1.1 Schematic cross section of a Drosophila embryo and representative images. The central schematic is of a cross-section of a Drosophila embryo showing the six general categories of DV gene expression with a few example genes listed. The surrounding images are example expression patterns of the nearest underlined genes; lateral views are shown. (Schematic reproduced from Hong et al. 2008b).

target genes in a concentration-dependent fashion (Stathopoulous *et al.* 2002; Zeitlinger *et al.* 2007).

Whole-genome ChIP-chip assays (see below) identified ~100 potential Dorsal target enhancers, and over 30 of these have been directly tested in transgenic embryos (e.g., Zeitlinger *et al.* 2007, Hong *et al.* 2008a). Altogether, these enhancers direct 6 distinct patterns of gene expression across the dorsal-ventral (DV) axis of precellular embryos (Figure 1.1). Dorsal works in a highly combinatorial manner to generate these diverse patterns (reviewed by Hong *et al.* 2008b). For example, Dorsal and SuH, a transcriptional effector of Notch signaling, activate *single-minded* (*sim*) expression in a single line of cells (CNS ventral midline) on either side of the mesoderm (Cowden and Levine 2002; Morel *et al.* 2003). In contrast, Dorsal works together with a different sequence-specific transcription factor, Pointed (an effector of EGF signaling), to activate gene expression within lateral stripes in intermediate regions of the future ventral nerve cord (Gabay *et al.* 1996).

1.4 Enhancer evolution

In principle, substitutions of "coactivator" binding sites within Dorsal target enhancers can alter the DV limits of gene expression. For example, replacing SuH binding sites with Twist sites results in expanded expression of the modified enhancer within the presumptive neurogenic ectoderm (Gray and Levine 1996; Zinzen *et al.* 2006). Analysis of Dorsal target enhancers in divergent insects, including mosquitoes (*Anopheles gambiae*), flour beetles (*Tribolium castaneum*), and honeybees (*Apis mellifera*), suggests that such changes might occur during evolution to produce distinctive dorsal-ventral (DV) patterning mechanisms (Zinzen *et al.* 2006).

One such example is seen for the ventral midline of the honeybee, *Apis mellifera* (*A. mellifera*). In *Drosophila*, the ventral midline is just two cells in width and arises from two lines of sim-expressing cells that straddle the mesoderm prior to gastrulation (Fig. 1.2). In contrast, the ventral midline of the *A. mellifera* CNS is considerably wider, encompassing ~5-6 cells. An expanded ventral midline is also seen in flour beetles, *Tribolium castaneum* (*T. castaneum*), suggesting that the broad pattern is ancestral, and the narrow midline of *Drosophila* (and the mosquito, *Anopheles gambiae*) is a derived feature of the dipteran CNS (Zinzen *et al.* 2006).

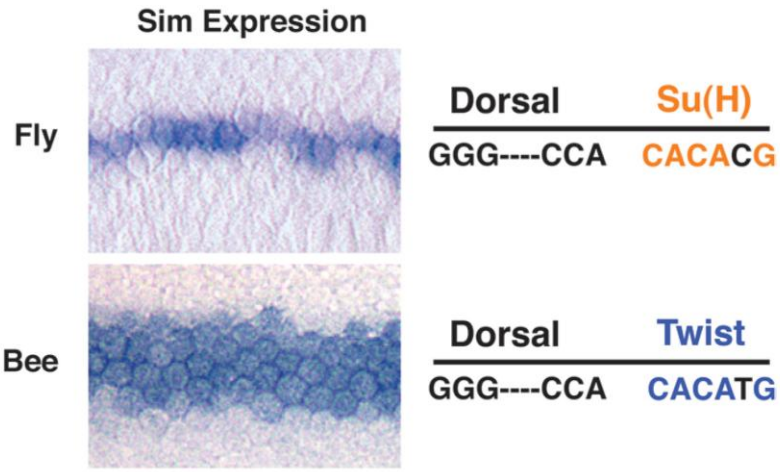


Figure 1.2 sim exhibits a broader pattern of expression in the honeybee CNS as compared with Drosophila. This expansion appears to result from the replacement of Twist binding sites with Suppressor of Hairless sites (Notch signaling) in the respective 5' sim enhancers. (Reproduced from Perry et al. 2009).

Expansion of the *sim* expression pattern is sufficient to account for the broad ventral midlines of the *A. mellifera* and *T. castaneum* CNS. In *Drosophila*, ectopic activation of sim expression using the *eve* stripe 2 enhancer results in the formation of an ectopic ventral midline throughout the neurogenic ectdoderm of transgenic embryos (Zinzen *et al.* 2006). The *sim* regulatory region contains two distinct enhancers. One mediates activation by Dorsal and Notch signaling (establishment enhancer), while the other mediates positive autofeedback through direct binding of the Sim transcription factor to the autoregulatory enhancer. Once Sim is misexpressed, the expanded pattern is maintained by autofeedback.

Sim establishment enhancers were identified in the 5' flanking regions of the *sim* loci in *Anopheles gambiae (A. gambiae)*, *T. castaneum*, and *A. mellifera*.

The *sim* enhancer from *A. gambiae* directs sharp lateral lines when expressed in transgenic *Drosophila* embryos. In contrast, the enhancers obtained from the *sim* loci of *T. castaneum* and *A. mellifera* produce broader expression patterns. The *A. gambiae* enhancer looks like the *Drosophila* enhancer in that it contains a series of Dorsal and SuH binding sites. However, the *T. castaneum* and *A. mellifera* enhancers contain Twist sites rather than SuH sites, and consequently, they direct broader patterns of gene expression (Zinzen *et al.* 2006).

1.5 Constancy of enhancer location

The *sim* enhancers of flies, mosquitoes, flour beetles, and bees lack simple sequence similarity. Despite this extensive sequence divergence, comparable enhancers are located in the same relative positions: in the immediate 5' flanking regions of the respective *sim* loci (e.g., Fig. 1.3).

Since this is a relatively common location for developmental enhancers, additional studies were done to determine whether enhancer locations are conserved for other critical DV patterning genes (Cande *et al.* 2009). These studies identified enhancers for 5 additional genes: *cactus*, *sog*, *twist*, *brinker* and *vnd*. *cactus* is a key component of the Toll signaling pathway that regulates Dorsal nuclear transport (Roth *et al.* 1991; Stein and Nüsslein-Volhard 1992). It is activated by high levels of the Dorsal gradient in the presumptive mesoderm of both *Drosophila* and *T. castaneum* embryos (Maxton-Kuchenmeister *et al.* 1999; Nunes da Fonseca *et al.* 2008). The enhancers that are responsible for these expression patterns are located in 3'

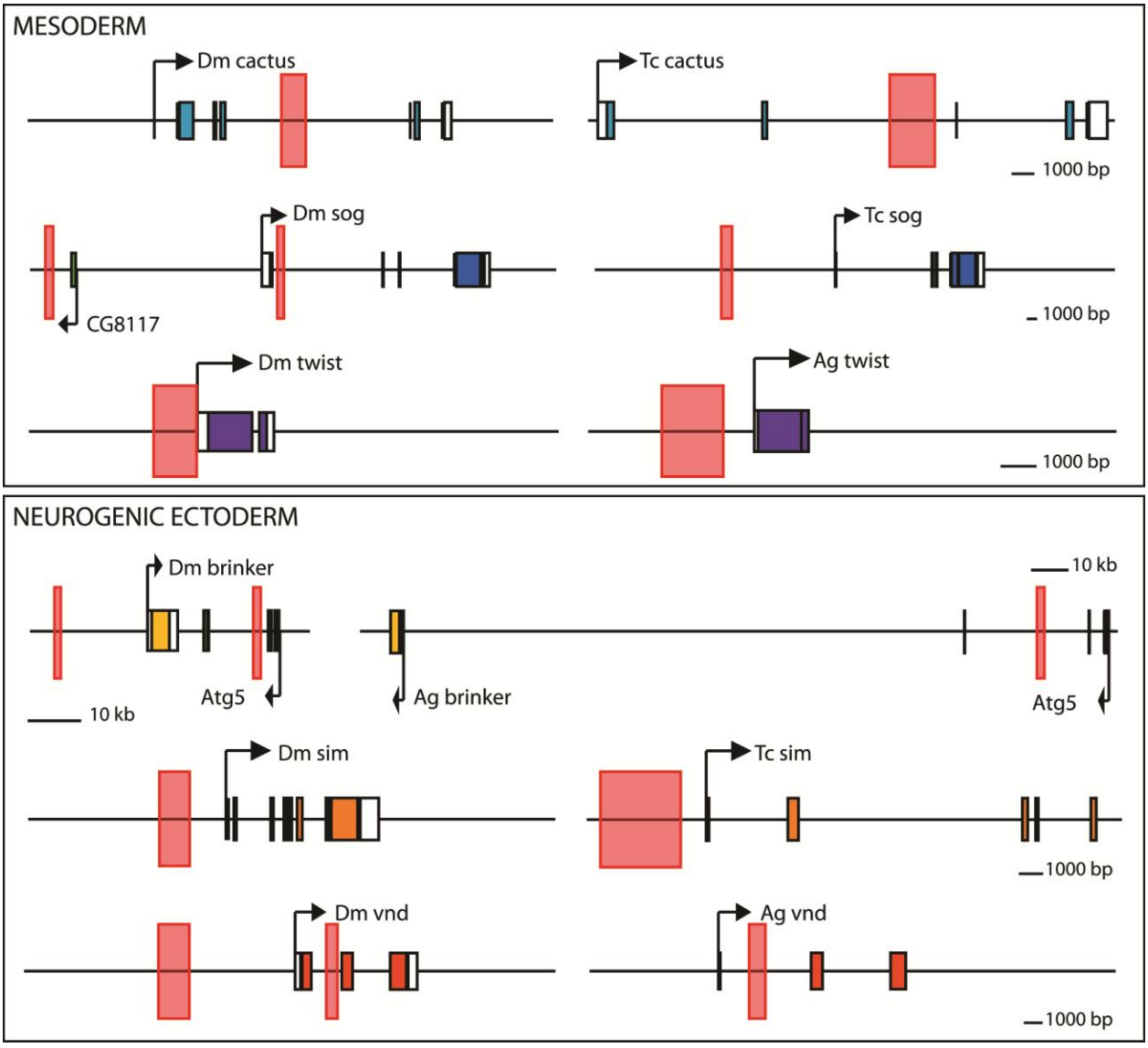


Figure 1.3 Conservation of enhancer location in divergent insects. Enhancers regulating the associated transcription units are indicated by pink boxes. Coding exons are indicated by colored rectangles. Note the conservation of a brinker enhancer within the intron of the neighboring Atg5 loci of flies and mosquitoes. Abbrevations: Ag, Anopheles gambiae; Dm, Drosophila melanogaster; Tc, Tribolium castaneum; sim, single minded; sog, short gastrulation; vnd, ventral nervous system defective. (Reproduced from Cande *et al.* 2009).

introns of the respective *cactus* transcription units (Fig. 1.3) (Cande *et al.* 2009).

Enhancer conservation at the *brinker* (*brk*) locus is even more dramatic. *brk* encodes a sequence-specific transcriptional repressor that helps restrict Dpp (BMP) signaling to the dorsal ectoderm (Jazwinska *et al.* 1999). In *Drosophila*, two separate enhancers regulate *brk* expression in the presumptive neurogenic ectoderm of pregastrular embryos (Hong *et al.* 2008a). One of the enhancers is located ~10 kb 5' of the *brk* transcription start site. The other is located 13 kb downstream of the start site, within the intron of a neighboring gene, *Atg5*. The major enhancer regulating *brk* expression in the *A. gambiae* embryo is located within the *Atg5* gene, even though the *brk* transcription unit is inverted relative to its orientation in *Drosophila* and Atg5 is located quite far, ~100 kb, from *brk* in the mosquito genome (Cande *et al.* 2009) (Fig. 1.3). Binding site turnover has been well documented in insect enhancers (reviewed by Ludwig 2002). Despite this turnover within existing enhancers, there might be constraints on the de novo evolution of developmental enhancers. We suggest that the evolution of novel patterns of gene expression depends primarily on the modification of ancestral enhancers rather than the invention of new ones.

1.6 Shadow Enhancers

ChIP-chip assays led to the comprehensive identification of Dorsal target enhancers in the *Drosophila* genome (Zeitlinger *et al.* 2007). These studies identified multiple enhancers at over one-third of the target genes that are directly regulated by the Dorsal gradient. For example, the *vnd* gene encodes a sequence-specific transcription factor that specifies the ventral-most neuronal cell identities of the ventral nerve cord (e.g. Weiss *et al.* 1998). It is activated by enhancers located in both the 5' flanking region as well as within the first intron of the transcription unit (Stathopoulous *et al.* 2002; Shao *et al.*, 2002; Zeitlinger *et al.* 2007). Similarly, *sog* is regulated by both a 5' enhancer and an intronic enhancer (Fig. 1.4), and as discussed above, *brk* is activated by enhancers located in both 5' and 3' flanking regions (Zeitlinger *et al.* 2007, Hong *et al.* 2008a).

We refer to the secondary enhancers located in remote 5' or 3' positions as shadow enhancers (Hong *et al.* 2008a). Preliminary studies suggest that they might help confer resilience in gene expression in response to genetic and environmental perturbations. For example, *vnd* and *sog* exhibit normal

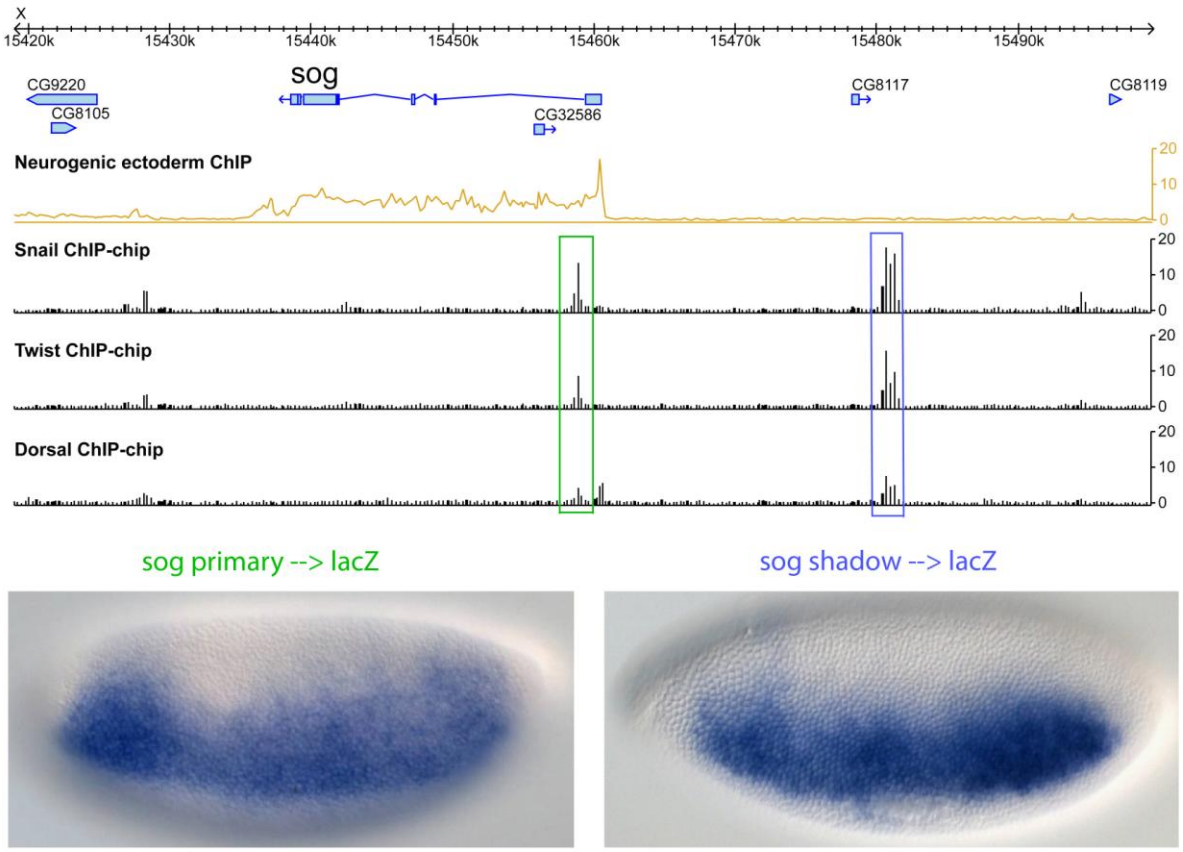


Figure 1.4 ChIP-chip assays identified two enhancers for the early *sog* expression pattern. The *sog* transcription unit is show in yellow, and the locations of Dorsal, Twist, and Snail binding sites are indicated below. There are two clusters of binding sites, in the first intron and over 20 kb 5' of the start site. The intronic cluster was previously shown to function as an enhancer for the *sog* expression pattern, and generates a similar pattern of expression when attached to a lacZ reporter gene and expressed in transgenic embryos (lower left). The distal "shadow" enhancer generates a similar pattern of expression (lower right).

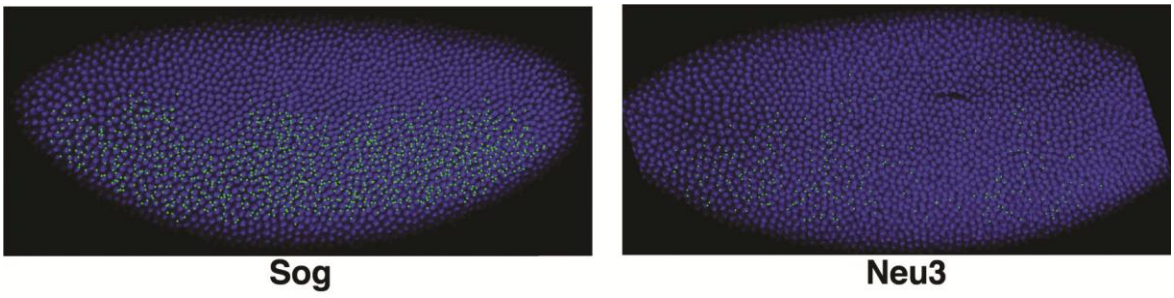


Figure 1.5 Onset of *sog* and *Neu3* expression in precellular embryos at the early phases of nuclear cleavage cycle 14. The embryos were collected from *dorsal*/+ females and therefore contain half the normal levels of the Dorsal nuclear gradient. The *sog* pattern is normal, but *Neu3* displays erractic activation – it should look more similar to the filled-in *sog* pattern, at left. *sog* contains a shadow enhancer while *Neu3* does not. A general correlation between the robustness of the pattern and the existence of shadow enhancers holds across several DV patterning gene examples. (Figure modified from Boettiger and Levine 2009).

patterns of transcriptional activation in embryos derived from dl/+ heterozygotes (half the normal dose of the Dorsal gradient), whereas *Neu3* and *rho* display erratic patterns of activation (Boettiger and Levine 2009) (Fig. 1.5). *vnd* and *sog* contain shadow enhancers, whereas *Neu3* and *rho* do not. It is possible that dual enhancers for a common expression pattern ensure accurate and reproducible activation in large populations of embryos subject to environmental fluctuations.

It is possible that shadow enhancers arise from "cryptic" duplication events. Of course, other scenarios can be envisioned, but regardless of mechanism, once they arise shadow enhancers might confer an adaptive advantage to a population by ensuring accurate activation of critical developmental control genes. Shadow enhancers offer an opportunity for producing novel patterns of gene expression without disrupting the core function of the primary enhancer and associated gene. According to this view, the evolution of shadow enhancers might come at a cost to the fitness of a population, but this cost could be compensated by the advantages conferred by the novel mode of gene expression. Perhaps at some frequency there are situations in which selection is relaxed, perhaps for isolated populations in more moderate environments, which could at least temporarily reduce the costs of modifying these enhancers for novel use.

1.7 Shadow Enhancer Conservation

In parallel to experiments designed to test enhancer function in *D. melanogaster* (described in the following chapters), I also began to clone homologous regions from other species of *Drosophila* to test for enhancer conservation. Enhancer function was evaluated using reporter genes integrated into transgenic *D. melanogaster*. The goal was to determine if any of these known enhancers have gone missing in any species, or if there are differences in the expression patterns driven. Cases where one or both enhancers in a pair have taken on novel function would be particularly interesting.

Initial stages of this effort were computational. Binding site cluster analysis using the program ClusterDraw2 (Papatsenko 2007) provided a means of assessing the presence or absence of clusters of specific binding sites. The method uses known consensus sequences that are recognized by specific activators or repressors and gives increasing weight to better matches and

closer associations of potential binding sites. Figure 1.6 shows an example of cluster analysis output for a region near the gene *sog*. Regions with strong cluster scores align well with ChIP-chip data for the same transcription factors (in this case Dorsal, twist, and snail) at the location of the known proximal and shadow enhancer (Hong *et al.* 2008a). The inclusion of zelda binding sites often increases the strength of signals from known enhancers and was also included.

Computational cluster analysis has the advantage of being informative for species in which genome wide ChIP data is unavailable. Clusters were predicted for both proximal and shadow enhancers for the gene sog across the Drosophila, though the signal is weak for the D. virilis proximal enhancer.

All of these enhancers were shown to be conserved through reporter experiments (Fig 1.7). A similar analysis was performed for the gene *snail* and its pair of enhancers, the subject of Chapter 2. In this case, only enhancers from one of the more distant species was tested (*D. virilis*) (Fig 1.8). Clusters of Dorsal, twist, snail, and zelda binding sites are present in all species examined. Enhancers from *D. virilis* drive mesodermal expression in *D. melanogaster*, indicating a conservation of enhancer position and at least partial conservation of function in patterning. Despite conservation of enhancer location, these two enhancers drive less similar patterns than the same two enhancers in *D. melanogaster* (Fig 1.9), see Chapter 2. The proximal enhancer drives much narrower expression than its shadow counterpart, which is also slightly narrower than *D. melanogaster* endogenous *snail*.

A final comparison was performed for an enhancer pair characterized further in Chapter 3, for the gene *hunchback* (Fig1.10). Again, like *snail*, even the "better" of the two enhancers falls slightly short of the endogenous *D. melanogaster* boundary. In this case the shadow drives a significantly restricted pattern of reporter expression.

For each of the three genes examined, *sog*, *sna*, and *hb*, functional enhancers exist in each homologous position in each species. Each contributes to the expression of the relevant gene. Despite some interesting differences in the extent of the pattern driven by especially the *snail* primary and the *hunchback* shadow, these enhancers are well conserved across species that diverged ~40 million year ago. This high degree of conservation implies a role for shadow enhancers in fitness and specific developmental function.

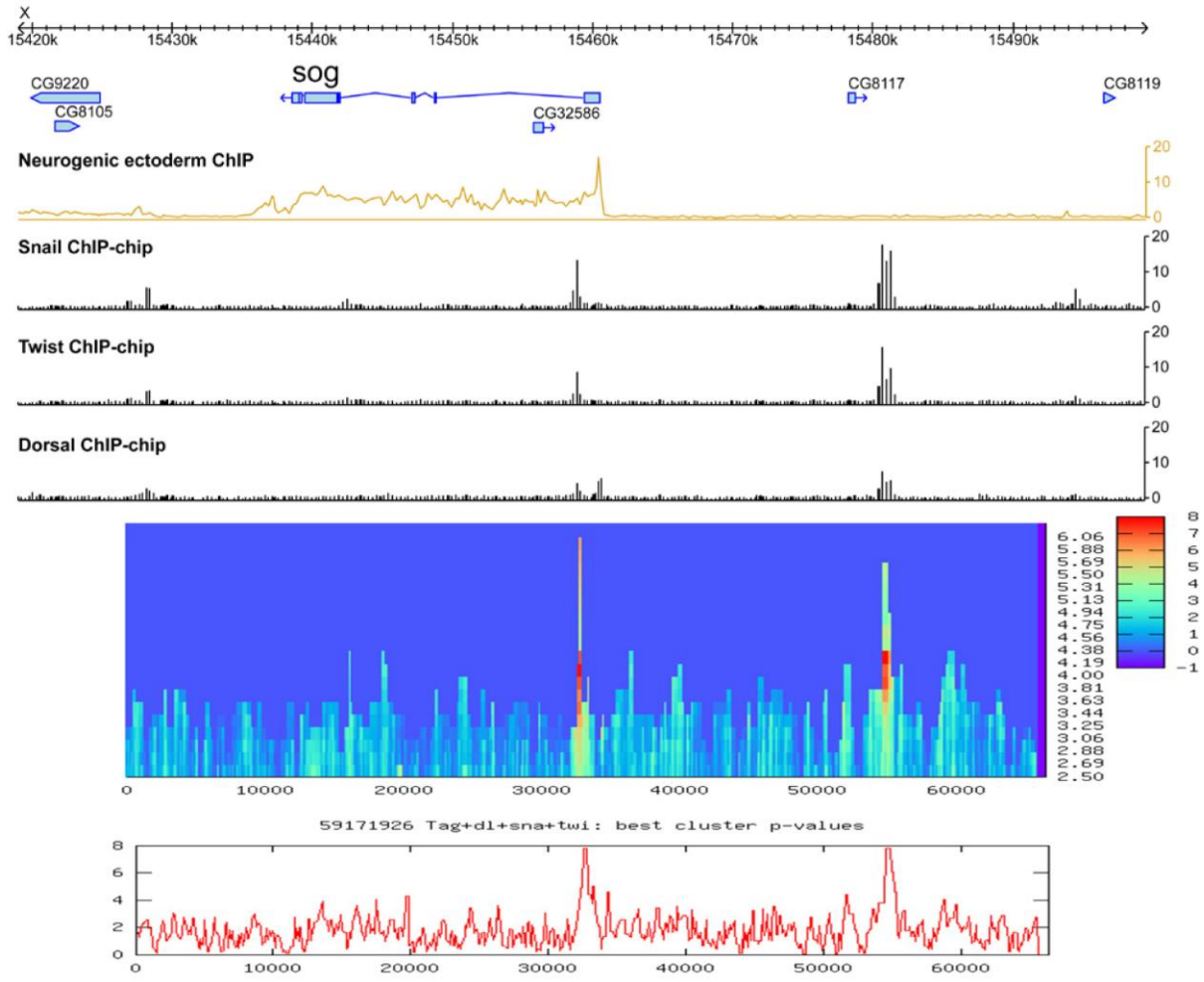


Figure 1.6 Binding site cluster analysis: a means of searching for signatures of enhancer conservation. Output from custom cluster analysis software is shown as a heat map aligned against the genomic region for the gene *sog* and DV transcription factor ChIP-chip data (as in Fig 1.4). Binding motifs used in the cluster analysis were for Dorsal, twist, snail, and zelda. This computational approach can be used when ChIP data is unavailable. Candidate peaks can then be tested further. Increasing cluster scores are shown as a range increasing from purple/blue to orange/red, key at right.

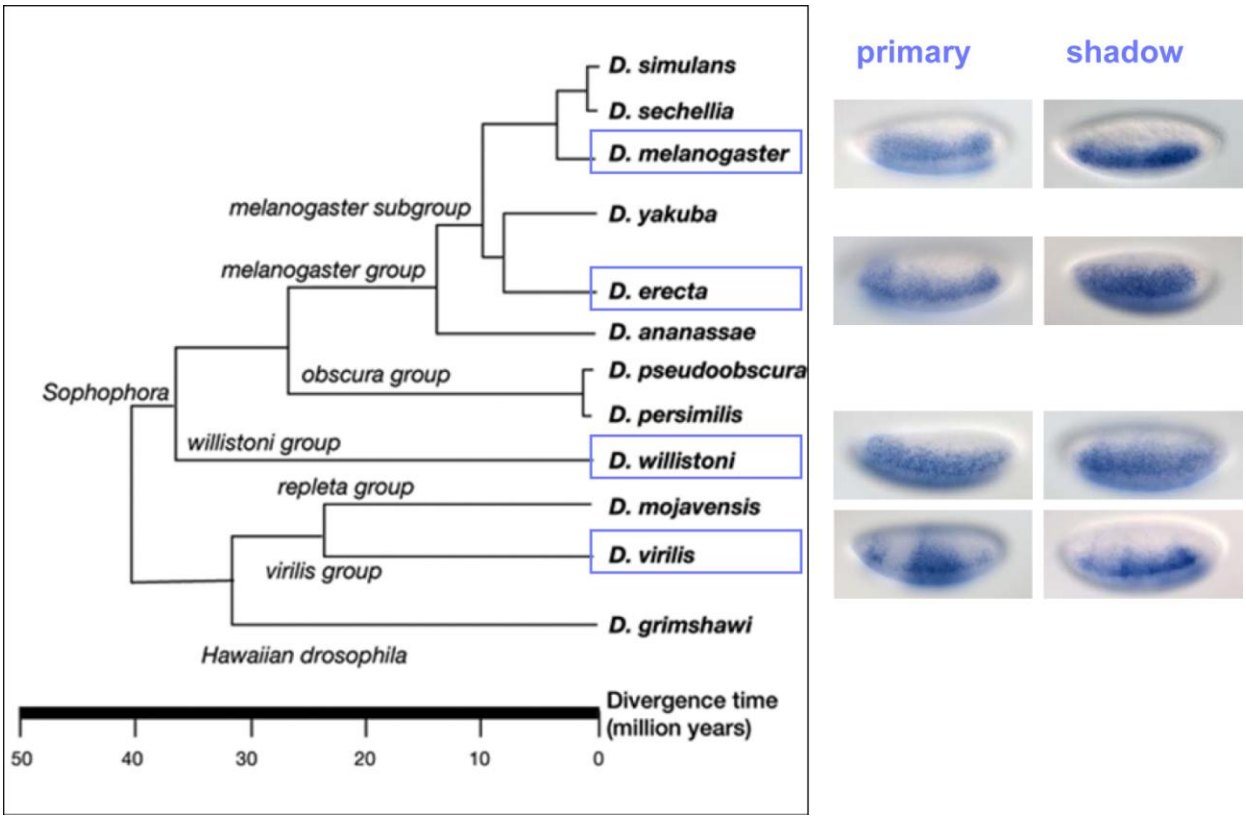


Figure 1.7 The *sog* primary and shadow enhancers are conserved across the *Drosophila*. Enhancer regions from selected species were tested in *lacZ* reporter assays by making transgenic *D. melanogaster* lines. Example embryos from each of these lines are shown at right.

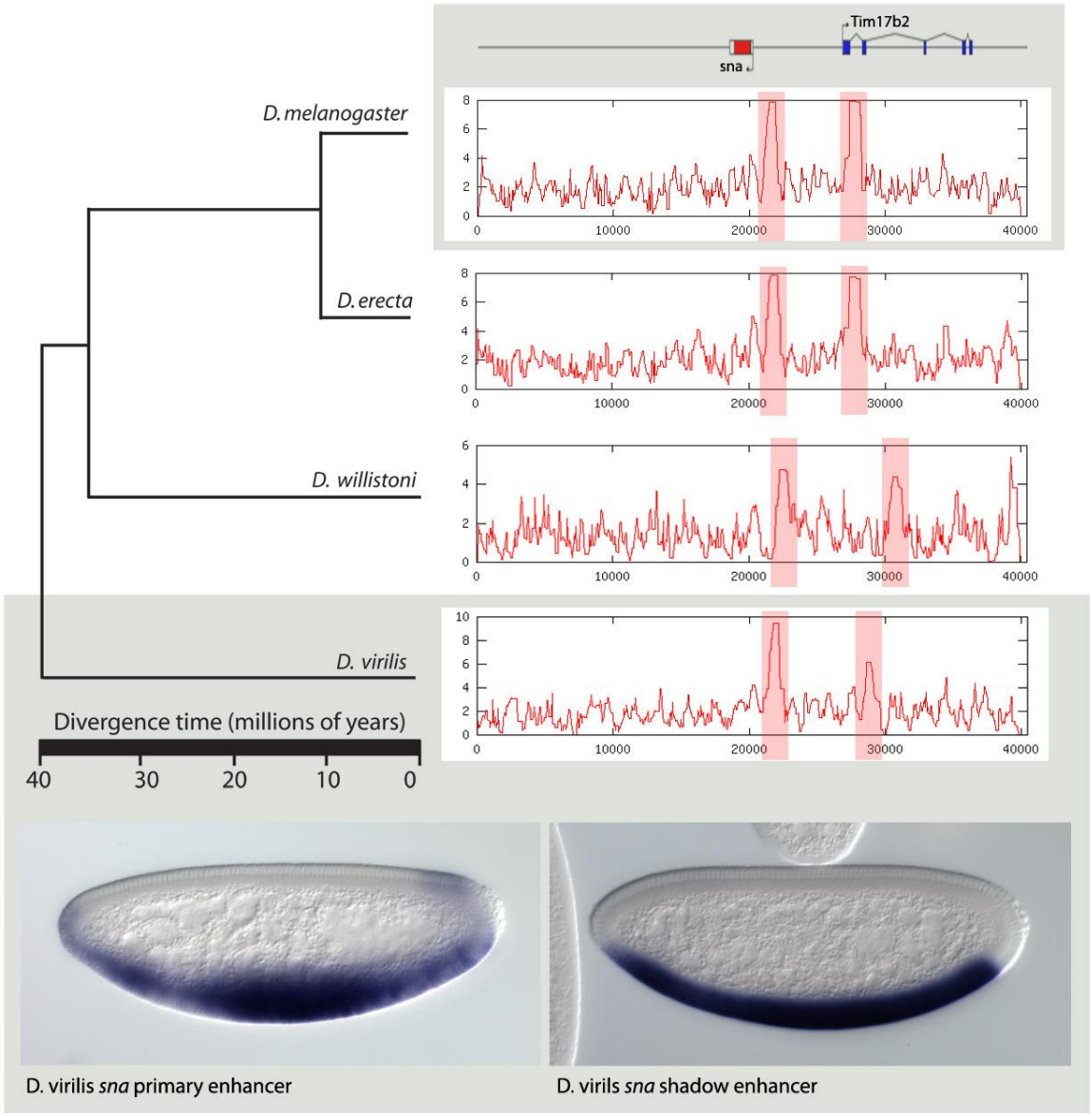


Figure 1.8 Cluster analysis indicates the conservation of the *snail* proximal and shadow enhancers. Cluster analysis as in Figure 1.6. Significant peaks are shown highlighted in light red. The *D. virilis* enhancers were tested further in *lacZ* reporter assays in transgenic *D. melanogaster* lines and shown to be functional mesodermal enhancers.

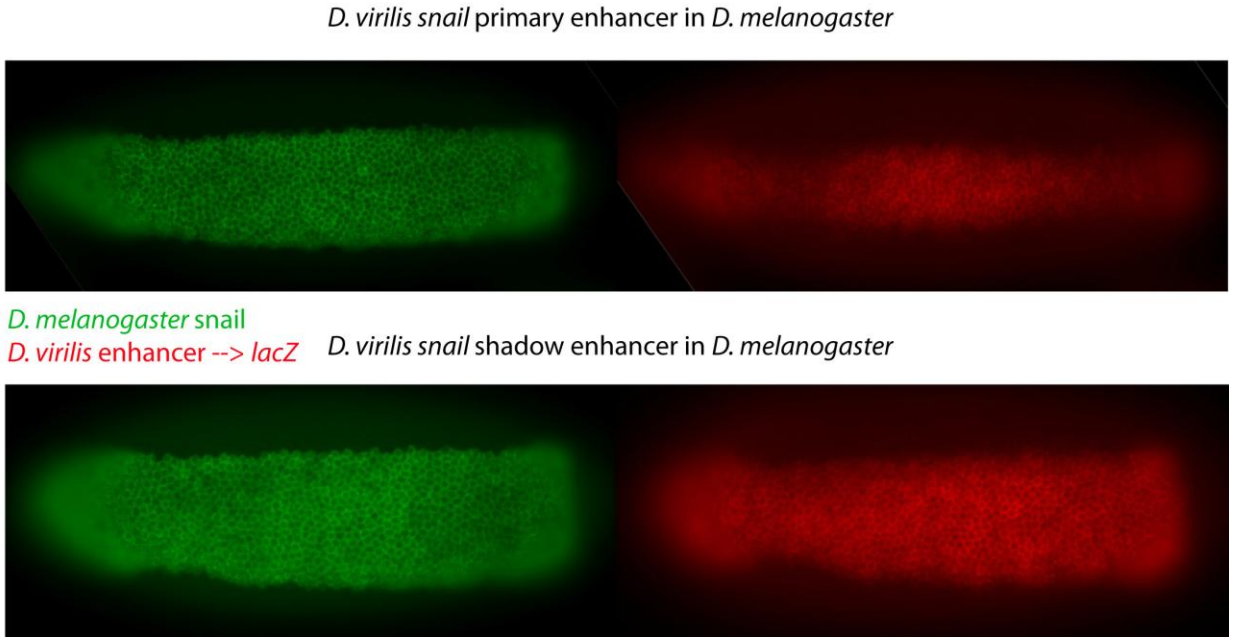


Figure 1.9 The *snail* proximal and shadow pair are more different in *D. virilis* than in *D. melanogaster*. In the top panel, the same D. melanogaster embryo is shown in two channels – endogenous *snail* in green, and a transgene containing a *D. virilis snail* primary enhancer driving reporter expression in red. Note that the *D. virilis snail* primary drives a much narrower pattern. The same comparision is shown in the bottom panel but with the *D. virilis snail* shadow. This pattern approaches the D. melanogaster endogenous *snail* pattern width more closely. Despite conservation of enhancer location, the two transgenes drive patterns that are more different from one another than do the *D. melanogaster* pair of *snail* enhancers, as will be seen in Fig 2.1.

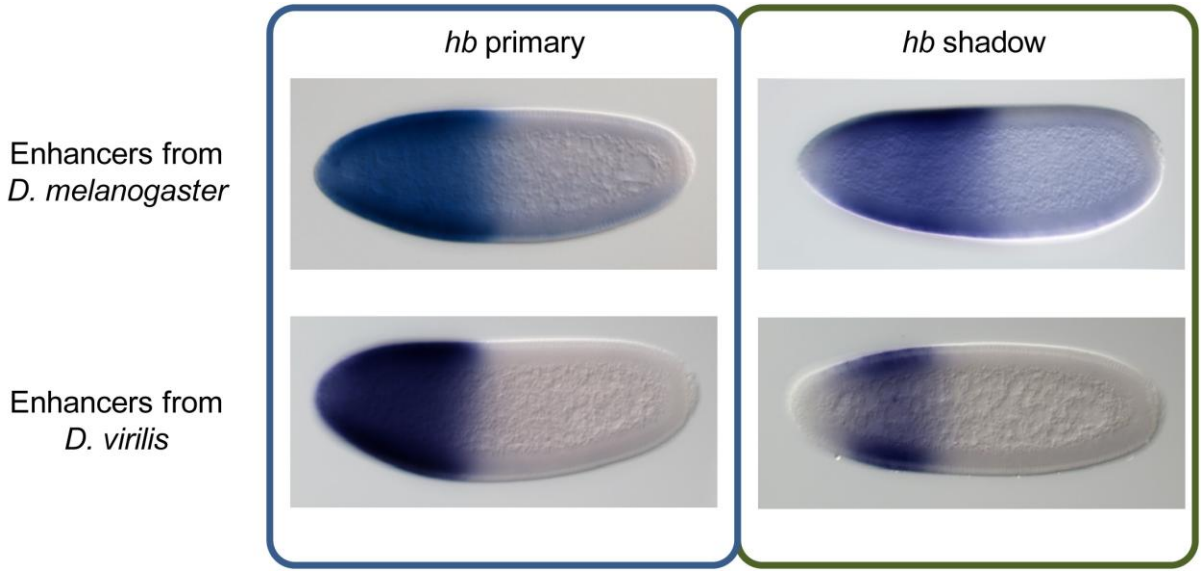


Figure 1.10 The *hunchback* primary and shadow enhancers are conserved in *D. virilis*, but are less similar. These enhancers in *D. melanogaster* are nearly identical and have similar regulatory logic. The same pair in *D. virilis* look more different from one another, and even the primary has an anteriorly shifted border when driving reporter expression in *D. melanogaster*. Compare to Fig 3.1A and B.

Chapter 2:

2.1 Chapter Overview

In this chapter I describe initial experiments designed to test shadow enhancer function. The work focused on the gene *snail*, and was published in Current Biology in 2010 (Perry *et al.* 2010).

2.2 Summary

Critical developmental control genes sometimes contain "shadow" enhancers that can be located in remote positions, including the introns of neighboring genes (Hong *et al.* 2008a). They nonetheless produce patterns of gene expression that are the same or similar as those produced by more proximal primary enhancers. It was suggested that shadow enhancers help foster robustness in gene expression in response to environmental or genetic perturbations (Boettiger and Levine 2009; Perry *et al.* 2009). We critically tested this hypothesis by employing a combination of bacterial artificial chromosome (BAC) recombineering and quantitative confocal imaging methods (Boettiger and Levine 2009; Venken *et al.* 2006). Evidence is presented that the *snail* gene is regulated by a distal shadow enhancer located within the neighboring *Tim17b2* locus. *snail* encodes a zinc finger transcription factor that has been implicated in epithelial/mesenchyme transitions (EMT) in a broad spectrum of developmental processes and cancers (Leptin 1991; Kosman *et al.* 1991; Barrallo-Gimeno and Nieto 2005). Removal of the proximal primary enhancer does not significantly perturb *snail* function, including the repression of neurogenic genes and formation of the ventral furrow during gastrulation at normal temperatures of development. However, at elevated temperatures there is sporadic loss of *snail* expression and coincident disruptions in gastrulation. Similar defects are observed at normal temperatures upon reductions in the levels of Dorsal, a key activator of *snail* expression (reviewed in Hong *et al.* 2008b). These results suggest that shadow enhancers represent a novel mechanism of canalization, whereby complex developmental processes "bring about one definite end-result regardless of minor variations in conditions…" (Waddington 1942).

2.3 Results and Discussion

Despite both intrinsic and environmental sources of noise, which introduce variability in complex developmental processes, the patterning of the *Drosophila* embryo unfolds with high fidelity (e.g., Gregor *et al.* 2007). It has been postulated that gene interactions in developmental regulatory networks can channel these variable inputs into faithful outcomes, as a ball bouncing inside of a funnel is channeled to the center, a process termed "canalization" (Waddington 1942). Here we present evidence that shadow enhancers (Hong *et al.* 2008a) are important mediators of canalization, ensuring reliable and robust expression of critical patterning genes.

snail is a key determinant of dorsal-ventral patterning (Leptin 1991; Kosman *et al.* 1991; Boulay *et al.* 1987; Ip *et al.* 1992). It encodes a zinc finger repressor that establishes a sharp boundary between the presumptive mesoderm and neurogenic ectoderm, and is essential for the formation of the ventral furrow and invagination of the mesoderm. Whole genome ChIP-chip assays identified a cluster of Dorsal and Twist (key activators of *snail* expression) binding sites in the immediate 5' flanking region of the *snail* transcription unit that coincide with the known enhancer (Ip *et al.* 1992; Zeitlinger *et al.* 2007). Unexpectedly, these studies also identified a second cluster of binding sites within the neighboring Tim17b2 locus, located ~7 kb upstream of *snail*. A small genomic DNA fragment (~1 kb) encompassing this second cluster of binding sites was attached to a lacZ reporter gene and expressed in transgenic embryos (Fig. 2.1). The fusion gene exhibits localized expression in the presumptive mesoderm, similar to that seen for the endogenous gene (e.g., Fig. 2.1) or obtained with the proximal enhancer (the first 2.8 kb of the 5' flanking region; see Ip *et al.* 1992). We arbitrarily refer to the newly identified distal enhancer as the shadow enhancer and the original, proximal enhancer as the primary enhancer as a means of distinguishing the pair (Hong *et al.* 2008a). In each case to date, the more distal enhancer has been the more recently discovered and will be referred to as the "shadow" enhancer.

A *snail* fusion gene containing only the primary enhancer rescues the gastrulation of at least some *snail* mutants in a population of mutant embryos (Hemavathy *et al.* 2004). Since *snail* is essential for the coordinated invagination of the mesoderm during early gastrulation, variability in expression could lead to occasional disruptions in morphogenesis. Perhaps the additional enhancer provides a mechanism for suppressing such

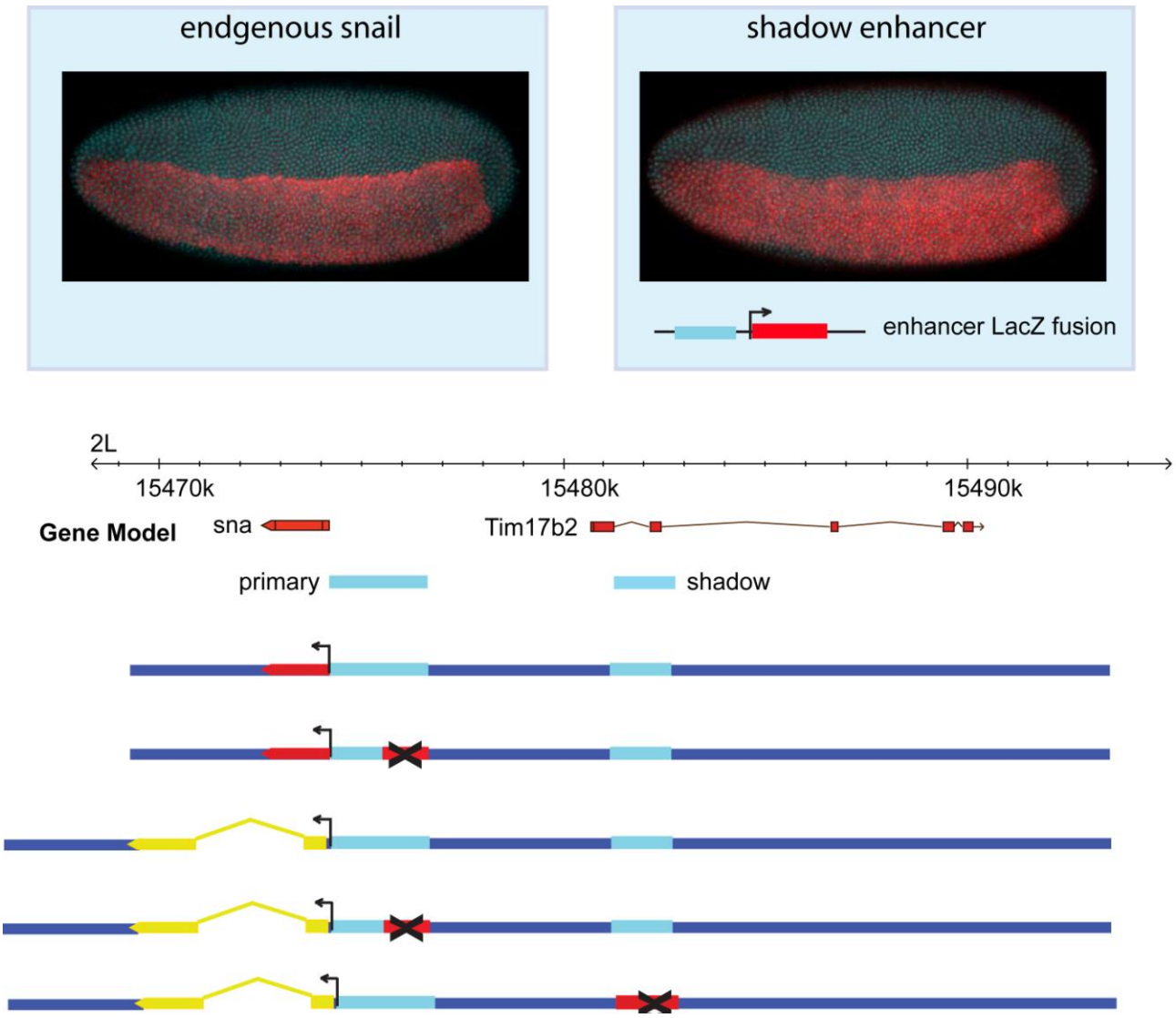


Figure 2.1 Identification of a *snail* shadow enhancer. The *snail* gene is expressed in the presumptive mesoderm, (top left in red). An intronic region in neighboring Tim17b2 was shown to be bound by transcription factors that regulate *snail* (Hemavathy *et al.* 2004) and here is shown to drive expression of a lacZ fusion gene in the mesoderm in a pattern qualitatively similar to the endogenous gene (upper right panel). Below, schematic representations of the BAC constructs used in subsequent experiments are aligned to the gene model. In all figures, anterior is to left, dorsal is at top, unless indicated.

variability, thereby ensuring robust expression in large populations of embryos.to changes in activator concentration than similar genes lacking shadows (Boettiger and Levine 2009).

An alternative view is that the proximal and shadow enhancers are primarily responsible for controlling distinct dynamic aspects of the *snail* expression pattern, rather than functioning in an overlapping manner during mesoderm invagination. An expectation of the former "robustness hypothesis" is that transgenes containing either enhancer alone should be sufficient to induce gastrulation in *snail* mutant embryos. We tested this possibility by creating a series of recombineered BACs (Venken *et al.* 2006; Venken *et al.* 2009) containing a ~25 kb genomic interval encompassing the *snail* and *Tim17b2* loci (Fig. 2.1). Comparable BACs were prepared that either contain or lack the proximal enhancer. This enhancer was not simply deleted, but a ~1 kb segment containing critical Dorsal activator elements was replaced with a "random" DNA sequence (see Experimental Procedures) in order to retain normal spacing of the regulatory region.

To measure the effect that different enhancers have on transcriptional activity we developed a reporter system for detecting nascent transcripts. The endogenous *yellow* gene is not transcribed until late in development and contains a large intron (e.g., (Jeong *et al.* 2008; Parkhurst and Corces 1986), making it an ideal reporter for the detection of *de novo* transcripts by *in situ* hybridization. In contrast, the *snail* transcription unit lacks introns and is therefore not amenable to quantitative in situ hybridization methods that rely on intronic probes. Consequently, a series of BACs were created that contain *yellow* in place of *snail*. These BACs contain both enhancers or have either the primary or shadow enhancer replaced with random DNA (Fig. 2.1). All of the aforementioned BACs were inserted in the same chromosomal location on 2L using phiC31 targeted integration (Venken *et al.* 2006; Groth *et al.* 2004; Bischof *et al.* 2007).

BACs containing the *snail* gene were crossed into a mutant background with a deletion spanning the entire *snail* transcription unit (*Df (2L)osp*^{29}) along with a marked balancer to identify homozygous *snail* null mutants. As noted earlier, the reciprocal situation, proximal enhancer without shadow, can sometimes rescue gastrulation (Hemavathy *et al.* 2004). Mutant embryos homozygous for the *snail* deficiency chromosome (*osp*^{29}) are easily

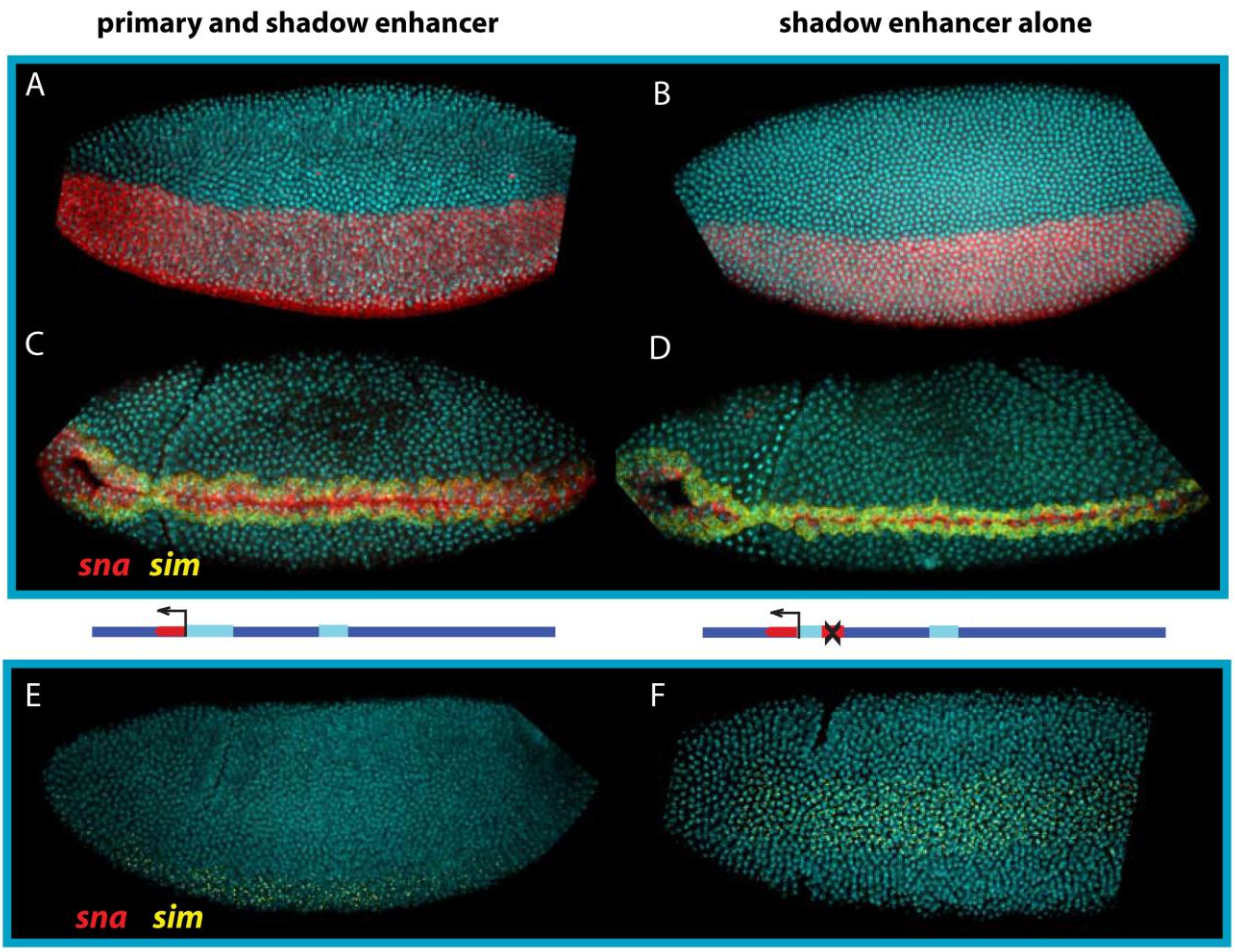


Figure 2.2 The *snail* shadow enhancer rescues gastrulation. A. The rescue BAC construct in a *sna* mutant background drives *sna* expression (red) uniformly throughout the mesoderm in cycle 14 embryos. B. The pattern driven only by the BAC with the primary enhancer deleted is qualitatively similar. C. During gastrulation, all *sna* expressing cells migrate into the interior of the embryo. A single row of cells flanking the *sna* domain express *sim*, shown in yellow. D. *sna* driven without the primary enhancer is sufficient to induce normal gastrulation and normal *sim* expression when these embryos are raised at 22C. E. In embryos lacking the snail BAC rescue construct, no *sna* is expressed. Instead, sim is expressed throughout the ventral region. Without *sna* there is no mesodermal invagination. Lateral view. F. Embryo as in (E), mesodermal view.

system that is normally excluded from the mesoderm by the Snail repressor Kasai *et al.* 1992; Nibu *et al.* 1998) (Fig. 2.2E,F).

There is neither a ventral furrow nor subsequent ingression of the mesoderm in these mutants (e.g., Leptin 1991; Kosman *et al.* 1991). BAC transgenes containing both enhancers (Fig. 2.2A) or just the shadow enhancer alone (Fig. 2.2B) rescue gastrulation of mutant embryos (Fig. 2.2C,D; compare with E,F). In both cases, a complete ventral furrow is formed, followed by invagination of the mesoderm indistinguishable from that seen in wild-type embryos. Both BACs restore *snail* expression in the presumptive mesoderm, and *sim* transcripts are restricted to lateral regions that form the ventral midline of the CNS after gastrulation. These observations, along with previous studies (e.g., Hemavathy *et al.* 2004), indicate that neither the primary nor shadow enhancer is necessary for the gastrulation of embryos raised at optimal, permissive conditions.

It should be noted that while these lines are capable of rescuing early development and gastrulation, we did not obtain full genetic rescue to adulthood. Something is missing from these relatively small BACs, presumably a more distant enhancer that is necessary for some function later in development. Embryos that have no native copies of the snail locus were identified by using a labeled balancer (this label marks the non-mutant embryos) as described in the Methods section, and for the purposes of these experiments this was sufficient. Each modified construct was compared directly to a control construct. It should also be noted that in a later work, another group was able to obtain full genetic rescue by taking a similarly sized BAC but shifting it slightly to encompass additional 3' sequence (Dunipace *et al.* 2011). This BAC did not incorporate as much 5' sequence, and this may have additional affects on gene expression, but was sufficient to make a stable line containing only this BAC and not a native functional copy of the *snail* gene.

Although the shadow enhancer is sufficient for generating a qualitatively normal pattern of *snail* expression, additional assays were done to determine whether there might be subtle changes in expression. Quantitative confocal imaging methods were used to investigate this possibility (see Boettiger *et al.* 2009). As mentioned earlier, BAC transgenes were prepared that contain the *yellow* reporter gene in place of the *snail* transcription unit. In situ hybridization assays with intronic probes permit direct detection of *yellow* de novo transcripts, and hence, precise measurements of *snail* transcription

with single cell (nucleus) resolution. At normal culturing temperatures, 22C, there is no discernible difference in the initial *de novo* transcription patterns of BAC transgenes containing both enhancers (Fig. 2.3A) or just a single enhancer, either the primary enhancer or shadow enhancer (Fig. 2.3B). In the majority of cases more than 90% of the nuclei in the presumptive mesoderm express *yellow* transcripts.

Less reliable expression is observed for BAC transgenes containing a single enhancer at elevated temperatures, 30C (Fig. 2.3C,D). More than 20% of the nuclei in the presumptive mesoderm lack *yellow* transcripts in over half of the embryos expressing the BAC transgene without the shadow enhancer. This effect is even more pronounced upon removal of the primary enhancer. The same cut-off value, absence of *yellow* transcripts in at least 20% of all mesodermal nuclei, occurs in over three-fourths of these embryos (Fig. 2.3). In contrast, the BAC transgene containing both the primary and shadow enhancers continues to display nearly complete patterns of *de novo* transcription at the elevated temperature.

Similar results were obtained in response to genetic perturbations (Fig. 2.4A,B). For example, the *yellow* transgene BAC containing both enhancers exhibits a normal pattern of expression in embryos derived from *dl*/+ mothers containing half the normal dose of the Dorsal gradient (Fig. 2.4A). The distribution of nuclei failing to maintain active expression is similar to that seen for wild-type embryos (Fig. 2.4C). However, the comparable BAC transgene containing only the shadow enhancer exhibits erratic patterns of activation in these embryos, particularly in lateral regions (Fig. 2.4B, quantification in C). These results, along with the preceding analysis of embryos grown at elevated temperatures, suggest that the *snail* shadow enhancer helps ensure accurate and reproducible patterns of gene expression in large populations of embryos subject to genetic and environmental perturbations.

The preceding results document quantitative changes in the variability and reliability of *snail* expression upon removal of the primary or shadow enhancer. We next asked whether such variation causes changes in cellular morphogenesis, particularly the formation of the ventral furrow and subsequent invagination of the mesoderm (Fig. 2.4D,E). *snail* mutant embryos carrying BACs with both enhancers (Fig. 2.4D) or just the shadow

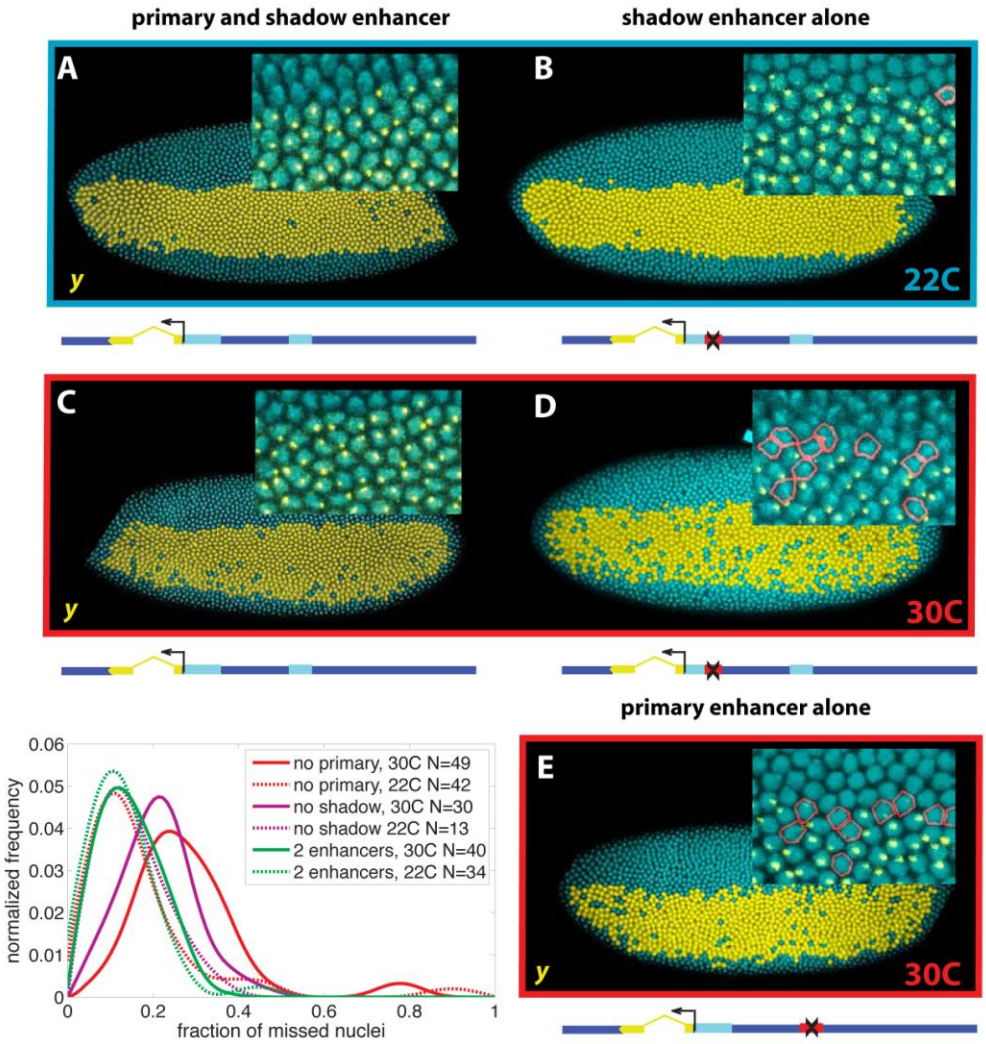


Figure 2.3 Multiple enhancers ensure robust gene expression under different thermal conditions. A. Visualization of expression of the *yellow* reporter gene from the BAC containing the *sna* locus, stained for the *yellow* intron. Cells actively transcribing the reporter are shown in yellow. Intronic probes show a single bright point of transcription inside actively transcribing nuclei (inset); all embryos are heterozygous and have one copy of the reporter. Nuclei that express the endogenous gene but not the reporter are outlined in red. A schematic representation of the BAC is shown below the embryo. **B.** At 22C a similar degree of uniform expression is exhibited by embryos carrying a *yellow* BAC lacking the primary enhancer. **C.** At 30C embryos with both enhancers still show straight boundaries and a small percent of inactive nuclei. **D.** Embryos lacking the primary enhancer at 30C show substantially more ragged boundaries of expression and a greater percent of inactive cells in the mesoderm. **E.** Embryos lacking the shadow enhancer are similar to those lacking the primary at both temperatures. Lower left: Frequency distributions of the fraction of cells in the *sna* expressing region that lack *yellow* expression are plotted for each of the 6 different embryo populations. N indicates the number of embryos in each population sample.

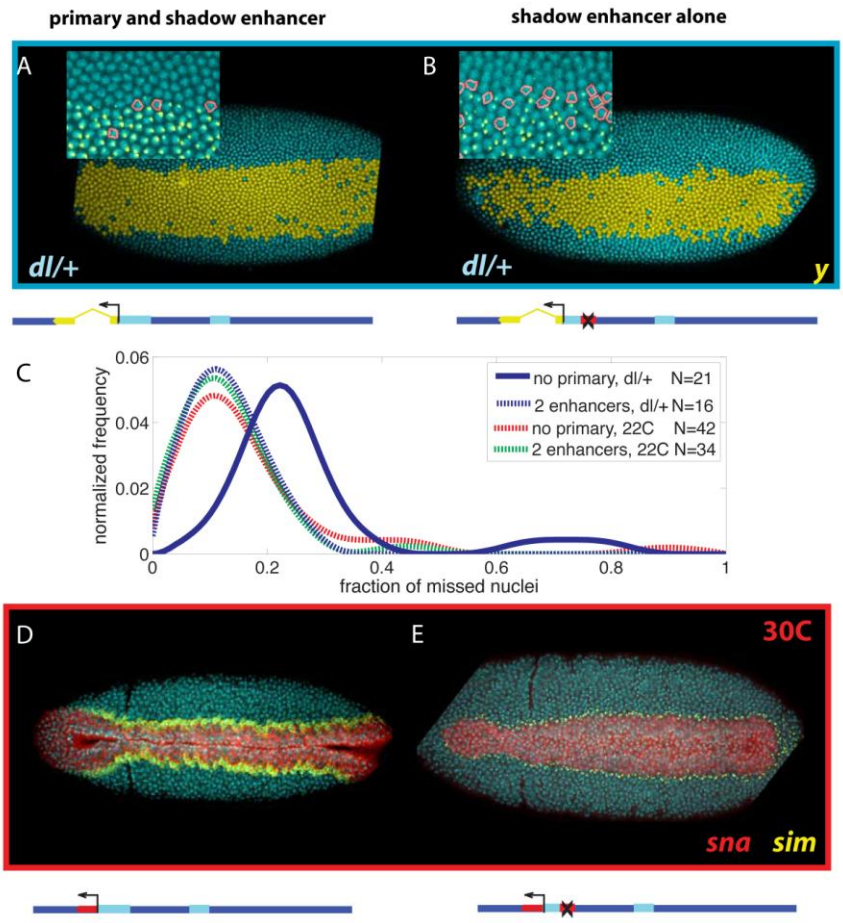


Figure 2.4 The effect of intrinsic and extrinsic variability. A. Embryos from dorsal heterozygote mothers raised at 25C show uniform yellow expression when driven with both enhancers. Only a few cells are lacking active expression (inset). B. Embryos with a single enhancer in this background show substantially greater loss of expression and ragged boundaries. C. The distribution nuclei which fail to maintain active transcription shifts to the right in the dorsal heterozygote background only for embryos lacking one of the enhancers. D. All observed embryos raised at 30C (N=28) from a population heterozygous for the BAC-constructs containing both enhancers gastrulate normally, forming a straight ventral furrow; note stage of development by presence of cephalic furrow. E. Some embryos from a similar population, but with only the single enhancer and raised at 30C show various defects in gastrulation (N=10 of 14). Note embryo stage by presence of cephalic furrow, yet lack of significant mesodermal invagination. The number of snail expressing cells anterior to the cephalic furrow is also reduced. Supplemental Figure 2 shows a range of defects observed in these embryos; a narrower pattern of anterior expression may result in delays in involution of anterior regions, and some exhibit a more erratic midline.

enhancer (Fig. 2.4E) were grown at elevated temperatures, 30C. Embryos carrying the transgene with both enhancers exhibit normal patterns of gastrulation (Fig. 2.4D). In contrast, comparable embryos lacking the primary enhancer display erratic patterns of gastrulation, including the formation of incomplete ventral furrows that do not extend along the entire germband (Fig. 2.4E) and disruptions in the symmetry of the involuted mesodermal tube (see Fig. 2.5). As shown earlier, such defects are not observed at normal temperatures, 22C (Figs. 2.2 and 2.5).

We have presented evidence that the *snail* shadow enhancer located within the *Tim17b2* locus helps ensure reliable and reproducible patterns of *snail* expression in the presumptive mesoderm during gastrulation. BAC transgenes lacking either the primary enhancer or the shadow enhancer display erratic patterns of *de novo* transcription at elevated temperatures. We propose that shadow enhancers come to be fixed in populations by ensuring robustness in the activities of key patterning genes such as *snail*. binding sites, but an additional enhancer could suppress this "noise" by increasing the probability of gene activation. This increased time of active transcription per cell might augment the overall levels of expression, which could be an important function of shadow enhancers.

Other critical dorsal-ventral determinants also contain shadow enhancers, including *brinker*, *vnd*, and s*og* (Hong *et al.* 2008a). The recent analysis of *shavenbaby* suggests that shadow enhancers are essential for the reliable morphogenesis of embryonic bristles in older embryos (Frankel *et al.* 2010). There is also evidence that shadow enhancers might be a common feature of vertebrate systems, such as zebrafish (Kikuta *et al.* 2007).

Shadow enhancers appear to represent a novel mechanism of canalization, whereby complex developmental processes lead to a fixed outcome despite genetic and environmental perturbations (Waddington 1942). Other mechanisms of canalization have been suggested, including recursive wiring of gene regulatory networks and "capacitors" such as hsp90 that suppress both altered folding of mutant proteins and transpositioning of mobile elements (Lott *et al.* 2007; Manu *et al.* 2009; Rutherford and Lindquist 1998; Specchia *et al.* 2010).

It is conceivable that primary and shadow enhancers mediate overlapping patterns of activity only during early embryogenesis. They might come to

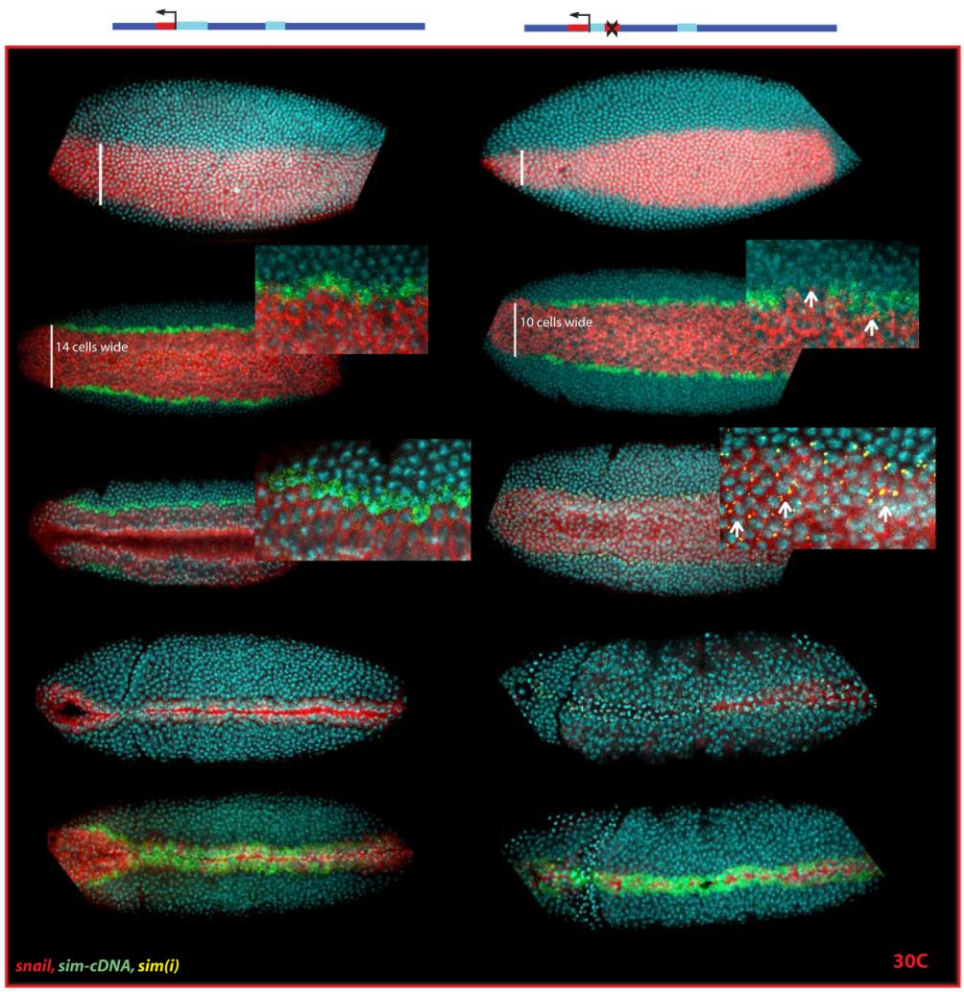


Figure 2.5 Effects of enhancer number on rescue of early snail function. Left panels show from pre-gastrulation to late gastrulation for *snail* mutant embryos rescued with the unmodified *snail* locus BAC (with *snail* labeled in red) when raised at 30C. Expression spans 14-18 cells throughout the ventral portions of the embryo. *sim* expression is induced in a single line of lateral cells (green). Apical constriction of cells along the midline leads to the formation of a ventral furrow uniform along the length of the embryo. Invagination is complete when the two lines of *sim* expressing cells meet in the midline. In some embryos rescued with the modified BAC lacking the primary enhancer (right panels), the anterior portion of *snail* expression is abnormally narrow. Some embryos show delayed onset of gastrulation at the anterior pole. The mesoderm invaginates completely, bringing together the two lines of sim expressing cells, sometimes in a less straight line. The completion of invagination in some embryos lags behind initiation of germband extension.

possess distinctive regulatory activities at later stages of development. Nonetheless, during the time when their activities coincide during gastrulation, they maintain reliable patterns of *snail* expression in response to environmental and genetic variability. Although either enhancer might be sufficient, both enhancers are required for consistently accurate and reliable patterns of expression in response to variability. This precise patterning enables rapid development, without delays arising from corrective feedback mechanisms. It is easy to imagine that delays in embryogenesis would result in selective disadvantages to the resulting larvae, which must compete for limiting sources of food. Regardless of the specific mechanisms that select for shadow enhancers, the occurrence of such enhancers provides an opportunity for the evolution of novel patterns of gene expression. As long as the two enhancers maintain overlapping activities during developmental "hotspots" such as gastrulation, they can drift or be selected to produce divergent patterns of gene expression. Some initial evidence of divergence of this particular pair was shown in Fig 1.8 and 1.9.

2.4 Experimental Procedures

2.4.1 Fly genetics

Positive BAC line males (labeled with w+) were crossed to yw; wg[Sp]/CyO; Pr,Dr/TM3,Sb,Ser virgins. Homozygous BAC lines were created by selfing the red-eyed, Sb,Ser flies from the F1 generation. Males from the BAC lines carrying the yellow reporter constructs were crossed to yw, dl^{6}/CyOvirgins.

To test for rescue of the BAC constructs, we generated a white eyed, double balancer strain carrying a CyO linked hunchback-LacZ reporter by crossing and back crossing wnt4/CyO, hb-lacZ (BSC 6650) to yw; wg[Sp]/CyO; Pr,Dr/TM3,Sb,Ser virgins. Positive BAC males were crossed into this line to create w; +/CyO, hb-lacZ; BAC[snail,w+]/TM3,Sb,Ser virgins. Simultaneously, w; *Df (2L)osp*^{29}/CyO, (BSC 3078) flies carrying a deletion spanning the snail gene were crossed to yw; wg^{Sp}/CyO; Pr,Dr/TM3,Sb,Ser virgins. The *Df (2L)osp*^{29}/wg^{Sp}, +/TM3,Ser males were crossed to the virgins containing the labeled balancer and the BAC. The progeny were selfed to create homozygous stable lines for the BAC carrying the snail deletion over the hb-lacZ marked CyO balancer. Populations still containing the Ser balancer or a wildtype chr III were analyzed also analyzed to test the effect of single copy rescue. The labeled balancer allowed for the reliable identification of embryos lacking a functional copy of endogenous *sna*.

2.4.2 Recombineering and transgenesis

Recombineering was performed as described in (Venken *et al.* 2006; Venken *et al.* 2009; Lee *et al.* 2001; Liu *et al.* 2003; Warming *et al.* 2005) with minor modifications. A detailed protocol can be found at http://flydev.berkeley.edu/cgi-bin/labpage/Levine_Lab/Resources.html; this protocol was modified primarily from detailed protocols found at the NCI-Frederick Biological Resources Branch web site by Søren Warming (Warming *et al.* 2005), and the P[acman] Resources web site at http://pacmanfly.org/ by Koen Venken (Venken *et al.* 2006; Venken *et al.* 2009).

In brief, CH321 or CH322 BAC clones that map to the region of interest were identified from end-sequenced BAC libraries which can be viewed on a browser at http://pacmanfly.org, and ordered from BacPac Resources (http://bacpac.chori.org/) (Venken *et al.* 2009). These BACs arrive already cloned into a vector containing attB sequence for targeted integration, mini-*white* cassette, chloramphenicol resistance, and are in the inducible copy number strain EPI300 (Epicentre Biotechnologies). BACs that do not require modification (such as control BACs for genetic rescue experiments) are ready for induction, preparation, and micro injection (see relevant section below). The experiments in this work involve the BAC CH321-18I14 containing the full *snail* and *Tim17b2* locus.

BACs requiring modification were first transformed into the recombineering strain SW102 (Warming *et al.* 2005), which was obtained from NCI-Frederick Biological Resources Branch. Cultures containing specific BACs were grown overnight and recombination functions were induced as described in (Venken *et al.* 2006; Warming *et al.* 2005). The induced bacteria were electroporated with targeting constructs that were prepared previously by PCR amplification. Targeting constructs were made using a pair of 90 base pair long oligonucleotides. These contained 25 base pairs specific to the region being amplified that was to be swapped into the BAC, and an additional 65 base pairs of sequence homologous to the target BAC flanking the region to be replaced. The homologous regions, or "homology arms", target the amplified sequence to the region of interest for recombination. A list of primers used is presented in Appendix I, Table 1. After electroporation and a one hour recovery period in 2XYT broth, bacteria were plated in a dilution series on LB plates with the appropriate antibiotic for overnight

incubation at 30C. Individual resulting colonies were screened by PCR for appropriate recombination at both homology arm locations.

Confirmed positive recombinant colonies were transformed back into EPI300 cells (Epicentre Biotechnologies) and reconfirmed by antibiotic marker selection and PCR; PCR products were sequenced for final confirmation. Restriction digests followed by gel electrophoresis were used to confirm that no significant rearrangements of the BAC had occurred during the recombineering process; SpeI and EcoRI fingerprinting digests of the original unmodified BAC were compared to digests of the modified BACs.

Oligonucleotides for amplification to make homology arm constructs (90 base pairs in length) were from Integrated DNA Technologies (IDT); shorter primers for colony screening PCR were from ELIM Biopharmaceuticals. Restriction enzymes were from New England Biopharmaceuticals. Qiagen products were used to isolate plasmid DNAs, gel-purify DNA fragments, and purify PCR products. Qiagen *taq* polymerase was used in colony PCR screening; Invitrogen Platinum *pfx* was used to amplify targeting constructs.

The ampicillin resistance cassette for the replacement of enhancer regions was PCR amplified from pBluescript (KS); construction of the *yellow* reporter fused to a kanamycin resistance cassette is described below.

2.4.3 Construction of the *yellow* reporter

Detecting sites of active transcription with high sensitivity requires that anti-sense RNA *in situ* hybridization probes be made against intronic regions of genes (where introns are available). Probes against exonic regions can initially detect these sites of transcription, but are often quickly masked from detection by the accumulation of mature mRNA. By targeting intronic regions, fluorescent *in situ* hybridization (RNA-FISH) signal is limited to near the site of transcription/splicing. Sensitivity of intronic probes, in our experience, is also often increased by the use of longer probes (up to 6kb) to maximize signal from these highly localized regions.

Traditional reporters fall short in one or both categories; for example, neither GFP nor lacZ contain introns and GFP is only ~700 base pairs long. In the interest of accurately detecting sites of active transcription, we developed a reporter that contains both intronic and exonic sequences, with a reasonably long intron (~2.7 kb), by using the adult pigmentation gene

yellow. This gene is not expressed in the early embryo, and as a lone component of an enzymatic pathway not normally present at this stage, has no noticeable effect on embryo development or phenotype. The intronic region was independently tested for enhancer activity by placing this sequence upstream of a traditional eve-lacZ fusion gene and did not drive expression in the early embryo.

For use in recombineering, the *yellow* sequence (~4.5kb) was PCR amplified from genomic DNA from a non-*yellow* mutant fly strain (the genome-sequenced strain has a C inserted into the ATG start codon, as does the original *yellow* mutant) and joined to a kanamycin resistance cassette (~1140bp region) that was PCR amplified from TOPO vector (Invitrogen) via fusion PCR. This combined fragment was then cloned into pBluescript (KS).

The *yellow*-kanamycin fragment was swapped into the place of the *snail* gene at the ATG start codon at the 5' end, leaving the *sna* 5' UTR intact. The endogenous 3' UTR was also left fully intact. In the BAC reporter constructs the large size of the *yellow*-kanamycin sequence (~5.6kb) created an overall insertion of about 1500 base pairs. This did not change the spacing between any regulatory elements at the locus known to be involved in patterning the early embryo. The primers used are listed in Appendix I, Table 1.

This reporter, when recombineered into the BAC in place of the endogenous *sna* gene and when integrated into fly line VK33 (Venken *et al.* 2006), gave excellent signal with low background using both full length RNA *in situ* probes and a probe specific to the intron of the gene (Fig. 2.6). It is interesting to note that a faint nuclear haze of signal accumulated within the nucleus but away from the distinct sites of transcription in these embryos. Perhaps either the splicing or the degradation rate can not keep up with production of these artificial transcripts/spliced intronic sequences in the early embryo. Despite this slight accumulation of faint signal, sites of active transcription remained exceptionally easy to detect.

2.4.4 Use of conditional origin of replication plasmids for the reduction of background

A certain percentage of colonies generated during each recombination experiment did not contain the appropriately modified BAC and were instead shown by PCR to contain the plasmid used as template for the targeting construct. This occurred despite the use of DpnI digestion after targeting

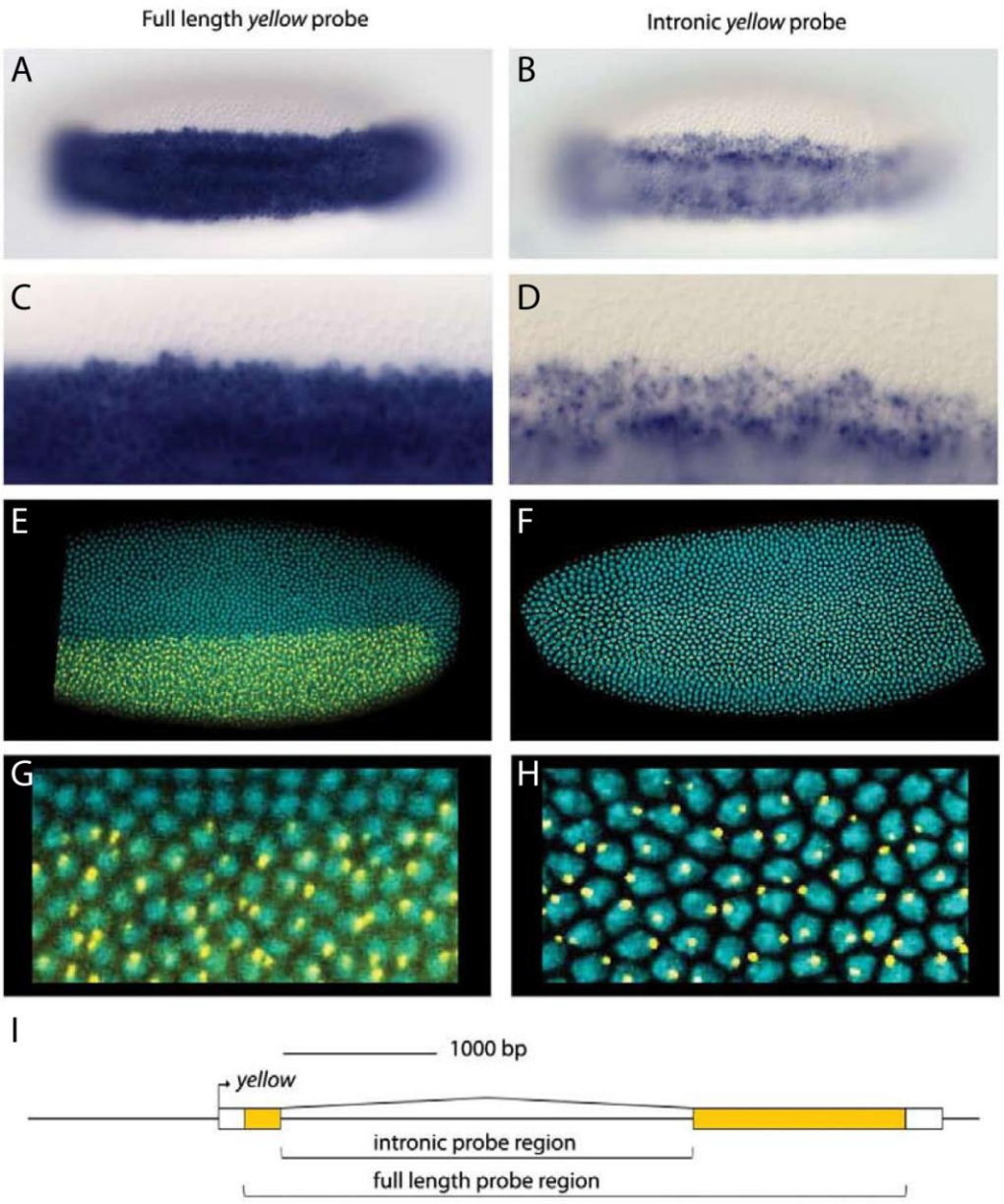


Figure 2.6 *yellow* reporter. A-B Colormetric *in situs* with the *yellow* full length probe (A), which gives robust expression and low background, and intronic probe (B), showing punctate dots from nascent expression. **C-D** Magnified views of embryos in A and B. **E** Fluorescent in situ with *yellow* full length probe, or, **F**, intronic probe. **G-H.** Magnified views of embryos in E-F, showing build up of cytoplasmic signal with the full length probe and isolated nascent transcripts with the intronic probe. **I.** Schematic of *yellow* gene and regions used to make probe templates.

construct PCR, where 2 uL of DpnI was added to each reaction and incubated at 37C for 2 hours before gel extraction. DpnI digestion should eliminate methylated (unamplified) plasmid DNA and gel extraction should further reduce background, though it in some cases remained a significant problem. Perhaps some plasmid escapes digestion; alternately, it has previously been suggested that this background is caused by aberrant PCR that produces, in effect, an unmethylated plasmid (Datsenko and Wanner 2000).

To further reduce background, regions to be targeted (such as the *yellow*-kanamycin fusion sequence) were subcloned into the multiple cloning site of a plasmid with a conditional origin of replication (oriR_γ) and this plasmid was maintained in a *pir*$^+$ host, as in Datsenko and Wanner, 2000). The conditional origin should prevent the SW102 strain (which lacks *pir*) from being able to replicate the plasmid containing the resistance marker, eliminating background colonies. In effect this did not eliminate all undesired background colonies, but did reduce the number generated by several orders of magnitude. The idea to use a plasmid with a modified origin of replication should be credited to Datsenko and Wanner, 2000, and was suggested to us by Russell Vance, who provided plasmid pSR47S [S1] and bacterial strain DH5α-pir.

2.4.5 BAC preparation for micro injection and phiC31 mediated integration

BACs were induced to high copy number using Epicentre BAC autoinduction solution, according to supplier's instructions, and grown overnight for 16-18 hours at 37C. DNA was prepared for micro-injection using the Invitrogen PureLink HiPure miniprep kit by following manufacturer instructions with described modifications for BACs and cosmids. DNA was diluted to a final concentration of ~300-400 ng/uL and 1x injection buffer. At least 200 embryos were injected per construct, either in-house or by BestGene Inc. (Chino Hills, CA).

All BACs were integrated into the same attP landing site to facilitate experimental comparison (Groth *et al.* 2004). Landing site line "VK33" (Venken *et al.* 2006) that has an attP site on chromosome 3 and that also contains the vasa promoter driving phiC31 integrase enzyme on the X chromosome (Bischof *et al.* 2007) was obtained from the Bloomington Drosophila Stock Center (stock number 24871).

2.4.6 Hybridization

Embryos that developed at 22C were collected at 2-4 or 2-6 hours of development (2-6 hours encompasses gastulation), and were fixed as described in (Kosman *et al.* 2004). Embryos that developed at 30C were collected from 1-4 hours to help account for more rapid development. All probes were made with digoxigenin conjugated haptens and biotin-conjugated haptins, and stained using sheep anti-digoxigenin and mouse anti-biotin primary antibodies (Roche Applied Sciences), followed by Alexa Fluor 555-donkey-anti-sheep and Alex Fluor 488 donkey anti-mouse secondary antibodies (Invitrogen, Molecular Probes).

2.4.7 Quantitative confocal imaging

Embryos were imaged on a Leica Scanning Confocal SL microscope with a 40x oil immersion objective. Confocal images were taken at 1024x1024 resolution with approximately 250 nm/pixel. Independent z-sections spanning the nuclear layer of the blastoderm embryo were sampled at 1/2 μm intervals to find any active transcripts in all nuclei (typically 14 sections for pre-gastrulating embryos, 20-40 sections for early gastrulation). Nuclei were counter-stained with Draq5 (Biostatus, Leicestershire UK) and simultaneously imaged with probes stained with AlexaFluor 488, and sequentially imaged for probes stained with AlexaFluor 555. The excitation and emission spectrum of Draq5 and AlexaFluor 488 are sufficiently separated to cause no measurable increase in background with our chosen emission bandpass, as confirmed experimentally through double staining. Embryos were staged by quantitative analysis of nuclear density (see below) and by developmental morphology. We analyzed embryos from the early moments of cleavage cycle 14 through the completion of gastrulation.

2.4.8 Automated image analysis

We wrote an automated image segmentation program in Matlab™R2008b (Mathworks) to identify and count all stained nuclei and detected probes. Z-stacks were projected into two-dimensional images by selecting the maximum intensity pixel in each stack, prior to further processing. The boundaries of each individual nucleus were determined using the DNA counter-stain, processed with a difference of Gaussians filter, which allows robust determination of nuclei using edge detection, size selection, and signal strength to inform classification. This was followed by an object dilation

algorithm to create a computational mask in which all pixels in an embryo are assigned to a uniquely identified nucleus. For each nucleus in the embryo (~2000 per embryo) the script determines the transcriptional activity state by colocalization of the *yellow* intronic in situ probe within the identified boundaries. True hybridization foci are identified by a difference of Gaussian filter algorithm. Nuclei and hybridization foci within 2 nuclear diameters (10-15 microns) from the edge of the embryo were systematically excluded from the analysis, as the slight curvature at the edge of the flattened embryo reduce reliability of detection in this region.

To address possible misclassification due to uncertainty in determining the nuclear boundary, transcripts localized to pixels adjacent to the nuclear boundary may be automatically reassigned to the neighboring nucleus if the original parent nucleus already contains an interior localized probe. This exploits the fact that the hetero allelic expression of the reporter should result in no more than a single transcript per nucleus. An iterative extension of this algorithm also insured reliable classification when several adjacent nuclei each had transcripts that localized to boundary pixels. Co-staining for the endogenous *snail* gene labels all of the cells in the snail expression region. Since *snail* has no introns and is only a short gene, probes do not give sufficiently reliable foci of nascent transcription. Instead, we use the sharply bounded region of mRNA accumulation to define the expression region for the endogenous transcript. Nuclei inside this region of expression which failed to express a *yellow* transcript were classified as "inactive". The script is implemented in a custom designed graphical user interface to allow user supervision and ensure appropriate classification at each step.

To estimate the uncertainty in our quantitative measurements we first checked the accuracy of our nuclear segmentation against manual nuclei determination. For all embryos examined, small adjustments to the filter parameters readily allow 100% of nuclei to be accurately segmented. The automated filter determinations routinely segment all but a handful of nuclei (1-2%) correctly prior to user input. We used double staining for overlapping probes to quantify the accuracy of our transcript localization filter and nuclear assignment routine. This analysis indicates 92-95% accuracy in identification and correct assignment of staining. 5-8% of transcripts are either not identified (due to failure of the probe to react, failure of the code to detect very weak signal or partial reaction, or assignment of a transcript to the wrong nucleus. This provides a measure of

the sensitivity limits of our assay to detect differences in uniformity of the transcription state.

The annotated Matlab codes for these files are available on our website:
http://flydev.berkeley.edu/cgi-bin/labpage/Levine_Lab/Resources.html

Chapter 3:

3.1 Chapter Overview

In this chapter I describe efforts to identify and test the function of more shadow enhancers, this time in a different patterning system – the "gap gene" patterning system critical for proper anterior-posterior patterning and segmentation of the early fruit fly embryo. This work focuses primarily on three of the four major gap genes: *hunchback*, *Kruppel*, and *knirps*. Further BAC experiments show that both enhancers are necessary for complete patterns of gene expression for *hunchback*, and a complementary set of experiments with plasmids showed a similar trend for *Kruppel* and *knirps*. In addition, we describe finding some key differences in these pairs of enhancers from what was described previously for *snail*. Unlike in the case of *snail*, results show that all shadow enhancers are not entirely equivalent to their counterparts – especially when it comes to dominant, long-range repression. While these enhancers seem to do a similar job in ensuring complete patterns of gene activation across a domain, they also sometimes require the action of a remote dominant repressor in their counterpart enhancer in order to set correct limits of gene expression. We published this work in PNAS in 2011 (Perry *et al.* 2011).

3.2 Summary

Segmentation of the *Drosophila* embryo begins with the establishment of spatially restricted gap gene expression patterns in response to broad gradients of maternal transcription factors, such as Bicoid. Numerous studies have documented the fidelity of these expression patterns, even when embryos are subjected to genetic or environmental stress, but the underlying mechanisms for this transcriptional precision are uncertain. Here we present evidence that every gap gene contains multiple enhancers with overlapping activities to produce authentic patterns of gene expression. For example, a newly identified *hunchback* (*hb*) enhancer (located 5 kb upstream of the classical enhancer) ensures repression at the anterior pole. The combination of intronic and 5' *knirps* (*kni*) enhancers produces a faithful expression pattern, even though the intronic enhancer alone directs an abnormally broad expression pattern. We present different models for

"enhancer synergy" whereby two enhancers with overlapping activities produce authentic patterns of gene expression.

3.3 Introduction

Recent studies identified "shadow" enhancers for genes engaged in the dorsal-ventral patterning of the early *Drosophila* embryo (Hong *et al.* 2008a). These enhancers are sometimes located within neighboring genes, and along with conventional, proximal enhancers, they produce robust patterns of gene expression in early embryos under stress (Frankel *et al.* 2010; Perry *et al.* 2010). For example, the *snail* gene exhibits erratic patterns of activation in embryos raised at 30°C when either the proximal or shadow enhancer is removed (Perry *et al.* 2010). It was proposed that shadow enhancers represent a mechanism of "canalization" (Waddington 1942), whereby populations of embryos develop normally even when subject to variations in temperature or genetic background.

In the present study we provide evidence that many of the genes controlling anterior-posterior patterning contain multiple enhancers with overlapping activities, including head patterning genes and gap genes, which initiate the segmentation gene network (e.g. Clyde *et al.* 2003; Jaeger *et al.* 2004). For example, the gap genes *hunchback* (*hb*), *Kruppel* (*Kr*), and *knirps* (*kni*) are each regulated by two distinct enhancers that control the initial bands of gene expression within the presumptive head, thorax, and abdomen. Evidence is presented that the two enhancers work together (enhancer synergy) to ensure uniform expression within correct spatial limits.

Previous studies have documented examples of enhancer autonomy and enhancer interference. Multiple enhancers often produce additive patterns of gene expression, as seen for the 7-stripe *even skipped* (*eve)* expression pattern arising from five separate enhancers (two located 5' of the *eve* transcription unit and three located downstream of the gene) (Small *et al.* 1996; Small *et al.* 1992; Fujioka *et al.* 1999). Sometimes, multiple enhancers interfere with one another when placed within a common regulatory region. For example, ventral repressors that delineate the *intermediate neuroblasts defective* (*ind)* expression pattern block the activities of a neighboring *eve* stripe 3 enhancer, and conversely, repressors that establish the posterior limit of the stripe 3 pattern interfere with *ind* (Stathopoulos and Levine 2005).

In the present study, evidence is presented that combining multiple enhancers in a common regulatory region can produce sharper and more homogeneous patterns of gene expression. We discuss potential mechanisms for such "enhancer synergy" and suggest that minimal enhancers producing aberrant patterns of gene expression might nonetheless contribute to authentic expression profiles in the context of their native loci.

3.4 Results and Discussion

3.4.1 Every gap gene contains multiple enhancers for a single gap pattern

Candidate gap enhancers were identified using ChIP-chip data (Li *et al.* 2008). Specifically, clustered binding sites for maternal and gap proteins were identified within 100 kb of every gap gene (see Materials and Methods). This survey identified each of the known enhancers, as well as putative "shadow", a potential distal shadow enhancer was identified for *hb,* located 4.5 kb upstream of the proximal transcription start site (designated "P2" in Margolis *et al.* 1995) and upstream of the later-acting distal promoter (designated "P1") (Fig. 3.1C).

A 400 bp genomic DNA fragment from this newly identified region was attached to a *lacZ* reporter gene and expressed in transgenic embryos (Fig. 3.1B). The resulting *hb/lacZ* fusion gene exhibits localized expression in anterior regions of the embryo similar to that seen for the endogenous gene and "classical" enhancer identified over 20 years ago (Driever *et al.* 1989; Struhl 1989) (Fig. 3.1B; compare with A). The classical proximal and distal shadow enhancers exhibit similar responses to increasing Bicoid copy number (Fig. 3.2A).

ChIP-chip data also identified potential pairs of enhancers for *Kr* (Fig. 3.1D-F) and *kni* (Fig. 3.1 G-I). There are two distinct clusters of transcription factor binding sites upstream of *Kr.* The previously identified *Kr* "CD2" enhancer contains the proximal enhancer but also part of the distal binding cluster (Hoch *et al.* 1990). Subsequent *lacZ* fusion assays identified each ChIP-chip peak and underlying binding sites as separable proximal and distal enhancers (Fig. 3.1D-F). Similarly, more refined limits were determined for

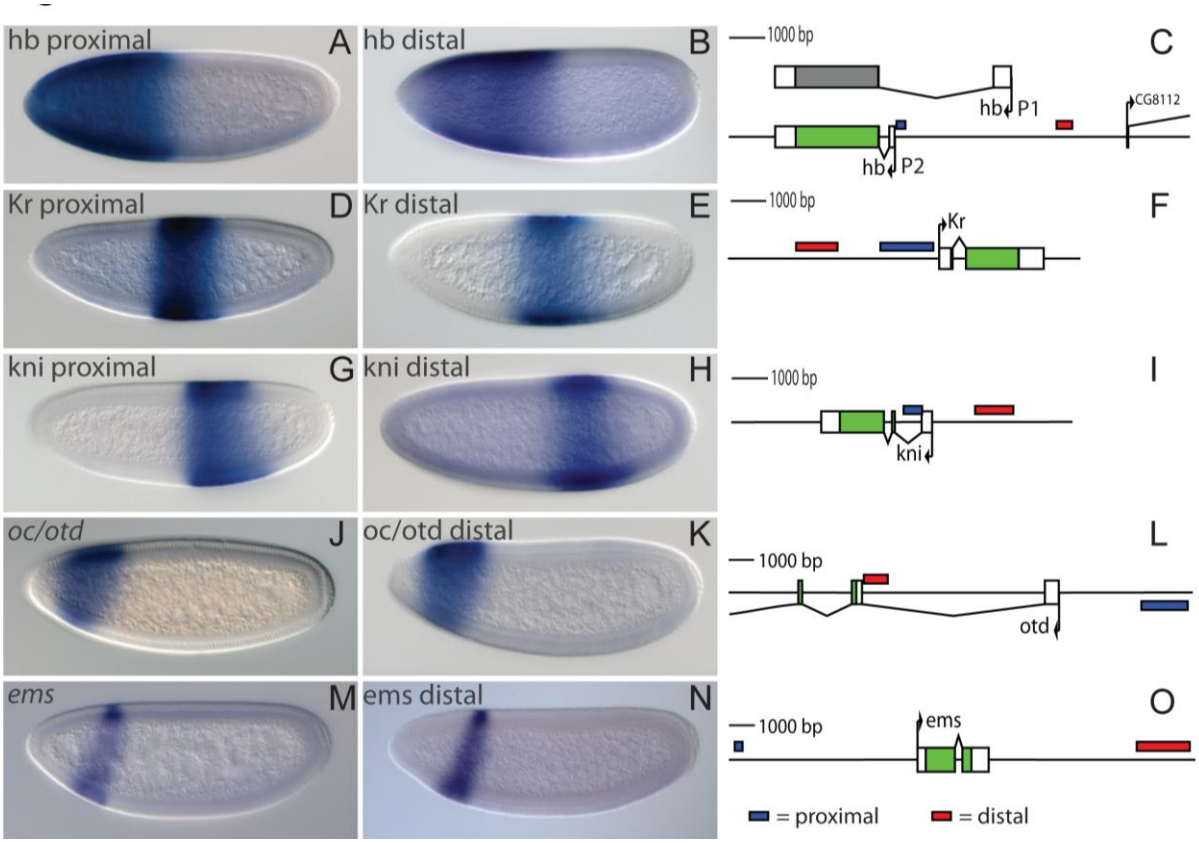


Figure 3.1 Activities of gap enhancers identified by in situ hybridization. (A-B) *hb/lacZ* transgenes containing the (A) proximal (classical) or (B) newly identified distal enhancer (B). The locations of these enhancers are shown in (C). (D-E) *Kr/lacZ* transgenes containing the (D) proximal or (E) distal enhancer. The locations of these enhancers are shown in (F). (G-H) *kni*/*lacZ* transgenes containing either the (G) proximal intronic enhancer or (H) the distal 5' enhancer (H). The locations of these enhancers are shown in (I). (J-K) Expression of endogenous *oc/otd* (J); *oc/lacZ* transgene containing an intronic enhancer (K). The locations of the *oc/otd* enhancers are shown in (L). Expression of endogenous *ems* (M); *ems/lacZ* transgene containing a distal enhancer (N). Locations of the enhancers shown in (O).

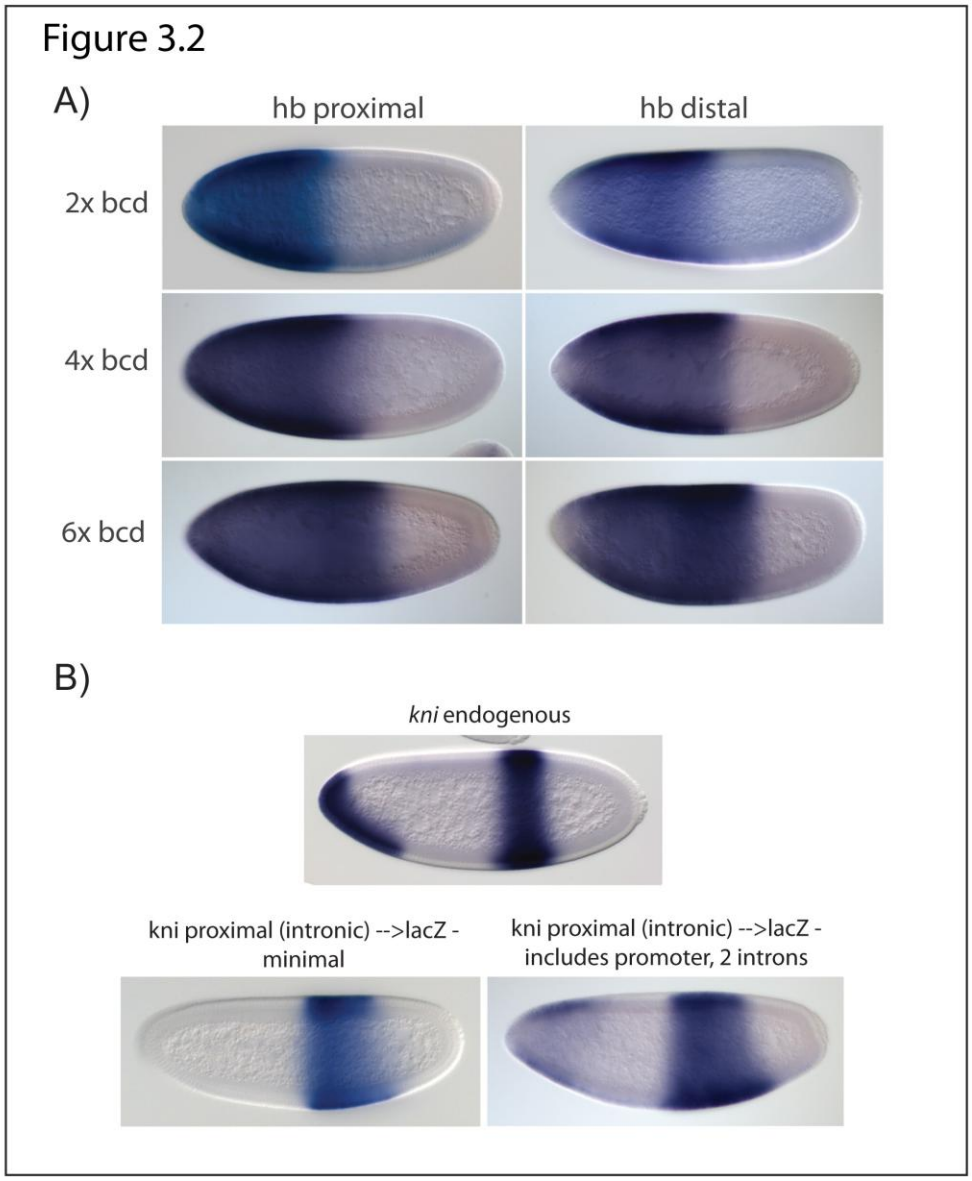


Figure 3.2 Response of the *hb* proximal and distal enhancer to varying Bcd concentration, and an expanded *knirps* construct.
(A) The *hb* primary and shadow enhancer driving a *lacZ* reporter were crossed into multicopy Bicoid backgrounds. The two enhancers show a similar response. Also note the lack of anterior terminal repression for the *hb* primary enhancer, present in the more distal enhancer. (B) Inclusion of a larger *kni* minimal proximal fragment does not change the ectopic expansion of expression, nor does inclusion of the endogenous promoter region.

the *kni* intronic enhancer (Fig. 3.1G,I; Fig. 3.2B; (Schroeder *et al.* 2004)), in addition to the previously identified 5' distal enhancer (Pankratz *et al.* 1992). Both the distal *Kr* enhancer and the intronic *kni* enhancer produce somewhat broader patterns of expression than the endogenous gene (Fig. 3.1E,G; see below). Additional gap enhancers were also identified for *giant*, including an additional distal enhancer located ~35 kb downstream within a neighboring gene (Fig. 3.3A and (Schroeder *et al.* 2004)).

The survey of gap and maternal binding clusters was extended to include the so- called "head" and "terminal" gap genes, critical for the differentiation of head structures and the non-segmented termini of early embryos (Fig. 3.1J-O; Fig. 3.3B-J). Additional enhancers were identified for *empty-spiracles* (*ems*) (original enhancer identified in Hartmann *et al.* 2001) (Fig. 3.1M-O), *huckebein* (*hkb*) (original enhancer in Häder *et al.* 2000) (Fig. 3.3E-G), and *forkhead* (*fkh*) (original enhancer in Schroeder *et al.* 2004) (Fig. 3.3B-D). More refined limits were also determined for the previously identified *ocelliless*/*orthodenticle (oc*/*otd*) intronic enhancer (Schroeder *et al.* 2004) (Fig 3.1J-L). For simplicity, we will hereafter refer to the two enhancers regulating a given gap gene as proximal and distal, based on their relative locations to the transcription start site.

3.4.2 Multiple *hb* Enhancers Produce Authentic Expression Patterns

BAC recombineering (Venken *et al.* 2006; Venken *et al.* 2009), phiC31 targeted genome integration (Groth *et al.* 2004; Bischof *et al.* 2007), and quantitative *in situ* hybridization assays (Perry *et al.* 2010; Boettiger and Levine 2009) were used to determine the contributions of the proximal and distal enhancers to the *hb* expression pattern (Fig. 3.4). BACs containing ~20 kb of genomic DNA encompassing the *hb* gene and flanking sequences were integrated into the same position in the *Drosophila* genome. The *hb* transcription unit was replaced with the *yellow* gene, which permits quantitative detection of nascent transcripts using an intronic hybridization probe (see Perry *et al.* 2010; Materials and Methods and Fig. 3.5). The modified BAC retains the complete *hb* 5' and 3' UTRs. Additional BACs were created by inactivating the proximal or distal enhancers by substituting critical regulatory elements with "random" DNA sequences (see diagrams above panels in Fig. 3.3A-C and Materials and Methods).

BAC transgenes lacking either the distal (Fig. 3.4A) or proximal (Fig. 3.4B) enhancer continue to produce localized patterns of transcription in anterior

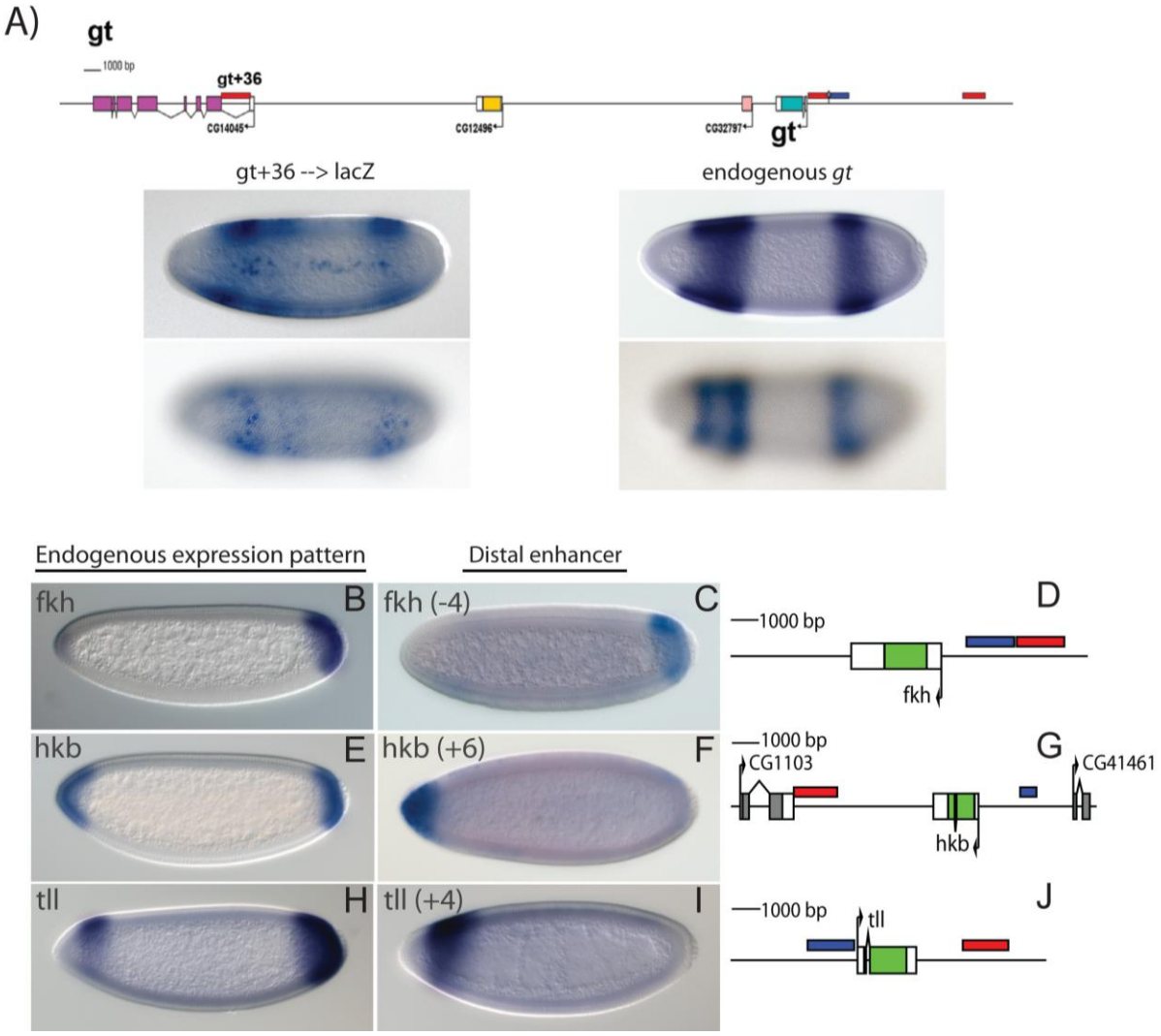


Figure 3.3 The gap gene *gt* has an even more distal additional enhancer; examples of "terminal" gap genes also have shadow enhancers. *gt* has additional enhancer in the intron of a neighboring gene ~36kb downstream. CG14045 is expressed constitutively at low levels, as are the intervening CG genes (data not shown). Patchy, stochastic expression was observed for this additional (3rd) enhancer for anterior and posterior patterns. Examples of terminal gap genes also have shadow enhancers; endogenous expression patterns for these genes are shown on the left. New enhancers were identified for *fkh* (B-D), and *hkb* (E-G). An original anterior enhancer for *tll* was identified in Rudolph *et al.* 1997; an additional enhancer for the anterior pattern of *tll* is shown in (J) (Ochoa-Espinoza *et al.* 2005, line provided by Steven Small). Schematics showing the locations of these enhancers are on the right.

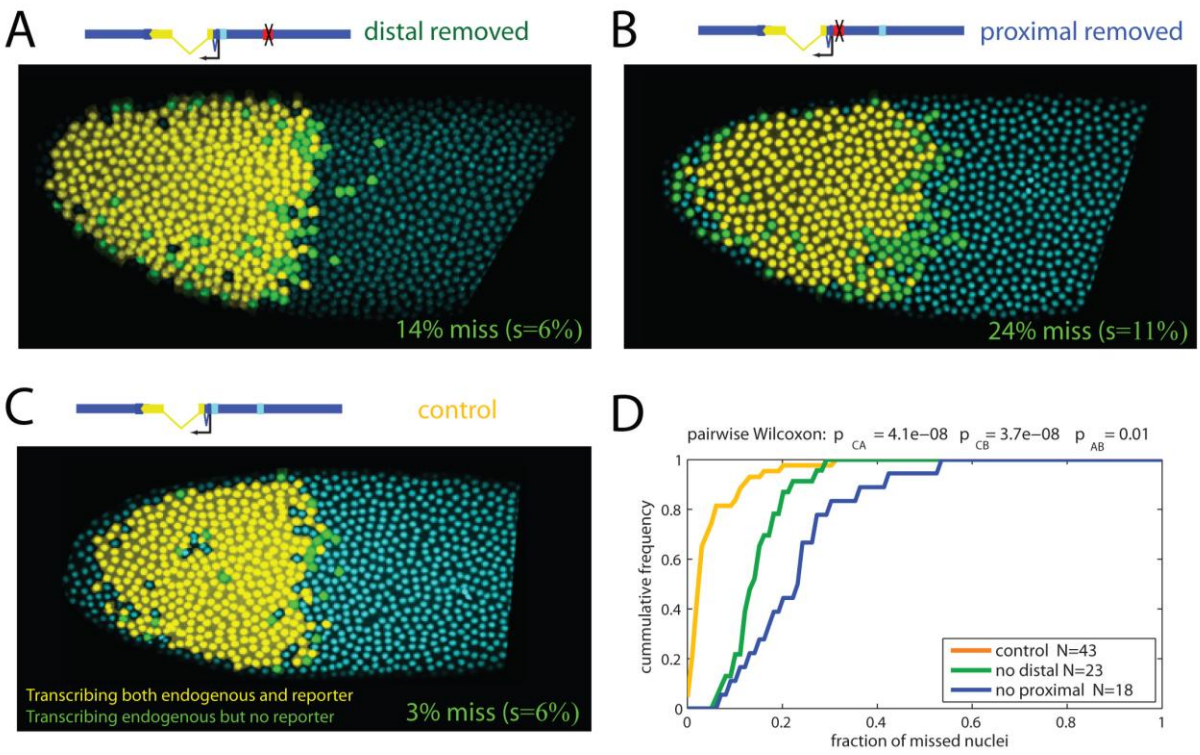


Figure 3. 4 Function of *hb* enhancers via BAC transgenesis. An ~20 kb BAC containing genomic DNA encompassing the *hb* locus was modified to remove specific proximal or distal enhancers. The *hb* transcription unit was replaced with the *yellow* reporter gene in order to identify *de novo* transcripts via *in situ* hybridization using a probe directed against the *yellow* intron (see diagram above images in panels A-C). (A) *hb*-BAC with distal enhancer inactivated (see X in diagram), (B) *hb*-BAC with proximal enhancer inactivated, and (C) *hb*-BAC with both enhancers intact. The median ratio of "discordance" is indicated beneath each image. This is the fraction of nuclei that express endogenous *hb* nascent transcripts, but not *yellow* transcripts. (D) Cumulative frequency distributions for the fraction of 'missing nuclei' in the three populations of embryos. The ordinate axis gives the probability of observing an embryo from this population with fewer than the abscissa fraction of nuclei transcribing the endogenous gene but not the reporter. Statistical comparisons between the distributions are presented above the panel, with subscripts matching the panel labels (i.e. pCA is the p value from the pair wise comparison of the distribution of embryos with the *hb* control BAC, (C), to those with the distal enhancer removed (A)).

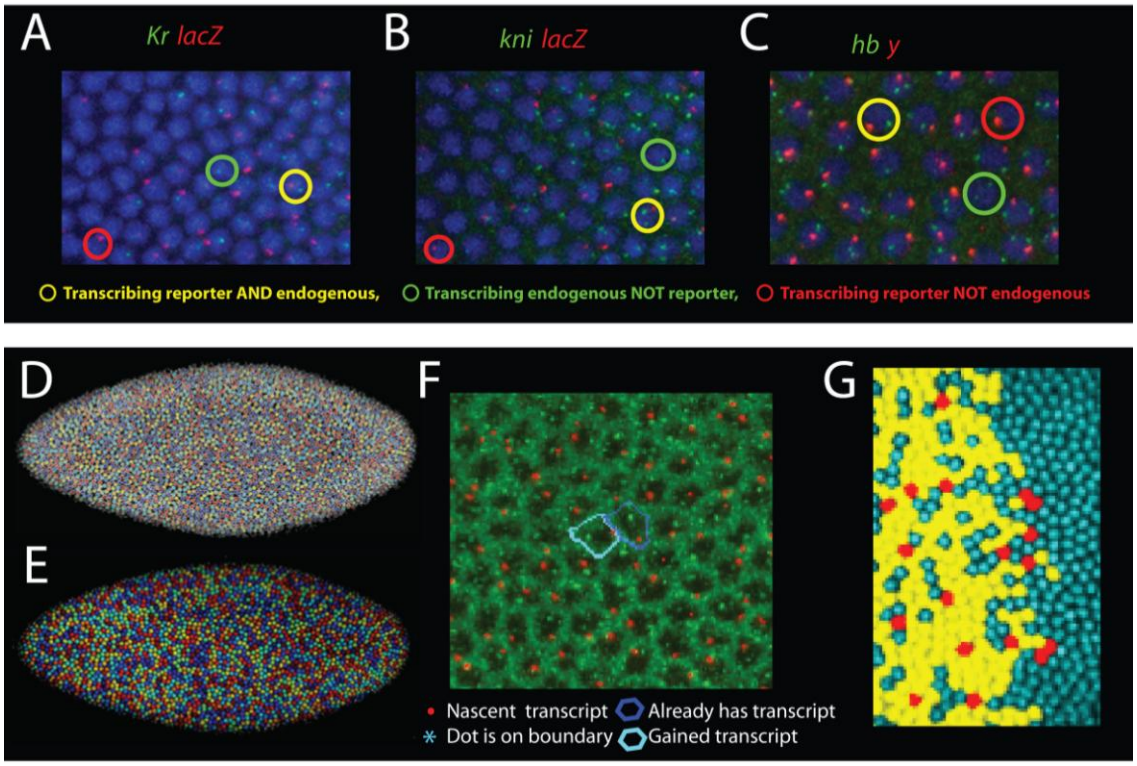


Figure 3.5 Examples from the image analysis process. (A) Nascent transcription foci detected for endogenous *Kr* (green channel) and for the *lacZ* transgene (red channel) as detected by fluorescent *in situ* using *Kr* full length mRNA probe and *LacZ* full length mRNA probe. Nuclei are counterstained in blue. Double staining allows detection of nuclei which transcribe both loci (yellow circles), just the endogenous gene (green circles), or just the reporter gene (red circles). (B) as in (A), with endogenous *kni*-full length probe. (C) as in (B), with intronic *yellow*-reporter in the red channel and endogenous full length *hb* probe in green. Since the *hb* embryos are imaged in cell cycle 13 instead of 14, the nuclei are also less dense. (D) Computational image processing algorithm segments all the different nuclei (representing unique nuclei by separate, randomly chosen colors). (E) These masks in (D) are then expanded in a space filling way so every pixel of the embryo is assigned to a nucleus. (F) The individual transcripts are segmented by a filter (red dots), and assigned to the containing nucleus (blue polygons). An extra filtering step resolves dots which lie on the border between two cells by comparing the presence of dots in neighboring cells. (G) The results are plotted by assigning the cell mask computed in (E) to a color code which represents its composite transcriptional state (i.e. transcribing reporter AND endogenous vs transcribing reporter NOT endogenous).

regions of transgenic embryos in response to the Bicoid gradient. However, the patterns are not as faithful as compared with the BAC transgene containing both enhancers (Fig. 3.4C). Embryos were double-labeled to detect both *yellow* and *hb* nascent transcripts. (cc) 13, a substantial fraction of nuclei (14%) expressing *hb* nascent transcripts lack *yellow* transcription upon removal of the shadow enhancer (Fig. 3.4A). An even higher fraction of nuclei (24%) lack *yellow* transcription when the proximal enhancer is removed (Fig. 3.4B). Control transgenic embryos containing both enhancers exhibit more uniform patterns of transcription, whereby only an average of ~3% of nuclei fail to match the endogenous pattern of transcription (Fig. 3.4C). The distribution of "missing nuclei" across the population of cc13 embryos is plotted in Fig 3.4D.

The pairwise Wilcoxon rank sum test (also called the Mann-Whitney U-test) was used to determine the significance of the apparent variation in gene expression resulting from the removal of either the proximal or distal enhancer (Fig. 3.4D). Control embryos containing the *hb* BAC transgene with both enhancers exhibit some variation in the number of nuclei that lack *yellow* nascent transcripts. Despite this variation, the statistical analyses indicate that the loss of either the proximal or distal enhancer results in a significant increase in the variability of *yellow* transcription patterns as compared with the control BAC transgene (p=4E-8).

3.4.3 The distal hb enhancer mediates dominant repression

The preceding analyses suggest that multiple enhancers produce more uniform patterns of *de novo* transcription than individual proximal or distal enhancers. Additional studies were done to determine whether multiple enhancers also help produce authentic spatial limits of transcription (Fig. 3.6).

hb expression normally diminishes at the anterior pole of cc13-14 embryos. This loss in expression has been attributed to attenuation of Bcd activity by Torso RTK signaling (e.g., Kim *et al.* 2010). However, the proximal enhancer fails to recapitulate this loss (Struhl *et al.* 1992) (Fig. 3.1A). In contrast, the distal enhancer is inactive at the anterior pole (Fig. 3.1B), and the two enhancers together produce a pattern that is similar to endogenous expression, including reduced expression at the pole (Fig. 3.4C).

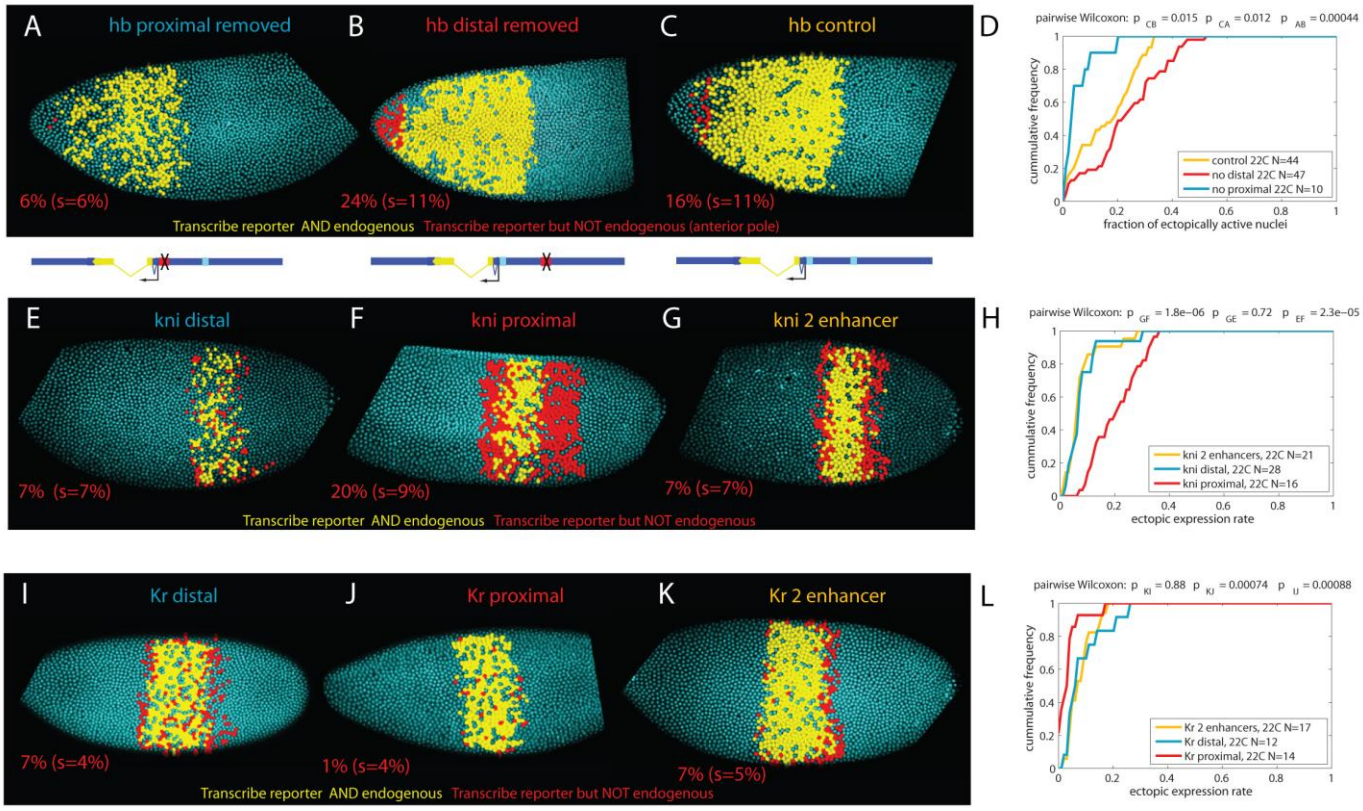


Figure 3.6 Enhancer synergy produces correct spatial limits.
Discordance of *yellow* (A-C) or *lacZ* (E-G, I-K) transgenes and endogenous gap gene nascent transcripts. Nuclei exhibiting ectopic transgene expression are indicated in red. Sites of concordant expression are indicated in yellow.
(A-C) BAC transgenes lacking the proximal (A) or distal (B) enhancer, or containing both enhancers intact (C). Nuclei in the anterior third of the *hb*-expression region which transcribe the reporter but not endogenous *hb* are shown in red. (D) Cumulative frequency of nuclei in the anterior third of the *hb*-expression domain containing *yellow*, but not endogenous *hb*, nascent transcripts. Median and standard deviations are shown on the corresponding panels. (E-F) *kni/lacZ* reporter genes driven by (E) distal enhancer, (F) proximal enhancer or (G) both enhancers. Median fractions of nuclei transcribing *lacZ* but not the endogenous gene are indicated below each image. (H) Cumulative frequency distributions for the fraction of ectopically active nuclei. (I-L) Similar analysis of *Kr*/*lacZ* transgenes containing the distal (I), proximal (J), or both (K) enhancers.

To examine the relative contributions of the proximal and distal enhancers in this repression, *yellow* nascent transcripts were measured in transgenic embryos expressing BAC reporter genes containing one or both *hb* enhancers (see Fig. Particular efforts focused on the early phases of cc14, when repression of endogenous *hb* transcripts is clearly evident (Fig. 3.6). transgene lacking the proximal, classical enhancer, but containing the newly identified distal enhancer, a median of 6% (std 6%) of nuclei exhibit expression of *yellow* nascent transcripts but lack expression of the endogenous gene (Fig 3.6A). In contrast, a median of 24% (std 11%) of nuclei displays a similar discordance upon removal of the distal enhancer. In control embryos, 16% (std 11%) of nuclei express *yellow*, but lack *hb*, nascent transcripts (Fig 3.6B-C). It should be noted that the BAC transgene lacking the proximal enhancer exhibits "super-repression" due to reduced activation at the anterior pole ($p = 0.012$) (Fig 3.6D).

These observations suggest that the distal enhancer contains repression elements that function in a dominant manner to attenuate the activities of the proximal enhancer at the anterior pole. There is a loss of repression when the distal enhancer driving *lacZ* is crossed into a *torso* mutant background (Fig. 3.7). This observation implicates one or more repressors functioning downstream of Torso signaling, including Tailless and Huckebein, which have been shown to function as long-range dominant repressors (Courey and Jia 2001). Indeed, whole-genome ChIP assays identify more potential Tll and Hkb binding sites in the distal vs. proximal enhancer (Li *et al.* 2008) (Fig 3.8A). The persistence of *hb* expression in anterior regions has been shown to be detrimental, causing defects in mouth parts and malformation of the gut (Janody *et al.* 2000).

3.4.4 Correction of the *kni* expression pattern

Kr/lacZ and *kni/lacZ* fusion genes containing either one or two enhancers were inserted into the same position in the *Drosophila* genome (Fig. 3.6). Transgenic embryos were double-labeled to detect the expression of the transgene (*lacZ*) as well as the endogenous gap gene.

The *kni* proximal (intronic) enhancer alone produces an abnormally broad pattern of expression, especially in posterior regions (Fig. 3.6F; see Fig. 3.1G and Schroeder *et al.* 2004). In contrast, the *kni* distal (5') enhancer produces erratic *lacZ* activation within nearly normal spatial limits (Fig. 3.6E). An essentially normal pattern of *lacZ* transcription is observed when both

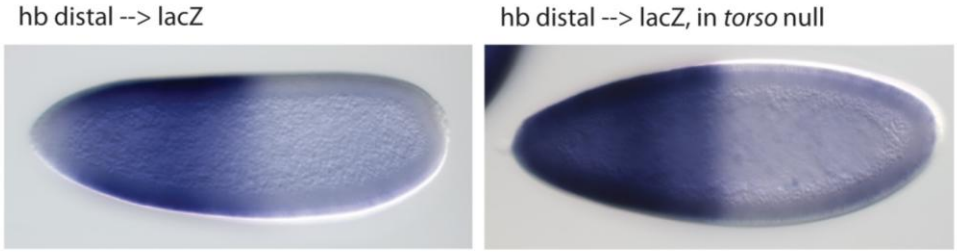
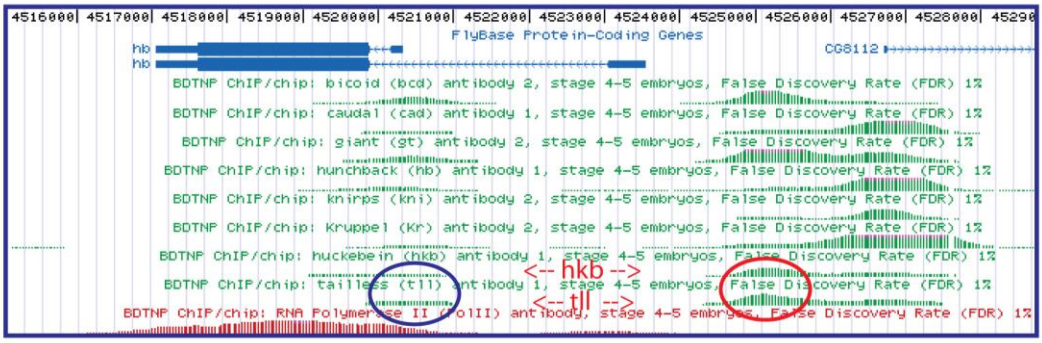


Fig 3.7 Repression of *hb* in anterior regions is mediated by *torso* signaling. The hb distal enhancer is shown in wild type (left) and *torso* null backgrounds (right).

A)



B)

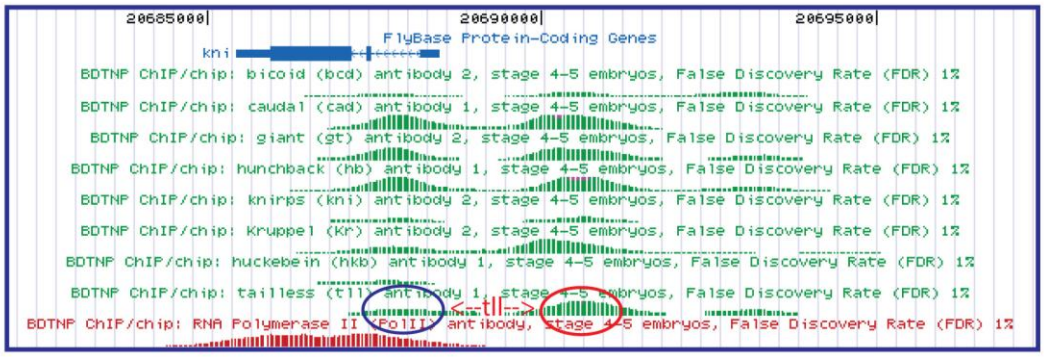


Fig 3.8 ChIP-chip data for dominant repressors. Gap gene ChIP-chip data from the Berkeley Drosophila Transcription Network Project (13) shows greater relative binding of putative dominant, long-range repressors at the more distal enhancer for both *hb* and *kni*. A) Gap gene binding near *hb* shows increased binding of both Hkb and Tll at the location of the distal enhancer, shown circled in red (proximal enhancer circled in blue). B) Gap gene binding near kni shows increased relative binding of Tll at the more distal enhancer (circled in red).

enhancers are combined in a common transgene (intronic enhancer 5' and distal enhancer 3' of *lacZ*; Fig. 3.6G). It appears that *lacZ* transcription is slightly broader than the endogenous pattern, but considerably narrower than the pattern observed for the intronic enhancer alone (Fig. 3.6J) (p = 1.8E-6), and not statistically different from the expression limits of the distal enhancer alone (p = 0.72) (Fig. 3.6L). There is no significant narrowing of the *Kr/lacZ* expression pattern when both the distal and proximal enhancers are combined within the same transgene (Fig. 3.6K,L) (p = 1.0). Perhaps additional *Kr* regulatory elements are required for the type of narrowing observed for the *kni* intronic enhancer. Alternately, all of these transgenes use the *eve* basal promoter and it is possible that promoter-specific interactions are important for establishing the normal limits of the *Kr* expression pattern.

As discussed earlier, long-range repressors bound to the distal *hb* enhancer might inhibit the activities of the proximal enhancer at the anterior pole of precellular embryos. The distal *kni* enhancer might function in a similar manner to sharpen the expression limits of the intronic enhancer. The spatial limits of gap gene expression patterns have been shown to depend on cross-repressive interactions (e.g. Kraut and Levine 1991; Manu *et al.* 2009a; Manu *et al.* 2009b). The *kni* intronic enhancer might lack critical gap repression elements since it produces an abnormally broad expression pattern. Indeed, whole-genome ChIP assays identify more putative Tailless binding sites in the distal vs. intronic enhancer (Li *et al.* 2008) (Fig 3.8B). These Tailless repression elements might function in a dominant fashion to restrict the limits of the intronic enhancer.

The modest anterior expansion of the expression pattern driven by the *kni* intronic enhancer is more difficult to explain since this boundary is probably formed by the Hb repressor (Yu and Small 2008), which is not known to function in a long-range and dominant manner. If the action of short-range repressors is also affected by stochastic processes (e.g. binding of the repressor to enhancer or looping of a bound enhancer to promoter), perhaps having two enhancers might improve the chances of maintaining proper repression.

We have presented evidence that the robust and tightly defined patterns of gap gene expression do not arise from the unique action of individual enhancers. Rather, these patterns depend on multiple and separable enhancers with similar, but slightly distinct regulatory activities. This

enhancer synergy produces more homogeneous patterns of transcriptional activity, as well as more faithful spatial limits of expression.

The enhancer synergy documented in this study is somewhat distinct from the proposed role of the shadow enhancer regulating *snail* expression in the presumptive mesoderm (Perry *et al.* 2010). The dual regulation of *snail* by the proximal and distal (shadow) enhancers was shown to ensure homogenous and reproducible expression in embryo after embryo in large populations of embryos even when they are subject to increases in temperature. In contrast, dual regulation of *hb* expression by proximal and distal enhancers appears to ensure homogenous activation in response to limiting amounts of the Bicoid gradient. They are used as an obligatory patterning mechanism rather than for buffering environmental changes. Despite these apparent differences, it is possible that dominant repression is also used as a mechanism of synergy for the regulation of *snail* expression. The distal enhancer contains repressor elements (e.g., Huckebein) that inhibit the expression of the proximal enhancer at the termini (Perry *et al.* 2010).

Different mechanisms can be envisioned to account for enhancer synergy. Perhaps the simplest is that there are fewer inactive nuclei within a given gap expression domain due to the diminished failure rate of successful enhancer-promoter interactions with two enhancers rather than one. If the rates at which enhancers fail to activate transcription are completely independent then one would expect the combined action of two enhancers to yield a multiplicative reduction in how often a given cell fails to express the gene within a given window of time. This sort of synergy does not require any direct physical or cooperative interactions between the enhancers. Nonetheless the effect can be significant (as seen for *hb*). For example, two enhancers each with a 10% uncorrelated failure rate may together be expected to have a 1% failure rate, a ten-fold reduction (see Fig 3.9A and B). For genes that produce strong bursts of mRNA expression, this change in frequency of transcription may have a dramatic effect on the variation of total mRNA levels.

A second but critical potential mechanism of enhancer synergy concerns long-range, dominant repression. Repressors (such as Tailless) bound to one enhancer are sufficient to restrict the spatial limits of the other enhancer. There is no need for long-range repressor elements to appear in both enhancers in order to achieve normal spatial limits of gene expression. It has

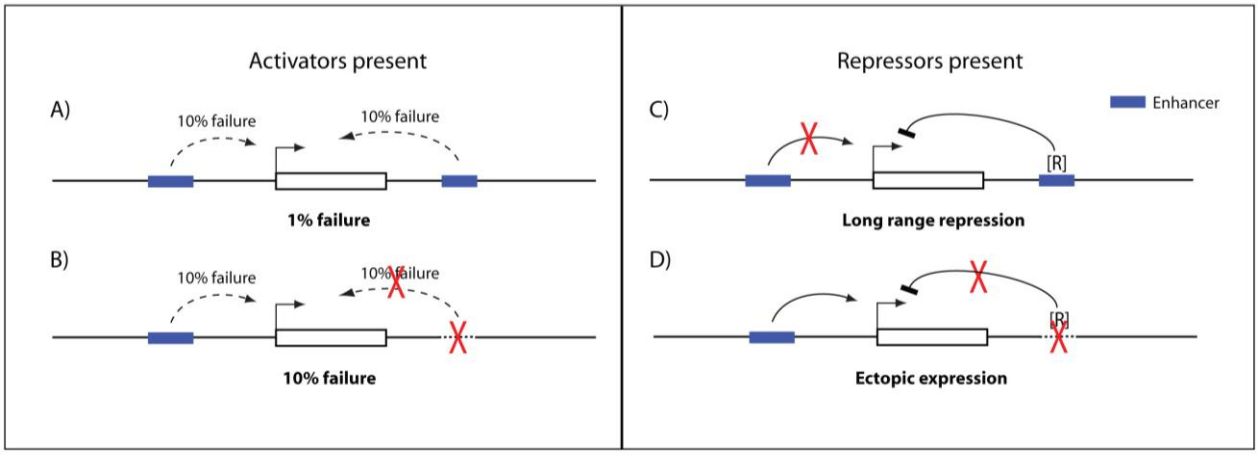


Fig 3.9 Models for enhancer synergy. (A-B) Activation of one promoter by two enhancers. If two independent enhancers each have a 10% failure rate in activating expression and transcription factor binding or enhancer looping is rate limiting, then the two enhancers should have a combined failure rate of 10% x 10% = 1% (A). Removing one enhancer increases the failure rate to 10% (B). (C-D) The binding of a long range, "dominant" repressor to one enhancer is sufficient to inactivate the other (C). Removal of this enhancer results in "ectopic" expression (D).

been suggested that long-range repressors, such as Hairy, mediate the assembly of positioned nucleosomes at the core promoter. Such repressive nucleosomes should block productive enhancer-promoter interactions, even for enhancers lacking repressor sites (Fig 3.9C and D).

Regardless of the detailed molecular mechanisms, the combined action of multiple enhancers helps explain why an individual enhancer sometimes fails to recapitulate an authentic expression pattern when taken from its native context. Enhancers that produce abnormal patterns of expression (e.g., *kni* intronic enhancer) can nonetheless contribute to homogeneous and robust patterns of gene expression in conjunction with the additional enhancers contained within the endogenous locus.

3.5 Materials and methods

3.5.1 Enhancer identification and testing

Prospective enhancers were identified near genes of interest using a combination of ChIP-chip data (provided for various maternal, gap, and pair rule genes by the Berkeley *Drosophila* Transcription Network Project, (Li *et al.* 2008) and sequence-based binding site cluster analysis. The cluster analysis was performed using the software ClusterDraw2 (Papatsenko 2007). The program and binding motif models used are available online at http://line.bioinfolab.net/webgate/submit.cgi.

Candidate regions (listed in Appendix I, Table 2) were tested *in vivo* using traditional *lacZ* reporter assays combined with targeted phiC31 transgenesis as adapted for use in *Drosophila* (Groth *et al.* 2004; Bischof *et al.* 2007). An nE2G backbone (Markstein *et al.* 2004) modified for targeted integration was used to test potential enhancers by placing them upstream of an *eve-lacZ* fusion gene. The same construct was used for the one vs. two enhancer experiments for *Kr* and *kni*; the second enhancer for the two enhancer constructs was added into a BstBI restriction site downstream of *lacZ* ~5kb away from the first enhancer. The landing site 51D (Bischof *et al.* 2007), Bloomington Stock Center number 24483, was used for *lacZ* assays, in part due to its lack of the "head stripe" artifact found in many other landing site lines when combined with enhancer-testing plasmids. Because of our interest in AP patterns (some of which are stripes in anterior regions), other lines were avoided in favor of this landing site. Unfortunately, constructs integrated into this landing site often have faint mesodermal background

instead, making them unsuitable for use with DV enhancers. A preferable landing site was later determined to be "VK33" (site described in Venken *et al.* 2009); it also seems to lack the head-stripe artifact for most integrated fragments yet seems otherwise clean of artifactual expression.

The two *hb* enhancer-*lacZ* constructs were crossed into a 4x or 6x maternal Bicoid copy number background using the BB9+16 fly line (Struhl 1989).

3.5.2 Recombineering and transgenesis

Recombineering was performed as described previously (Perry *et al.* 2010)(see also Venken *et al.* 2006; Venken *et al.* 2009; Lee *et al.* 2001; Liu *et al.* 2003). The *yellow* reporter (used to detect sites of nascent transcript by using an intronic *in situ* probe) was integrated as a *yellow-kanamycin* fusion that left the native *hb* UTRs intact. The *bcd* binding site clusters and surrounding regions of the primary or shadow enhancers were removed via replacement with an ampicillin resistance cassette taken from pBlueScript. Primers used for construct building and recombineering are listed in Appendix I, Table 2. BAC CH322-55J23 (Venken *et al.* 2009) was the basis for all subsequent modifications. All BACs were integrated into landing site VK37 on chromosome 2 (Venken *et al.* 2006), Bloomington Stock Center number 24872.

3.5.3 Whole-mount *in situ* hybridization

Embryos were fixed using standard methods. Fluorescent or colormetric *in situ* hybridization was performed as described in (Perry *et al.* 2010; Kosman *et al.* 2004). Probes were generated with the primers listed in Appendix I, Table 2, and *in vitro* transcription. Reporter genes were labeled with digoxigenin-tagged antisense probes, sheep anti-dig primary antibodies (Roche), and donkey anti-sheep Alexa 555 secondary antibodies (Invitrogen). Endogenous genes *hb, Kr* and *kni* were labeled with biotin-tagged probes, mouse anti-bio primary antibodies (Roche), and donkey anti-mouse Alexa 488 secondaries (Invitrogen). Nuclei were counterstained with DRAQ5 (Biostatus Ltd.).

3.5.4 Confocal image acquisition and computational image processing

1024x1024 3-color image stacks were acquired using a Leica SL Laser Scanning Confocal microscope as described in (Perry *et al.* 2010). Image

segmentation and analysis was performed as described in (Perry *et al.* 2010), with minor modification. Nuclei were segmented using a Difference of Gaussians filter optimized with size selection (Fig. 3.5D). A space-filling, segment dilation algorithm was used to assign all pixels in the embryo to one of the segmented nuclei, created a final nuclear mask (Fig. 3.5E). All nuclear masks were manually checked to confirm accurate segmentation. Nascent transcripts were localized for both the reporter and the endogenous genes, also using Difference of Gaussians filters, this time optimized to detect the bright transcripts corresponding to sites of transcription, see Fig. 3.5. Intensity thresholds and dot size thresholds reduced spurious counts. Segmentation results were curated by the user. This segmentation enabled the nucleus by nucleus analysis of transcriptional state of reporter and endogenous gene as described in the text and shown in figures 2-3. This analysis scripts were wrapped in a Graphical User Interface implemented through Matlab's software GUI Design Environment (GUIDE). The source code for this analysis is available in the supplemental material of (Perry *et al.* 2010). Updated versions of our image segmentation routines can be found on our Github page for image processing tools: https://github.com/AlistairBoettiger/Image_Analysis. All of the source code used to compute and plot the results from this publication is available on the Github source page for this project.

Chapter 4:

4.1 Chapter Overview

This chapter presents results from a final project that dissects the logic and function of a third *hunchback* enhancer. While this third enhancer is not a shadow enhancer *per se*, it helps integrate the earlier input of the *hb* proximal/shadow pair, is only partially distinct in time and space, and is also involved in the canalization and reliable output of the *hunchback* expression pattern despite variations in input.

The project is largely but not quite entirely complete. Reporter gene expression, evaluation of downstream target genes of Hb, and the phenotypes produced in genetic rescue experiments are properly compared to control construct lines. One comparison remains to be completed – protein expression from modified genetic rescue lines is compared here to endogenous protein expression patterns and not yet to the proper control, an unmodified genetic rescue line. Even with this caveat in mind the data look promising and the function of this enhancer and its role in producing a sharper, less variable, and more precise *hunchback* boundary is clear.

4.2 Chapter Summary

Activation of the gap gene Hunchback by the maternal Bicoid gradient is one of the most intensively studied gene regulatory interactions in animal development. Nonetheless, key issues remain unresolved, including the mechanisms responsible for suppressing variations in the Bicoid gradient to produce a reliable and reproducible border of Hunchback (Hb) expression (canalization). Hb is activated by a proximal enhancer in response to Bicoid, and this expression is reinforced by a recently identified distal "shadow" enhancer. Here we present evidence that a long-neglected "stripe" enhancer is critical for suppressing potential variations in the definitive Hb expression pattern during the time when gap genes establish pair-rule stripes of gene expression. Removal of the central stripe enhancer results in a significant increase in the variability of Hb expression, sometimes causing cuticular defects in the mesothorax (T2) due to abnormal patterns of segmentation gene expression. Characterization of the central stripe enhancer reveals that

it is subject to extensive repression by the gap repressors Kruppel and Knirps, as well as Hb itself. These findings suggest that canalization of the Hb expression pattern is mediated by a dedicated enhancer, separate from those responsible for initiating expression by Bicoid. We discuss the advantages of cross-repression for canalizing gene expression and argue that this helps account for the prevalence of repression in the patterning of the precellular embryo.

4.3 Introduction

Hunchback is the premiere gap gene of the segmentation regulatory network. It is activated in the anterior half of the precellular embryo, within 20-30 min after the establishment of the broad Bicoid regulatory gradient during nuclear cleavage cycles 9 and 10 (~90 min following fertilization) (Driever and Nüsslein-Volhard 1988a, 1988b; Struhl *et al.* 1989; Gregor, *et al.* 2007a). The initial hb transcription pattern exhibits a reasonably sharp on/off border at the future head and thorax (Struhl *et al.* 1989; Driever *et al.* 1989; Treisman *et al.* 1989; Porcher *et al.* 2010; Perry *et al.* 2011). However, by the midpoint of nuclear cleavage cycle 14 (cc14) the Hb protein is distributed in a steep gradient extending into the presumptive thorax and anterior abdomen, where it coordinates the expression of the gap genes *Kruppel* (*Kr*), *knirps* (*kni*), and *giant* (*gt*) (Hulskamp *et al.* 1990; Struhl *et al.* 1992).

The positioning of the Hb border is critical since embryos containing 4 copies of the gene are lethal due to the homeotic transformation of the first abdominal segment (A1) into a duplicated metathorax (T3) (Fig 4.1). Previous studies have demonstrated that the definitive Hb border is highly reproducible from embryo to embryo, despite variations in Bicoid (Houchmandzadeh *et al.* 2002; Gregor *et al.* 2007; He *et al.* 2008). There is at least 2-fold less variance in the distribution of Hb as compared with the input Bicoid gradient. This suppression in Hb variability (canalization) is due to repression by Kr and kni (Manu *et al.* 2009). Here we sought to determine the cis-regulatory basis for this repression.

4.4 Results and Discussion

The *hb* transcription unit is controlled by two promoters, P2 and P1, and three enhancers (Fig. 4.2A) (Driever *et al.* 1989; Margolis *et al.* 1995). The "classical" proximal enhancer (Driever *et al.* 1989; Treisman *et al.* 1989) and distal shadow enhancer (Perry *et al.* 2011) mediate activation in response to

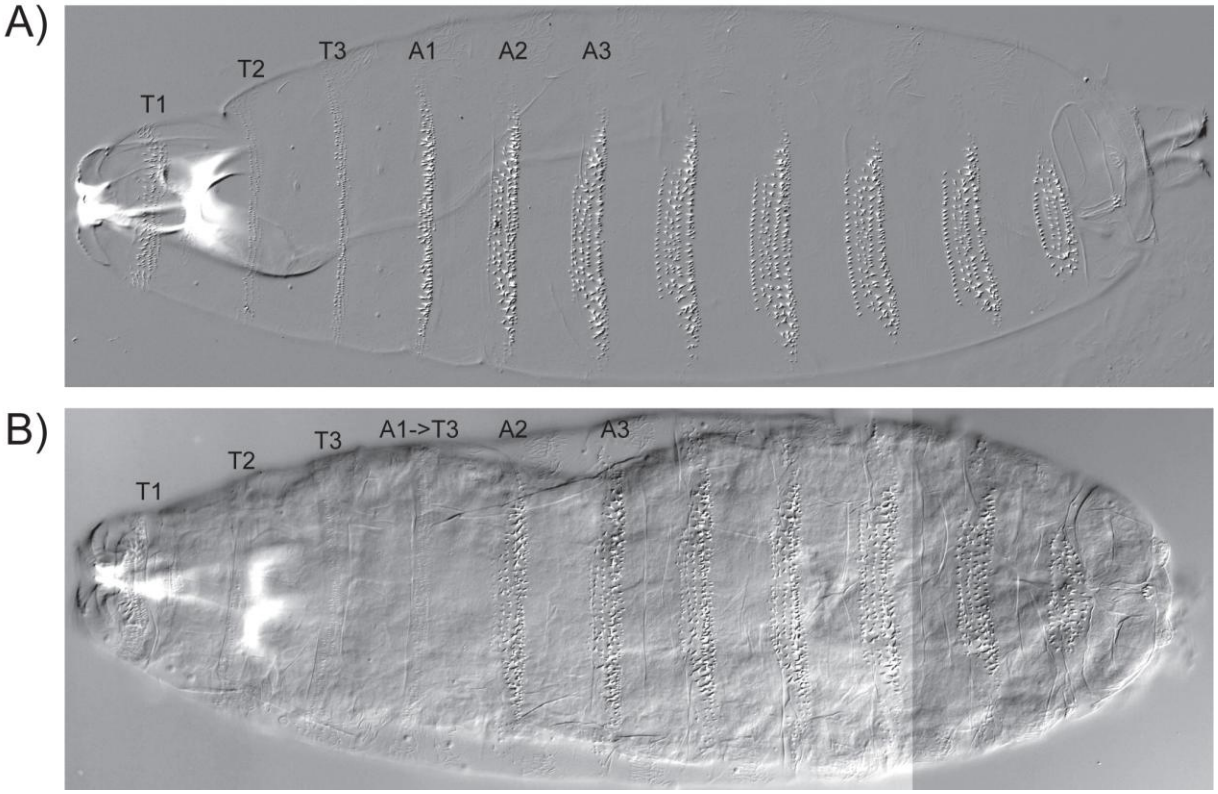


Figure 4.1 Four copies of hunchback causes an A1 to T3 transformation. Presumably, since hb regulates ubx, and expansion of Hb protein downregulates ubx expression in this segment, causing a homeotic transformation.

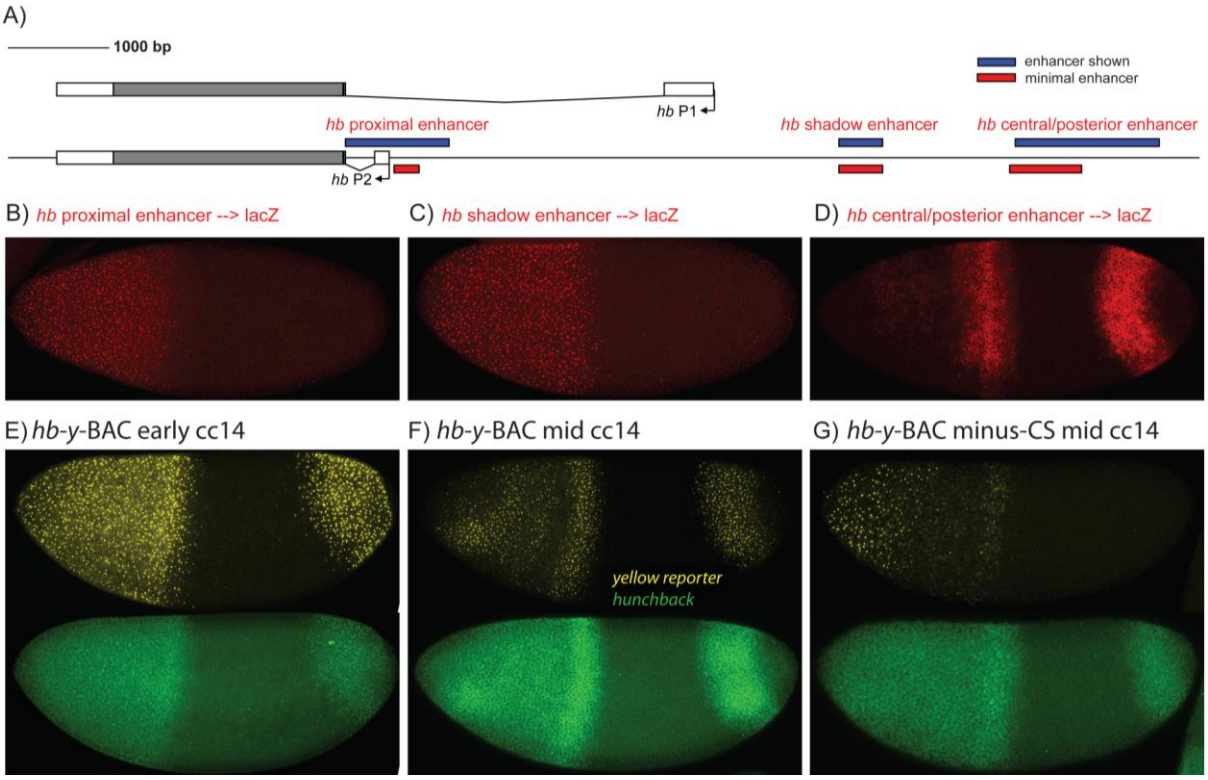


Figure 4.2 Summary of hunchback regulation and the y-hb-BAC reporter. A) Schematic of the hb locus showing the locations of the three enhancers that regulate pre-gastrulation *hb* expression. B)-D) Each of these enhancers driving the lacZ reporter in transgenic embryos. E)-G) Transgenic embryos expressing two forms of the y-hb-BAC; E) and F) are two slightly different stages in early cc14 showing endogenous *hb* expression in green and the *y* reporter in yellow. G) The *y* reporter is not expressed in the posterior stripe and is almost completely gone in the central stripe upon deletion of the stripe enhancer. Residual expression in the central region may come from the hb shadow enhancer, which drives some expression in this region on its own.

the Bicoid gradient. Expression is also regulated by a third enhancer, the central stripe/posterior stripe enhancer, which is located over 5 kb upstream of the P2 start site (Margolis *et al.* 1995). Each of these enhancers was separately attached to a lacZ reporter gene and expressed in transgenic embryos. As shown previously, they direct expression in anterior regions of cc12-13 embryos (Fig. 4.2B,C) (Driever *et al.*, 1989; Struhl *et al.*, 1989; Perry *et al.*, 2011), and in two stripes during cc14 (Fig. 4.2D) (Margolis *et al.* 1995). The hb "stripe" enhancer (Fig. 4.2D) is evocative of the eve stripe 3 + 7 enhancer, which is regulated by the combination of the Hb and Kni repressors (see below) (Small *et al.*, 1996; Yan *et al.*, 1996; Clyde *et al.*, 2003; Struffi *et al.* 2011).

BAC transgenesis was used to determine the contribution of the stripe enhancer to the complex *hb* expression pattern. The *hb* transcription unit was replaced with the *yellow (y)* reporter gene, which contains a large intron permitting detection of nascent transcripts (as in Perry *et al.* 2010). The resulting *y*-BAC mimics the endogenous expression pattern (Fig. 4.2E,F), including augmented expression at the posterior limit of the anterior expression pattern. However, removal of the stripe enhancer from an otherwise intact *y*-BAC transgene leads to a significant reduction in expression at this border and in posterior regions (Fig. 4.2G). The functional impact of removing the stripe enhancer was investigated by genetic complementation assays.

A BAC transgene containing 44 kb of genomic DNA encompassing the entire *hb* locus and flanking regulatory DNAs fully complements deficiency homozygotes carrying a small deletion that cleanly removes the *hb* transcription unit (see Fig 4.3). The resulting adults are fully viable and fertile and indistinguishable from normal strains. Embryos obtained from these adults exhibit a fully normal Hb protein gradient, including a sharp border at the presumptive head/thorax boundary, and a steep decline in the distribution of the protein in the presumptive thorax (Fig. 4.4B). By mid-cc14 a normal eve expression pattern is observed, with the sharp Hb border located within the interstripe separating *eve* stripes 2 and 3.

The Hb BAC transgene lacking the stripe enhancer fails to complement hb- embryos, due to the absence of the posterior hb expression pattern (Fig. 2A) and fusion of the 7th and 8th abdominal segments (Fig. 4.4E) (see Margolis *et al.* 1995). In addition, the anterior Hb expression pattern lacks the sharp "stripe" at its posterior limit, and there is a consistent anterior expansion of

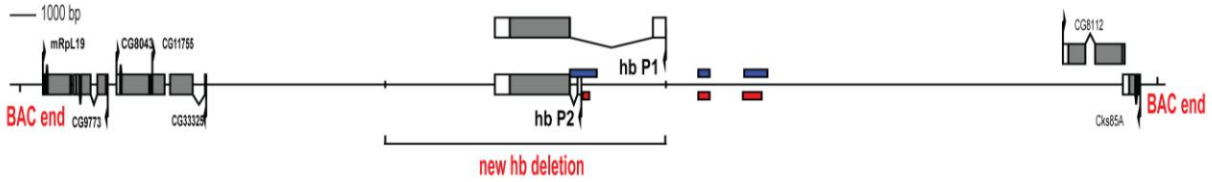


Figure 4.3 44kb hb-BAC region used for genetic rescue and fresh hb deletion. The deletion was generated using FRT mediated recombination to remove a tight region around the hb coding sequence.

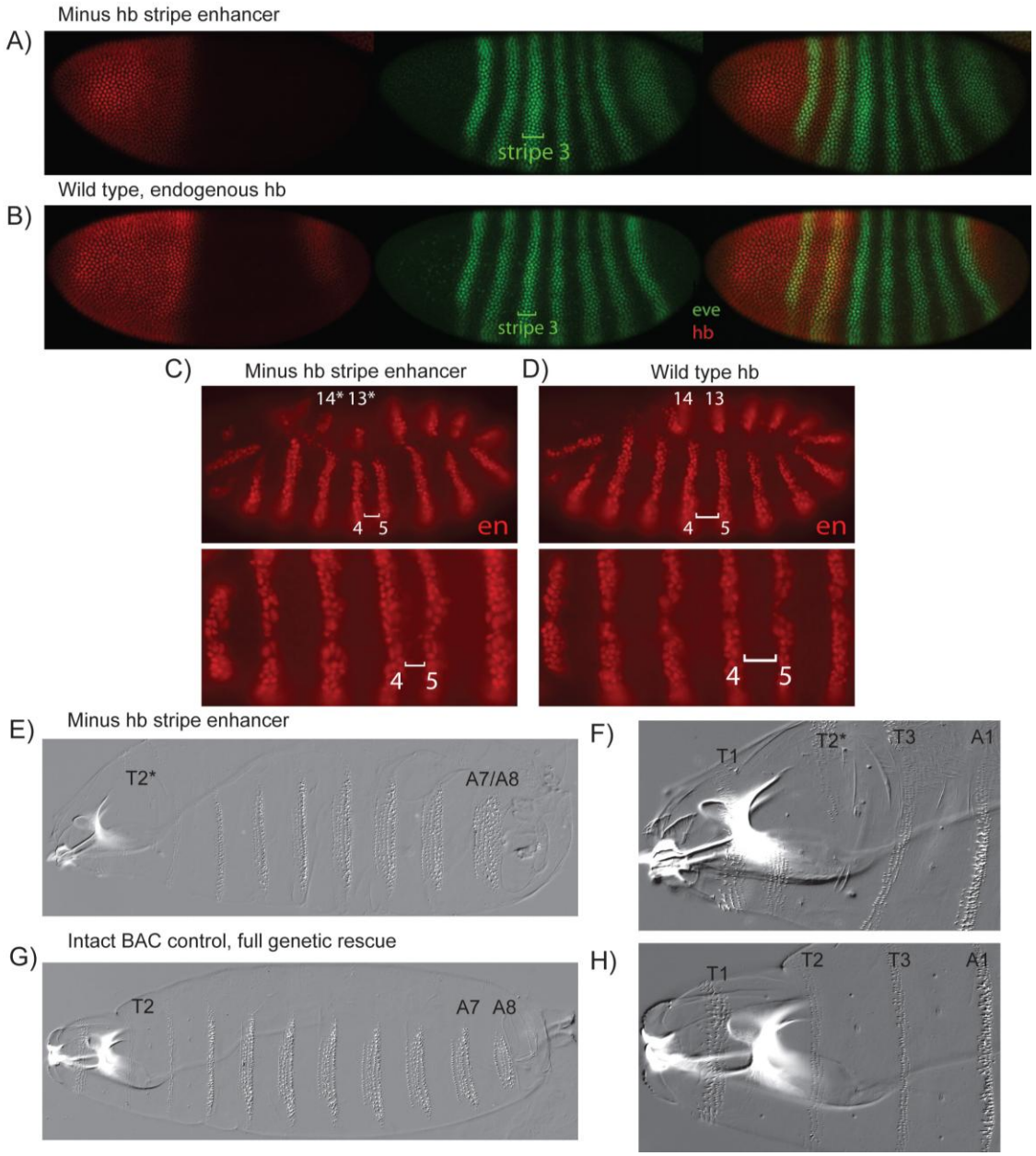


Figure 4.4 Effect removing the hb stripe enhancer on the downstream targets eve and engrailed, and the resulting larval cuticle phenotype. A-B) Removal of the stripe enhancer results in a shifted boundary of Hb protein expression and loss of high levels near the boundary as well as the posterior stripe. eve stripe 3 is consistently wider in this background. C)-D) engrailed stipes 4 and 5 are compressed; note the partial loss of stripe 13 in the posterior. E)-H) There is occasional loss of the ventral denticles of thoracic segment 2 (T2) (in approximately 25% of larvae examined).

eve stripe 3 (Fig. 4.4A; compare with B). The irregular spacing of eve stripes 2 and 3 results in narrowing of engrailed stripes 4 and 5 (Fig. 4.4D; compare with C), and variable losses of ventral denticles in the mesothorax (Fig. 4.4 E,F; compare with G,H). Thus, the hb stripe enhancer is important for the normal expression of eve and en in the presumptive thorax.

Quantitative measurements indicate significant alterations in the posterior limit of the anterior Hb expression pattern upon removal of the stripe enhancer (Fig. 4.5). There is an anterior shift of the final boundary by ~3% embryo length (EL) as compared with the wild-type Hb BAC transgene (averages at 44% and 47.2%EL, respectively). Moreover, the boundary is not as sharp even in the posterior-most regions of the pattern. The pattern becomes especially variable in a region around 40%EL. Most importantly, there is a 2-fold increase in the embryo-to-embryo variation in the distribution of the Hb protein at the posterior limit where there are diminishing and variable levels of the Bicoid gradient (measured at 50% max intensity). We therefore conclude that the stripe enhancer is essential for suppressing this variation to produce a relatively invariant Hb expression pattern with a sharp, precise boundary.

Previous studies implicated "redundant" repression by Kr and Kni in canalizing the Hb expression pattern (Clyde *et al.* 2003; Manu *et al.* 2009). Double mutants exhibit variable Hb borders, similar to those seen upon removal of the stripe enhancer (Manu *et al.* 2009). Computational analysis and whole-genome ChIP-Seq assays identified numerous Kr and Kni binding sites in the stripe enhancer, as well as a large number of Hb binding sites (Figure 4.6A, see Methods). There are just a few such sites within the the proximal and shadow enhancers, and the lone Hb site in the proximal enhancer might help augment expression in response to the Bicoid gradient (Figure 4.6B). To determine the mechanism by which the central stripe enhancer canalizes the Hb expression pattern we mutagenized a number of the Hb, Kr, and Kni sites (Fig. 4.7). Most of our efforts focused on the Hb binding sites since previous studies suggested that Hb positive autofeedback might be important for establishing a sharp on/off pattern of expression (Margolis *et al.* 1995; Treisman *et al.* 1989; Lopes *et al.* 2008). Autofeedback might help produce sharper boundaries by increasing the level of expression in previously "on" cells and not increasing the levels in a cell under a certain threshold. This mechanism is not thought to help limit variability, however – too little initial Bicoid activator results in less Hb, and if Hb activates itself,

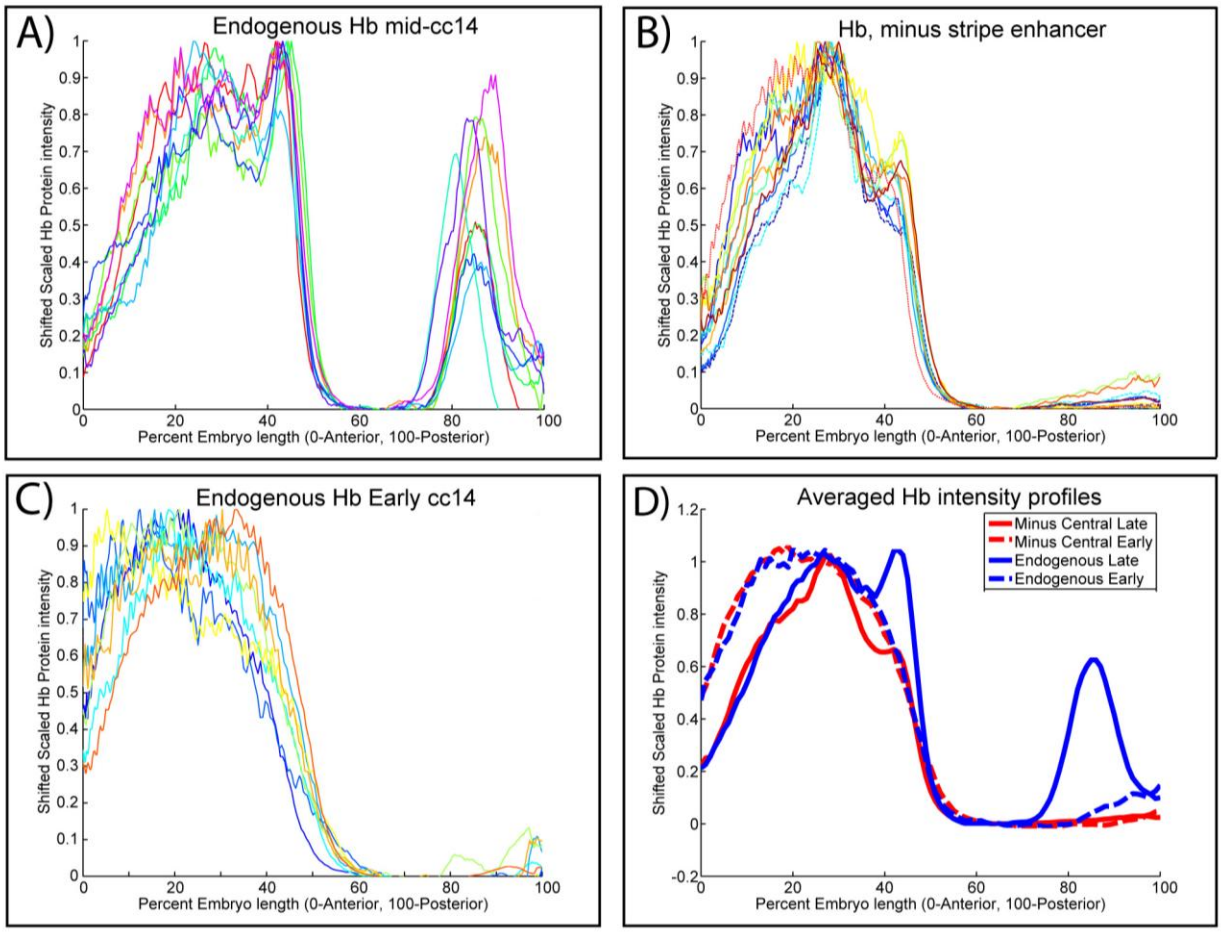


Figure 4.5 Quantification of Hb protein levels. Fluorescence intensity was quantified for multiple embryos for endogenous Hb and compared to embryos that are lacking the stripe enhancer. A) Endogenous Hb protein is quantified by staining intensity along the AP axis in pre-gastrulation embryos in mid-cc14 in which stripes of eve protein have fully resolved. B) Shows the same quantification in embryos minus the central stripe enhancer. C) Hb protein intensity in early cc14 is much more graded (shallower slope from maximum to minimum) than it is with central stripe enhancer input by mid cc14. D) Average curves for each data set. Embryos minus the central stripe enhancer produce more of a gradient than a boundary, much like the endogenous Hb in early cc14 embryos.

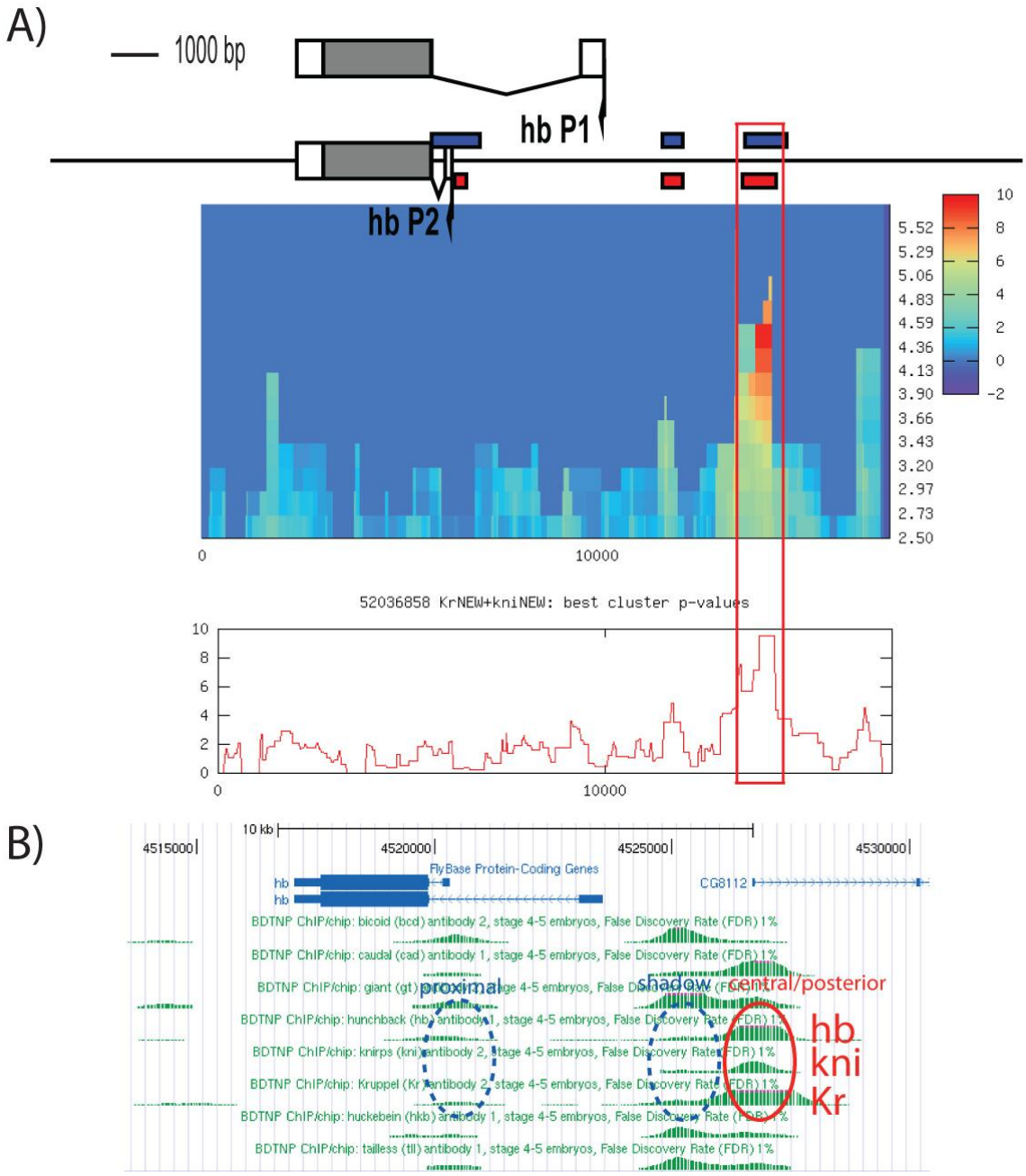


Figure 4.6 Binding site cluster analysis and gap gene ChIP data. A) Most Kr and kni binding sites at the hb locus are at the position of the stripe enhancer. B) ChIP-chip data for the gap genes shows the most relative Kr, kni, and hb binding occurs at the central stripe enhancer. (Gap gene ChIP data provided by the Berkeley Drosophila Trancription Project, Li *et al.* 2008).

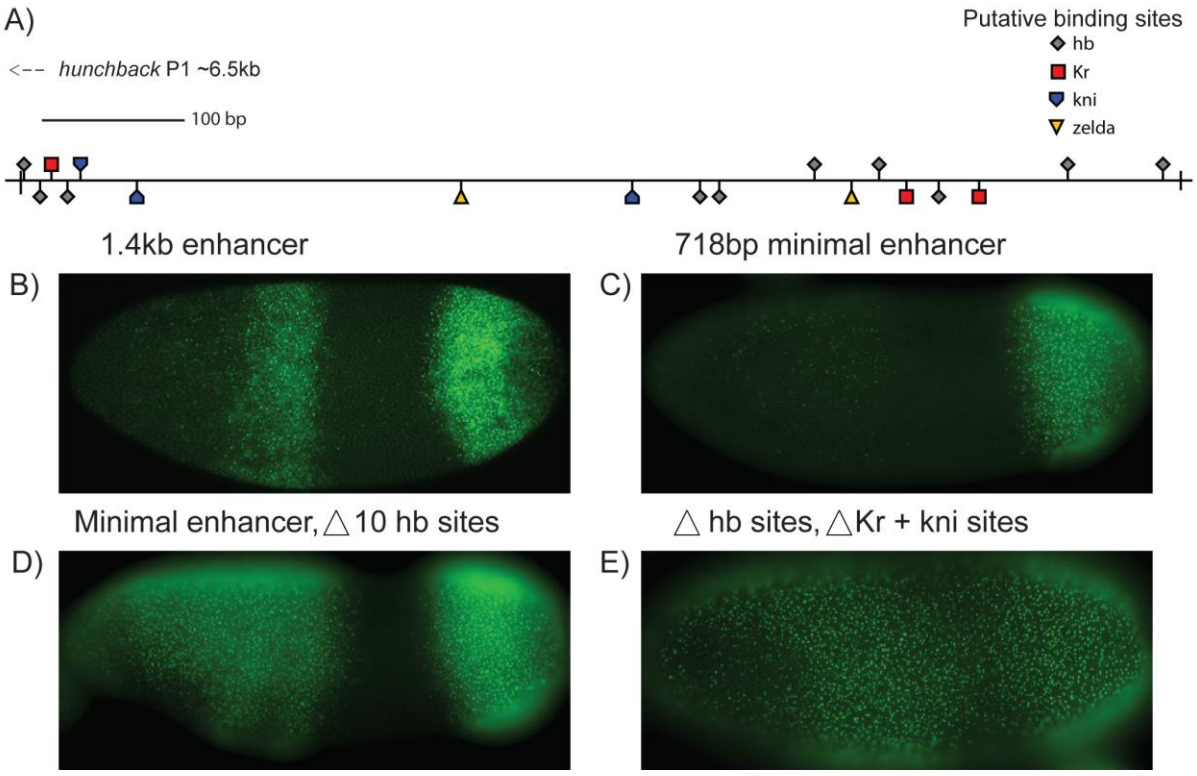


Figure 4.7 Mutagenesis of the minimal hb stripe enhancer. A) The distribution of repressor binding sites in this fragment. B) A 1.4kb enhancer recapitulates the portion of the pattern driven by this enhancer. C) A minimal 718bp enhancer was more amenable to mutagenesis, but drove less expression in central regions. D) This central pattern is much more strongly activated upon mutagenesis of 10 hb binding sites. E) Upon deletion of Kr and kni binding sites the early pattern expands into central regions. This construct was designed using an older position weight matrix for kni sites, and may have overlooked several sites. A new line (still in progress) may show cleaner loss of repression in these regions. Overall, it is clear that this enhancer is primarily shaped by repression and not activation. It is possible that further mutagenesis (for example, of a conserved STAT site) may help determine what activates it throughout the embryo.

less Hb initially means less Hb later given autoactivation. Too much initial Bicoid results in too much Hb, and too much autoactivation. The initial variability is at least preserved if not amplified by an subsequent downstream autoactivation.

Due to the large number of putative gap repressor sites we attempted to identify a minimal stripe enhancer. A 1.4 kb region of the enhancer produces a nearly normal pattern of expression when attached to a lacZ reporter gene and expressed in transgenic embryos (Fig. 4.2D; 4B). This is the smallest fragment we tested that recapitulates the central Hb stripe and posterior expression pattern. However, it contains too many binding sites for systematic mutagenesis (the 1.4kb region contains >35 potential hb sites, for example). A 700 bp fragment containing many fewer sites was examined since it produces weak but reproducible expression (Fig. 4.7C). We were surprised to find that mutagenesis of the Hb binding sites resulted in a substantial expansion in the expression pattern, rather than reduced expression as predicted by positive autofeedback models (Fig. 4.7D). Mutations in both the Hb and Kr binding sites resulted in further expansion of the expression pattern (Fig. 4.7E), consistent with the idea that gap repression is essential for suppressing variations in the Hb gradient. However, this repression is not limited to Kr and Kni, but also includes Hb itself.

The Hb stripe enhancer is probably regulated by the combination of ubiquitous activators, such as STAT, and localized gap repressors, particularly Hb, Kr, and Knirps. This strategy governs the regulation of the eve stripe 3+7 enhancer, which mediates a similar pattern of expression (Struffi *et al.* 2011). We suggest that the altered patterns of Hb expression observed in hb mutants is due to expanded expression of Kr and Kni, rather than direct positive autofeedback as previously suggested (Figure 4.8) (Margolis *et al.* 1994; Luckowitz *et al.* 1994; Treisman *et al.* 1995; Lopes *et al.* 2008; Holloway *et al.* 2011).

The dynamic regulation of the zygotic Hb expression pattern can be explained by the combinatorial action of the proximal, shadow, and stripe enhancers. The proximal and distal shadow enhancers mediate activation of Hb transcription in response to the Bicoid gradient in anterior regions of cc10-13 embryos. The initial border of Hb transcription is rather sharp, but the protein that is synthesized from this early pattern is distributed in a broad and shallow gradient. During cc14 the stripe enhancer mediates

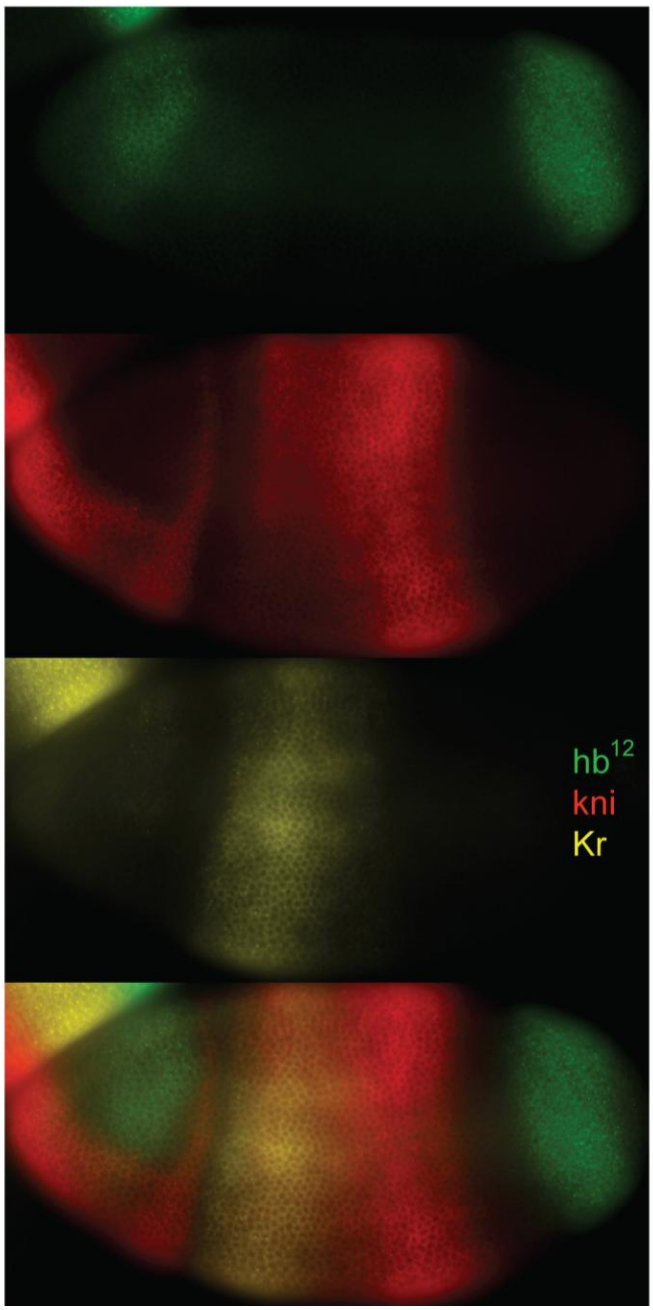


Figure 4.8 Expansion of other gap repressors in a hb mutant embryo. In the null point mutant background hb^{12}, non-functional *hb* transcript can still be seen, shown here in green. Both *Kr* and *kni* clearly expand into the normal *hb* domain. *hunchback* itself is restricted to a small anterior domain.

transcription in a domain that extends just posterior of the initial Hb border. Gap repressors, including Hb itself, restrict this second wave of zygotic Hb transcription to the region when there are variable levels of the Bicoid gradient, in a stripe that encompasses 44-47% egg length. The protein produced from the stripe enhancer is distributed in a sharp and steep gradient in the anterior thorax. It has been previously suggested that the steep Hb protein gradient is a direct readout of the broad Bicoid gradient. However, this sharp Hb protein boundary is never formed upon removal of the stripe enhancer, indicating that the Bicoid target enhancers are not sufficient to establish the definitive Hb pattern.

We suggest that Hb autorepression provides a potent mechanism for correcting variations in the Bicoid gradient by restricting expression from the stripe enhancer to the region where it is needed most to suppress variations in the Hb border. In contrast, Hb positive autofeedback would tend to preserve or augment variations in the Hb gradient, rather than dampening it. We propose that the pervasive use of cross-repressive interactions serves to limit variations in gene expression in the precellular embryo, prior to the deployment of corrective cell-based signaling mechanisms.

4.5 Materials and Methods

Recombineering was performed as described in previous chapters. The y-reporter was inserted so that native hb UTRs were left intact. Genetic rescue experiments were performed using a newly created hb deletion using flanking p-element insertions containing FRT sites and FLP-mediated recombination, as in Ryder *et al.* 2007. Protein staining intensity was measured as in Houchmandzadeh *et al.* 2002.

4.6 Concluding Remarks and Future Directions

4.6.1 The role of selection in shadow enhancer conservation

I have described experiments designed to test the function of a set of shadow enhancers as well as the *hb* stripe enhancer. In each case, though sometimes in slightly different ways, they seem to play a role in ensuring the precision and robustness of gene expression patterns at the right place and time during development. This seems particularly true in varying thermal environments and when sensitized genetically, but it is difficult to tell what about these mechanisms is shaped directly by selection or how strong that selection

might be. Under normal laboratory conditions we may never have observed a difference between the various experimental *snail* lines. Putting these same lines into a Dorsal heterozygote background made differences more apparent, but such a treatment is much more extreme than any embryo would normally encounter in the wild. Developmental biologists have long used such means to sensitize backgrounds and make experimental effects more obvious. Given the deep conservation of many of these enhancers, however, selection is clearly able to operate efficiently on much less dramatic differences. Perhaps selection only occurs on individual embryos which experience thermal extremes (out of a larger population of embryos in a range of environments), but selective pressure on a portion of the population is enough to maintain these additional seemingly redundant enhancers. Whatever the exact role of selection, these enhancers are deeply conserved and it seems likely that the role characterized here in providing robustness and precision of gene expression plays an important part.

4.6.2 The ease (or difficulty) of modification for novel function

In some cases, such as for several of the gap genes, it could be argued that at least one of the pair may be more constrained than the other, as they can play a more inseparable role. The enhancer in the *hb* or *kni* pair that is bound by dominant, long-range repressors could not easily be lost without an expansion of the overall pattern. Then again, the second enhancer of the pair could be lost without changing the boundaries of the pattern but it's loss would come with a reduction in completeness of pattern; many cells within the domain that normally stay active would not be "on" and actively transcribing the gene as much of the time. This might make for more ragged boundaries or reduced overall levels of expression. The loss of either enhancer could come at a cost. This might constrain the system in a way that would make it more difficult for either enhancer to take on novel roles (one of the ideas initially proposed in Hong *et al.* 2008). This cost could potentially be compensated for by the advantages conferred by the novel mode of expression, but it may be that this could only occur in situations in which selection is relaxed, perhaps for isolated populations in more moderate environments. Such situations could at least temporarily reduce the costs of modifying these enhancers for novel use. Unfortunately, it is unclear how often such situations occur or even if any of these enhancers has ever taken on truly novel function, or if that process was made any more likely by it being in a shadow enhancer pair. There is at least some flexibility in some species, some back and forth, as seen in Chapter 1 (Figs 1.9, 1.10).

It is possible that changes to underlying variation in a shadow enhancer pair are more neutral than changes to binding sites in a lone enhancer. For example, an accumulation of mutations that in effect add an additional long-range repressor binding site to the second enhancer in a pair may or may not change the overall pattern – it would be largely neutral in effect. It is easy to imagine a situation in which the gain of a dominant repressor binding site in one enhancer would subsequently allow for the loss of a similar site in the second enhancer. These trade-offs in underlying regulatory structure might be neutral in terms of gene expression patterns, phenotype, and fitness, but would change the distribution of sites that are available for modification in the future. Evolution of a novel pattern might occur by the modification of one of the two enhancers in one species, but the other of the pair in another species. An increase in underlying variation that is initially hidden from selection could help provide novel material for selection to act on after subsequent modifications cause underlying differences to be exposed.

4.6.3 The origin of shadow enhancers

Very little is known about how enhancers evolve *de novo*, and even less about shadow enhancer origin. In situations where selection is reduced and the system is inherently more flexible, it may be that random mutation can lead to the occasional accumulation of a few chance binding sites that affect an overall pattern. These can then be selected for or against, as is likely in some cases for Drosophila pigmentation (Kalay and Wittkopp, 2010). This leads to the ready gain and loss of enhancers and changes in enhancer position. For more constrained developmental patterning systems it may be more common to modify existing enhancers than for new ones to arise, as discussed in Chapter 1 and Cande *et al.* 2009. Given these apparently fundamental differences between types of patterning it would be interesting to identify and study shadow enhancers in more terminal patterning systems, where it may be easier to determine how a shadow enhancer has arisen and if it has increased (or decreased) the flexibility of the system or its ability to be modified further.

Shadow enhancers might arise in several ways. Accumulation of mutations that by chance result in binding sites that resemble activator and repressor consensus sequences in combination could in theory build a sort of minimal enhancer. Given a set of basic rules that are required for all known enhancers (e.g. some minimal combination of activator binding sites with the

right spacing for cooperativity and at least one repressor to help set spatial limits) it quickly becomes clear that this would be a rare event, especially when these limits must match the pattern of an existing enhancer for a given gene. Alternately, existing enhancers could be modified such that two distinct enhancers take on increasingly overlapping patterns in space and time during development. A third possible mechanism would involve the duplication of existing enhancers, potentially leading to a relatively cryptic change in function but which might help provide some of the potential benefits described previously in providing robustness of expression, especially in varying conditions. Selection could then fine-tune the pair, with functionally redundant dominant repressor sites coming or going, and they could begin to either diverge slightly or remain constrained, depending on functional requirements and selective pressures. Such duplicated enhancers could provide immediate value for fitness given the right conditions. However the new enhancers arise, they might then facilitate the evolution of novel patterns of gene expression, or become incorporated into core developmental machinery to help produce precise patterns of gene expression.

References

Barolo, S., and Levine, M. 1997. Hairy mediates dominant repression in the *Drosophila* embryo. *EMBO J* **16**: 2883-2981.

Barolo, S. 2011. Shadow enhancers: Frequently asked questions about distributed *cis*-regulatory information and enhancer redundancy. *Bioessays* epub ahead of print.

Barrallo-Gimeno, A., and Nieto, M.A. 2005. The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* **132:** 3151-3161.

Berman, B.P., Nibu, Y., Pfeiffer, B.D., Tomancak, P., Celniker, S.E., Levine, M., Rubin, G.M., Eisen, M.B. 2002 Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl. Acad. Sci. USA* **99**:757-762.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. 2007. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. USA* **104**: 3312-3317.

Boettiger, A.N. and Levine, M. 2009. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* **325**: 471-473.

Boulay, J.L., Dennefeld, C., and Alberga, A. 1987. The *Drosophila* developmental gene *snail* encodes a protein with nucleic acid binding fingers. *Nature* **330**: 395-398.

Cande, J., Goltsev, Y., and Levine, M. 2009. Conservation of enhancer positions in divergent insects. *Proc. Natl. Acad. Sci. U.S.A.* **106**: 14414-14419.

Clyde, D.E., Corado, M.S., 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.

Courey, A.J., and Jia, S. 2001. Transcriptional repression: the long and the short of it. *Genes Dev* **15**: 2786-2796.

Cowden, J., and Levine, M. 2002. The Snail repressor positions Notch signaling in the *Drosophila* embryo. *Development* **129**: 1785-1793.

Datsenko, K.A., and Wanner, B.L. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**: 6640-6645.

Driever, W., and Nüsslein-Volhard, C. 1988a. A gradient of bicoid protein in Drosophila embryos. *Cell* **54:** 83-93.

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

Driever, W., and Nüsslein-Volhard, C. 1989. The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature* **337**: 138-43.

Driever, W., Thoma, G., and Nüsslein-Volhard, C. 1989. Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen. *Nature* **340:** 363-367.

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 .

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* **126**: 2527-2538.

Gabay, L., Scholz, H., Golembo, M., Klaes, A., Shilo, B.Z., and Klambt, C. 1996. EGF receptor signaling induces pointed P1 transcription and inactivates Yan

protein in the *Drosophila* embryonic ventral ectoderm. *Development* **122**: 3355-3362.

Goldstein, R.E., Jiménez, G., Cook, O., Gur, D., Paroush, Z. 1999. Huckebein repressor activity in *Drosophila* terminal patterning is mediated by Groucho. *Development* **126**: 3747-3755.

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

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

Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W., and Tank, D. W. 2007. Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* **130**: 141-52.

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.

Häder, T., Wainwright, D., Shandala, T., Saint, R., Taubert, H., Brönner, G., and Jäckle, H. 2000. Receptor tyrosine kinase signaling regulates different modes of Groucho-dependent control of Dorsal. *Curr Biol* **10**: 51-54.

Hartmann, B., Reichert, H., Walldorf, U. 2001. Interaction of gap genes in the *Drosophila* head: *tailless* regulates expression of *empty spiracles* in early embryonic patterning and brain development. *Mech Dev* **109**: 161-72.

He, F., Wen, Y., Deng, J., Lin, X., Lu, L. J., Jiao, R., and Ma, J. 2008. Probing intrinsic properties of a robust morphogen gradient in Drosophila. *Dev Cell* **15**: 558-67.

Hemavathy, K., Hu, X., Ashraf, S.I., Small, S.J., and Ip, Y.T. 2004. The repressor function of snail is required for *Drosophila* gastrulation and is not replaceable by Escargot or Worniu. *Dev. Biol.* **269**: 411-20.

Hobert, O. 2010. Gene regulation: enhancers stepping out of the shadow. *Curr Biol* **20**:R697-699.

Hoch, M., Schröder, C., Seifert, E., and Jäckle, H. 1990. cis-acting control elements for *Krüppel* expression in the *Drosophila* embryo. *EMBO J* **9**:2587-2595.

Holloway, D., Lopes, F., Fontoura Costa, L. da, Travençolo, B., Golyandina, N., Usevich, K., and Spirov, A. 2011. Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in Drosophila segmentation. *PLoS Comp Bio* **7**: e1001069.

Hong, J.W., Hendrix, D.A., and Levine, M.S. 2008a. Shadow enhancers as a source of evolutionary novelty. *Science* **321**: 1314.

Hong, J.W., Hendrix, D.A., Papatsenko, D., and Levine, M.S. 2008b. How the Dorsal gradient works: insights from postgenome technologies. *Proc. Natl. Acad. Sci. USA* **105**: 20072-20076.

Houchmandzadeh, B., Wieschaus, E., and Leibler, S. 2002. Establishment of developmental precision and proportions in the early Drosophila embryo. *Nature* **415:** 798-802.

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.

Ip, Y.T., Kraut, R., Levine, M., and Rushlow, C.A. 1991. The dorsal morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in *Drosophila*. *Cell* **64**: 439-446.

Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K., and Levine, M. 1992. dorsal-twist interactions establish snail expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev.* **6**: 1518-30.

Janody, F., Reischl, J., Dostatni, N. 2000. Persistence of Hunchback in the terminal region of the *Drosophila* blastoderm embryo impairs anterior development. *Development* **127**: 1573-1582.

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

Jaźwińska, A., Kirov, N., Wieschaus, E., Roth, S., and Rushlow, C. 1999. The *Drosophila* gene *brinker* reveals a novel mechanism of Dpp target gene regulation. *Cell* **96**: 563-573.

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., and Carroll, S.B. 2008. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* **132**: 783-93.

Kalay, G., and Wittkopp, P. 2010. Nomadic enhancers: tissue-specific *cis*-regulatory elements of *yellow* have divergent genomic positions among *Drosophila* species. *PLoS Genet* **6**: e1001222.

Kasai, Y., Nambu, J.R., Lieberman, P.M., and Crews, S.T. 1992. Dorsal-ventral patterning in *Drosophila*: DNA binding of snail protein to the single-minded gene. *Proc. Natl. Acad. Sci. USA* **89**: 3414-3418.

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 *Drosophila* embryo. *Curr Biol* **20**: 446-451.

Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A. Z., Engström, P. G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K, *et al.* 2007. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Research* **17**: 545-555.

Kosman, D., Ip, Y.T., Levine, M., and Arora, K. 1991. Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. *Science* **254:** 118-22.

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.

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

Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A., and Copeland, N.G. 2001. A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**: 56-65.

Leptin, M. 1991. twist and snail as positive and negative regulators during *Drosophila* mesoderm development. *Genes Dev*. **5**: 1568-76.

Li, L.M., and Arnosti, D.N. 2011. Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes. *Curr Biol* **21**: 406-12.

Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks C.L., Chu, H.C., Ogawa, N., Inwood, W., Sementchenko, V., Beaton, A., Weiszmann, R., Celniker, S.E., Knowles, D.W., Gingeras, T., Speed, T.P., Eisen, M.B., Biggin, M.D. 2008. Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* **6**: e27.

Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., and Epstein, D.J. 2006. A functional screen for *sonic hedgehog* regulatory elements across a 1Mb interval identifies long-range ventral forebrain enhancers. *Development* **133**: 761-772.

Liu, P., Jenkins, N.A., and Copeland, N.G. 2003. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome research* **13**: 476-484.

Lopes, F., Vieira, F., Holloway, D., Bisch, P., and Spirov, A. 2008. Spatial bistability generates hunchback expression sharpness in the *Drosophila* embryo. *PLoS Comp Bio* **4:** e1000184.

Lott, S.E., Kreitman, M., Palsson, A., Alekseeva, E., and Ludwig, M.Z. 2007. Canalization of segmentation and its evolution in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **104**: 10926-10931.

Ludwig, M.Z. 2002. Functional evolution of noncoding DNA. *Curr. Opin. Genet. Dev.* **12**: 634-639.

Lukowitz, W., Schroder, C., Glaser, G., Hulskamp, M., Tautz, D. 1994. Regulatory and coding regions of the segmentation gene *hunchback* are functionally conserved between *Drosophila virilis* and *Drosophila melanogaster*. *Mech Dev* **45**: 105-115.

Manu, Surkova, S., Spirov, A.V., Gursky, V.V., Janssens, H., Kim, A., Radulescu, O., Vanario-Alonso, C.E., Sharp, D.H., Samsonova, M., and Reinitz, J. 2009a. Canalization of gene expression in the *Drosophila* blastoderm by gap gene cross regulation. *PLoS Biology* **7**:e1000049.

Manu, Surkova, S., Spirov, A.V., Gursky, V.V., Janssens, H., Kim, A., Radulescu, O., Vanario-Alonso, C.E., Sharp, D.H., Samsonova, M., and Reinitz, J. 2009b. Canalization of gene expression and domain shifts in the *Drosophila* blastoderm by dynamical attractors. *PLoS Comp Biol* **5**:e1000303.

Margolis, J.S., Borowsky, M.L., Steingrímsson, E., Shim, C.W., Lengyel, J.A., and Posakony, J.W. 1995. Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element. *Development* **121**: 3067-3077.

Markstein, M., Zinzen, R., Markstein, P., Yee, K.P., Erives, A., Stathopoulos, A., and Levine, M. 2004. A regulatory code for neurogenic gene expression in the *Drosophila* embryo. *Development* **131**: 2387-94.

Maxton- Küchenmeister, J., Handel, K., Schmidt-Ott, U., Roth, S., and Jäckle, H. 1999. Toll homologue expression in the beetle *Tribolium* suggests a different mode of dorsoventral patterning than in *Drosophila* embryos. *Mech. Dev.* **83**: 107-114.

Morel, V., Le Borgne, R., and Schweisguth, F. 2003. Snail is required for Delta endocytosis and Notch-dependent activation of *single-minded* expression. *Dev. Genes Evol.* **213**: 65-72.

Nibu, Y., Zhang, H., and Levine, M. 1998. Interaction of short-range repressors with *Drosophila* CtBP in the embryo. *Science* **280**: 101-104.

Nunes da Fonseca, R., von Levetzow, C., Kalscheuer, P., Basal, A., van der Zee, M., and Roth, S. 2008. Self-regulatory circuits in dorsoventral axis formation of the short-germ beetle *Tribolium castaneum*. *Dev. Cell* **14**: 605-615.

Ochoa-Espinosa, A., Yucel, G., Kaplan, L., Pare, A., Pura, N., Oberstein, A., Papatsenko, D., AND Small, S. 2005. The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **102**: 4960-4965.

O'Meara, M., Bigelow, H., Flibotte, S., Etchberger, J.F., Moerman, D.G., and Hobert, O. 2009. Cis-regulatory mutation in the *Caenorhabditis elegans* Homeobox gene locus *cog-1* affect neuronal development. *Genetics* **181**: 1679-1686.

Papatsenko, D. 2007. ClusterDraw web server: a tool to identify and visualize clusters of binding motifs for transcription factors. *Bioinformatics* **23**: 1032-1034.

Pankratz, M., Busch, M., Hoch, M., Seifert, E., Jackle, H. 1992. Spatial control of the gap gene knirps in the *Drosophila* embryo by posterior morphogen system. *Science* **255**: 986-989.

Parkhurst, S.M., and Corces, V.G. 1986. Interactions among the gypsy transposable element and the yellow and the suppressor of hairy-wing loci in *Drosophila melanogaster*. *Mol. Cell Biol.* **6**: 47-53.

Perry, M.W., Cande J.D., Boettiger A.N., and Levine, M. 2009. Evolution of insect dorsoventral patterning mechanisms. *Cold Spring Harbor Symposium* **74:** 275-279.

Perry, M.W., Boettiger A.N., Bothma J.P., Levine, M. 2010. Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr Biol* **20**: 1562-1567.

Perry, M.W., Boettiger, A.N., and Levine, M. 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc. Natl. Acad. Sci. USA* **108**: 13570-13575.

Porcher, A., Abu-Arish, A., Huart, S., Roelens, B., Fradin, C., and Dostatni, N. 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**: 2795-804.

Rankin, S., Li, Z., and Isberg, R.R. 2002. Macrophage-induced genes of *Legionella pneumophila*: protection from reactive intermediates and solute imbalance during intracellular growth. *Infect Immun.* **70**: 3637-3648.

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

Roth, S., Stein, D., and Nüsslein-Volhard, C. 1989. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell* **59**: 1189-1202.

Roth, S., Hiromi, Y., Godt, D., Nüsslein-Volhard, C. 1991. *cactus*, a maternal gene required for proper formation of the dorsoventral morphogen gradient in *Drosophila* embryos. *Development* **112**: 371-388.

Rushlow, C.A., Han, K., Manley, J.L., and Levine M. 1989. The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in *Drosophila*. *Cell* **59**: 1165-1177.

Rutherford, S.L., and Lindquist, S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* **396**: 336-342.

Schroeder, M.D., Pearce, M., Fak, J., Fan, H., Unnerstall, U., Emberly, E., Rajewsky, N., Siggia, E.D., and Gaul, U. 2004. Transcriptional control in the segmentation gene network of *Drosophila*. *PLoS biology* **2**:E271.

Shao, X., Koizumi, K., Nosworthy, N., Tan, D. P., Odenwald, W. and Nirenberg, M. 2002. Regulatory DNA required for vnd/NK-2 homeobox gene expression pattern in neuroblasts. *Proc. Natl. Acad. Sci. USA* **99**:113 -137.

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. 1996. Regulation of two pair-rule stripes by a single enhancer in the *Drosophila* embryo. *Developmental biology* **175**: 314-324.

Specchia, V., Piacentini, L., Tritto, P., Fanti, L., D'Alessandro, R., Palumbo, G., Pimpinelli, S., and Bozzetti, M.P. 2010. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* **463**: 662-665.

Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M., and Levine, M. 2002. Whole-genome analysis of dorsal-ventral patterning in the *Drosophila* embryo. *Cell* **111**: 687-701.

Stathopoulos, A., and Levine, M. 2005. Localized repressors delineate the neurogenic ectoderm in the early *Drosophila* embryo. *Dev Biol* **280**: 482-493.

Stein, D., and Nüsslein-Volhard, C. 1992. Multiple extracellular activities in *Drosophila* egg perivitelline fluid are required for establishment of embryonic dorsal-ventral polarity. *Cell* **68**: 429-440.

Steward, R. 1989. Relocalization of the dorsal protein from the cytoplasm to the nucleus correlates with its function. *Cell* **59**: 1179-1188.

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* **138**: 4291-4299.

Struhl, G., Struhl, K., and Macdonald, P. M. 1989. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell* **57**: 1259-73.

Struhl, G. 1989. Differing strategies for organizing anterior and posterior body pattern in *Drosophila* embryos. *Nature* **338**: 741-744.

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

Treisman, J., and Desplan, C. 1989. The products of the *Drosophila* gap genes hunchback and Krüppel bind to the hunchback promoters. *Nature* **341**: 335-337.

Venken, K.J., He, Y., Hoskins, R.A., and Bellen, H.J. 2006. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **14**: 1747-51.

Venken, K.J., Carlson, J.W., Schulze, K.L., Pan, H., He, Y., Spokony, R., Wan, K.H., Koriabine, M., de Jong, P.J., White, K.P., *et al.* 2009. Versatile P[acman] BAC libraries for transgenesis studies in *Drosophila melanogaster*. *Nature Methods* **6**: 431-4.

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

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A., and Copeland, N.G. 2005. Simple and highly efficient BAC recombineering using galK selection. *Nucleic acids research* **33**: e36.

Weiss, J.B., Von Ohlen, T., Mellerick, D.M., Dressler, G., Doe, C.Q., and Scott, M.P. 1998. Dorsoventral patterning in the *Drosophila* central nervous system: the intermediate neuroblasts defective homeobox gene specifies intermediate column identity. *Genes Dev*. **12**: 3591-3602.

Werner, T., Hammar, A., Wahlbuhl, M., Bosl, M.R., and Wegner, M. 2007. Multiple conserved regulatory elements with overlapping functions determine *Sox10* expression in mouse embryogenesis. *Nucleic Acids Res.* **35**: 6526-6538.

Winkler, C.J., Ponce, A., and Courey, A.J. 2010. Groucho-mediated repression may result from a histone deacetylase-dependent increase in nucleosome density. PLoS One **5**: e10166.

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**: 421-30.

Yu, D., and Small, S. 2008. Precise registration of gene expression boundaries by a repressive morphogen in *Drosophila*. Curr Bio **18**: 868-876.

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

Zinzen, R., Cande, J.D., Papatsenko, D., Levine, M. 2006. Evolution of the ventral midline in insect embryos. *Dev. Cell* **11**: 895-902.

Appendix I: Oligonucleotides

Table 1: Chapter 2 oligos

yellow-kanamycin fusion PCR.	
303YkanF1	cgaggatccggcgcgccATGTTCCAGGACAAAGGGTGGATCCTTG
304YkanR1	tagcagcccttgcgccctgaTTAACCTTGATGCTGATGATGCCACC
305KanYF1	atcatcagcatcaaggttaaTCAGGGCGCAAGGGCTGCTAAAGGAAG
306KanYR1	atttgcggccgcACTCTTCCTTTTTCAATTCAGAAGAAC
Y-kan insertion into snail BAC; from ATG through end of coding (leaves UTRs intact). Homology arm (65 bp) followed by yellow or kan primer (25 bp).	
423snaYkanLHA	GATCTCCCGATTTACCCATCTCGATCAGTACCGGAAACTAAAACTTAATCAC ACACACATCAAAA ATGTTCCAGGACAAAGGGTGGATCC
424snaYkanRHA	CGCCAGCGGAATGTGAGTTTGCTTAGGTAATTGTGTCCTGCTAAGGGATTC ATATGTCGAGAATC ACTCTTCCTTTTTCAATTCAGAAGA
Removal of the snail primary and shadow using the amp cassette; ampicillin from pBlueScript. Homology arm (65 bp) followed by Amp primer (25 bp).	
425snaPriLHA	CAAGGTGCAAAAATGGGACGGTCCTATTCTCAGCAAAAATTGACAAGAACA ACAACAATGTCTAT GGGGTCTGACGCTCAGTGGAACGAA
426snaPriRHA	GACAAGGAATAGACGAAGAGTGTCAAGTCTCCTGTGAGCTTGGATCGAAGA GATCTCCTGTTCCC GTGGCACTTTTCGGGGAAATGTGCG
427snaShLHA	TTAAACAAAACGGTGTTTGTTTTAAACAAAATACTTCAAAATTCGTAAAATAC GATTATCGTTAA GGGGTCTGACGCTCAGTGGAACGAA
428snaShRHA	TGCAGTATTTCCATTTAAGCTGTTTTTGCAAAGGAACCAATTTATTAAATTTC AAACAATTTAAC GTGGCACTTTTCGGGGAAATGTGCG
Recombination screening.	
324YScR1	AACCGGATTTATGAGAGGTCGCCACG
326KanScF1	GACTGGCTGCTATTGGGCGAAGTGC
343AmpRecScR1	CCAAACGACGAGCGTGACACCACGA
344AmpRecScF1	TGGTTATGGCAGCACTGCATAATTC
439snaYRecScF1	GAGTTATTACGGGTGGGTTGTGAGC
440snaKanRecScR1	TATTCACATCTGTCTGTATTTGAGC
447snaShLHAchF	TAGCCTCAACATCCCTGATAAATGG
448snaShLHAchR	AGCGAAGTCGAAGGTAGGACCAAGG
449snaShRHAchF	ACAGGTAAACATTTGCCAATATCTC
450snaShRHAchR	AGGCCACTGCAACTAGCTAAATTCC
451snaPrLHAchF	TATGCGAATATTTAATAAGAAGAAC
452snaPrLHAchR	ATATCCGTTGAAGTACCTGACTTGC
453snaPrRHAchF	TGGAGGTGCTAGCTGGGAAAGTTGTG
454snaPrRHAchR	ATGACACCTTCGCAGAGAGGTGTTC

Table 2: Chapter 3 oligos and candidate regions

Gap enhancer primers

Description	Primer name	Primer sequence
hb-5 hb distal enhancer, minimal	Hb-5F2	TACAAAACAACAGAGCAAACAAATCGC
	Hb-5R2	GCAACGCTAATCGGGCTGATCAGCGAAC
hb-5 hb distal enhancer, non-minimal	Hb-5F1	ATTCATTATTCATAGTTCATTGGTAA
	Hb-5R1	GGTTTCTAATTGATTCAATACAGGATTAC
hb-1 hb proximal enhancer	HbEndgF1	CCTGCTGTCGACTCCTGACCAACGTAATC
	HbEndgR1	TGTCGTCTCCCAGTTCTGCATCTTGGC
Kr proximal enhancer	370KrCD2modF1	attgcggccgcGTAAGTTCCCATATTTCGGACCTTATC
(modified from "CD2")	371KrCD2modR1	attgcggccgcTGGGTACTTCGCTGAGTTGAGTGAGTTG
Kr distal enhancer	372KrCD1F1	attgcggccgcGGATCCTAAGTTAACTATAATCCAGG
(modified from "CD1")	373KrCD1R1	cggactagtGTTGTTTTTGTACATATAACAATATCC
kni distal (5' "kd" enhancer)	514kniPBstF1	agcttcgaaAATTCGGGTGCTCTTGTTTGGCTCG
(with BstBI sites used to place 3' of lacZ)	515kniPBstR1	agcttcgaaAAAGTCAACTTCCAACATAATTTCTC
kni proximal minimal	368Kni+1F3	attgcggccgcAGGAAACTGGGAAAAACTAGACAGGATG
	369kni+1R3	cggactagtGAATCACTCAGACTGATCGAAAAGTGC
kni proximal expanded (includes	Kni+1expF1	GGTGGTGCGGTTCTTCTTGTCGATGATG
endogenous promoter)	Kni+1expR1	GAGTGAGTGAGAAATCCAGCCGCCCTTAG
oc/otd intronic "distal" enhancer	361Otd+7F	attgcggccgcTGAACAATGTCACCTATGCTACTGTAG
	367Otd+7R2	cggactagtTAGTTCCGTTTTCCGTATTCCATGTGC
ems distal 3' enhancer	DmelEms3'ShaF1	GGAACATTGAGAATCAGCCAAGAGATC
	DmelEms3'ShaR1	CGCATGTTGCCACATGAGGCAGAGG
gt distal-most +36kb enhancer	DmelGt+36F1	CGATGCATCCGTGTCATGGTTGGCATG
	DmelGt+36R1	GCAGTTCGCCAGGCGACGTAGTCG
fkh distal -4kb enhancer	fkh-4F1	CCACCATGGGACATCGGACGATGGAC
	fkh-4R1	CTGTTTGGGTTGCCTTTGAGGGGTTCC
hkb distal +6kb enhancer	hkb+6F1	CGTTTGTTTCCTAATTGGCCCATCATG
	hkb+6R1	GAACTGAAATAGAAGCAGTGCGAAAACC
(tailless enhancer previously described)		

***in situ* probe primers**		
hb full length probe	400hbFLPRF1	TTCGTTCGATTCGAATTCGCTTTCAAC
	401hbFLPRR1	AGTCAGTCACGAGTTTGTTACCACTG
Kr full length probe	402KrFLPRF1	ATCAGTCGTGATTTGGCTCTGTCAGC
	403KrFLPRR1	TATTTGAAGCAGCAACAATTAAACC
kni full length probe	404kniFLPRF1	TGCCAATCTTTAGACATACCATAAC
	405kniFLPRR1	AGTAGCCGAAGTGTCAGTTACACAC
gt full length probe	406gtFLPRF1	TTGAAATTCATTCCTTAGGGCTATG
	407gtFLPRR1	AGTTTGCGTTCGACATCGTCAGCGTG
yellow intronic probe	353YintF1	GTAAGTGGAAGTTAAATATGAAGCCCT
	354YintR1	CTAAAAAATAAAATAGGAGAATTAGCAG
yellow full length probe	301YcstF1	CTTGTGACCCTGATCACCTTGGTGACG
	302YcstR1	TTAACCTTGATGCTGATGATGCCACCAC
otd probe	549otdF1	GCAGTATTGGAAGAGTAGATTCTACAC
	550otdR1	AGCAGTCGAATAGCTTGAGCAGCTCG
ems probe	551emsF1	TACTTACTTCCAGGGAATCGGTGTGG
	552emsR1	ATCGAACGTCGTCGAGAGCACACGTC
hkb probe	547hkbFLF1	TCAACTCTCAGATATAACCATGATCG
	548hkbFLR1	AGGAGAGCTCAGTTCGAGTCGAGATC
fkh probe	553fkhT7F1	CCAATGCCGAACTGTGCCTCAGCCA
	554fkhR1	AGCTAGATATTCACGTTCGTGCGTCG
Recombineering primers		
Homology arm primers used to put yellow/kanamycin reporter in place of hb coding sequence	417hbYkanLHA	GAATATGCCCACTAACCCCACTCTCTCTGTTTTCTTATCCATTACAGCCGTCTAGAGCCGCCAAG ATGTTCCAGGACAAAGGGTGGATCC
(sequence after space is to amplify amp, etc.)	418hbYkanRHA	TTCTGGACAACGATTATATGATAATAGTGATAAATAATAATAACAAGGTGATGGTGATGGGGAAC ACTCTTCCTTTTTCAATTCAGAAGA
Homology arm primers used to replace hb proximal enhancer with ampicillin	419hbPriLHA	GATTTTTCAACAAAAACATTTTTTGTGTGGCGCATTTTCTGCGTTTTCGAATTTTTCCATTTTTG GGGGTCTGACGCTCAGTGGAACGAA
	420hbPriRHA	GAACAATTGCAACAGGCATTAGTTTATATATCGCTCAGGTAGACGGATGCACGCGTCAAGGGATT GTGGCACTTTTCGGGGAAATGTGCG

Homology arm primers used to replace the hb distal enhancer with ampicillin	421hbShLHA	TTCCCCTCGAAGGAACTTCTGTTGAGACAGCCGCAGGCCAAAGGTTAAGGATAAGGATCATTTGC GGGGTCTGACGCTCAGTGGAACGAA
	422hbShRHA	ACAATTAGCATTCGAAATTCCAAATCTCCTTAAGGTAGTTTGTGCATATTGGGTGTATTCCACCT GTGGCACTTTTCGGGGAAATGTGCG
Recombination screening primers		
	343AmpRecScR1	CCAAACGACGAGCGTGACACCACGA
Internal ampicillin primers (test for proper recombination, used with specific external primers)	344AmpRecScF1	TGGTTATGGCAGCACTGCATAATTC
	520AmpRcScR2	GGTAACTGTCAGACCAAGTTTACTC
hb proximal recombination screening	429hbPriRecScF1	AATGGCAAATAGTTCGGCACTTGCC
	430hbPriRecScR1	TGGTCGTGGCTGTCGTCTCCCAGTTC
hb distal recombination screening	431hbShRecScF1	CAGCTTCCAAAACCTTTAAGTATGC
	432hbShRecScR1	GCAGAAATCGAAATGAGAGAGCAGC
hb yellow-kanamycin screening	433hbYRecScF1	TCCCGTCACCTCTGCCCATCTAATC
	434hbKanRecScR1	AGAACTGAGTGTTATGCGCATATAC
	324YScR1	AACCGGATTTATGAGAGGTCGCCACG
	326KanScF1	GACTGGCTGCTATTGGGCGAAGTGC
Alternate screening primers, hb distal (unrecombined product screening)	455hbShLHAchF	ATCCTTGGCATGAAATCAGCTTACG
	456hbShLHAchR	TGCTGTGTGGATCCTACCCAGTTACG
	457hbShRHAchF	TGTCATTAAACCTAGTTAGAACAATCG
	458hbShRHAchR	TTTGGGCAACTTTAAGCCCAGACAC
Alternate screening primers, hb distal (unrecombined product screening)	459hbPrLHAchF	AACCGCAAATGACATTGAATCACTTC
	460hbPrLHAchR	TTAGTGGCGTGGTGTGGCCACAACG