DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Enhancer Function and Evolution in Drosophila Embryos

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:40046404

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Enhancer Function and Evolution in Drosophila Embryos

A dissertation presented

by

Benjamin James Vincent

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Genetics and Genomics

Harvard University

Cambridge, Massachusetts

April 2017

© 2017 Benjamin James Vincent

All rights reserved.

Dissertation Advisor: Dr. Angela DePace Benjamin James Vincent

Enhancer Function and Evolution in Drosophila Embryos

Abstract

As development proceeds from fertilization of a single egg cell, regulatory sequences called enhancers generate specific gene expression patterns in space and time. These patterns give rise to the many different cell types present in an adult organism. Changes in enhancer sequence underlie morphological differences between species as well as disease states in humans. It is critical, therefore, to determine how sequence variation within enhancers affect the overall expression pattern of their target gene.

The enhancers that control expression of the Drosophila gene *even-skipped* are an excellent system for investigating regulatory logic and evolution. Decades of genetic experiments have characterized many of the transcription factors that regulate these enhancers, and quantitative gene expression measurements in Drosophila embryos have enabled computational models of enhancer function. Despite these advantages, the failure of genetic reconstitution experiments suggest that we do not currently understand these enhancers well enough to build them. This shortcoming makes it difficult to identify the sequence features that are critical for enhancer function, or to reason about how enhancer sequence may evolve.

In this work, I challenged computational models of *even-skipped* enhancer function using genetic manipulations and quantitative measurements in *Drosophila melanogaster* embryos. I found that one expression pattern is generated by two different enhancers that respond differently to genetic perturbation of a particular transcription factor, *hunchback*. This work suggests that the same expression pattern can be built in different ways, and this feature may be important for the robustness of developmental expression patterns. I also found that the transcription factor Caudal acts as a Hunchback counter-repressor in the *even-skipped* stripe 2 enhancer, and that this interaction appears conserved in orthologous sequences. Finally, I present computational and experimental evidence that individual enhancers may diverge while

maintaining the expression pattern of the whole locus. As a whole, my work illustrates the astonishing plasticity of regulatory information, which is consistent with the rapid pace of regulatory evolution.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF MAIN FIGURES

LIST OF MAIN FIGURES (Continued)

LIST OF GENE NAMES AND SYMBOLS

even-skipped	*eve*
hunchback	*hb*
caudal	*cad*
bicoid	*bcd*
giant	*gt*
Krüppel	*Kr*
zelda	*zld*
knirps	*kni*
tailless	*tll*
dorsal	*dl*
twist	*twi*
Enhancer of split m4	*E(spl)m4*
shaven	*dPax2*
decapentaplegic	*dpp*
sloppy-paired	*slp*
huckebein	*hkb*
Signal-transducer and activator of transcription protein at 92E	*STAT92E*
snail	*sna*
short-gastrulation	*sog*
fushi-tarazu	*ftz*
zerknült	*zen*
capicua	*cic*
Dichaete	*D*
Trithorax-like	*Trl*
runt	*run*
nanos	*nos*
Mediator complex subunit 29/intersex	*MED29/ix*
moira	*mor*

LIST OF ABBREVIATIONS

TFs	Transcription Factors
UAS	Upstream Activating Sequence
eve2	*even-skipped* stripe 2 enhancer
AP	anterior-posterior
bp	base pair
kb	kilobase
eve3+7	*even-skipped* stripe 3+7 enhancer
eve2+7	*even-skipped* stripe 2+7 enhancer
WT	wild-type
shRNAs	short hairpin RNAs
AUC	area under the receiver operating characteristic curve
sna::hb	*hb* misexpressed along the ventral surface of the embryo using the *sna* promoter
SD	standard deviation
sog::gt	*gt* misexpressed along the lateral surface of the embryo using the s*og* promoter
PWM	position weight matrix
eve4+6	*even-skipped* stripe 4+6 enhancer
eve5	*even-skipped* stripe 5 enhancer
MCC	Matthews correlation coefficient
hbP2	*hunchback P2* enhancer
PIC	preinitiation complex
sna::kni	*kin* misexpressed along the ventral surface of the embryo using the *sna* promoter
CRISPRi	clustered regularly interspaced short palindromic repeats – interference
gRNAs	guide RNAs
dCas9	dead Cas9
YFP	yellow fluorescent protein

ACKNOWLEDGEMENTS

I am grateful to so many for their scientific and personal contributions to this work. The lessons I have learned from them have made me a better scientist and, more importantly, a better person.

Working in the DePace Lab has been one of the most edifying and fulfilling experiences of my life. Through her mentorship and leadership, Angela has taught me how live as a scientist with humility, integrity and empathy. She is the reason I choose to persist in science and education, and I will work as hard as I can for as long as I can to pay her investment forward. Under the supportive lab environment that Angela enables, my fellow lab members have become lifelong collaborators and friends. I'm especially grateful to Zeba Wunderlich for her years of compassionate leadership and quantitative help, and my rotation mentor, Tara Lydiard-Martin, for teaching me how to function in the lab. Meghan Bragdon deserves special thanks for collecting an astonishing fraction of the data contained in this thesis, as well as working with Kelly Eckenrode to define and perpetuate our early (and current) lab culture through art, food and drink. Finally, Max Staller has been my closest collaborator and a dear friend since I joined the lab. He was (and is) the patient older brother who always lets me boss him around, even though he taught me everything I know.

I'm also lucky to have mentored, recruited and collaborated with generous and talented students, including Clarissa Scholes, Kelly Biette and Olivia Foster. Each has donated countless hours to brainstorming and experimental help along with free therapy and party planning. The same can be said of all current lab members, including Jeehae Park, Javi Estrada, Lital Bentovim, Tim Harden and Anna Cha. In addition to current members, I've mentored many other graduate students and summer students who have helped me formulate and test the ideas in this work. They include Phoebe Hseih, Scott Gallagher, Chiara Ricci-Tam, Jonathan White, Adam Carte and Kate Shulgina. This dissertation is essentially a collection of rotation and summer projects, and to all of my mentees, it's as much yours as it is mine.

ACKNOWLEDGEMENTS (Continued)

Outside of my lab, I'm thankful for the larger scientific communities at Harvard, including the Systems Biology Department, the Program in Genetics and Genomics, and the Program in Biological and Biomedical Sciences. Each has given me exceptional mentors, colleagues, helpers and friends. I'm grateful to Kathy Buhl, Barb Grant and Jennie Epp for somehow being willing to ignore more pressing responsibilities every time I drop in on them unannounced. Systems Biology has given me too many friends to list all of them here, but some deserve special thanks, including Tom Hiscock, Zach Collins, Ian Swinburne and Sean Megason for dealing with constant, well-meaning harassment from your sister lab; Anna Green, Adam Riesselman, John Ingraham and Debbie Marks for providing a (comparatively) quiet space and many Scrabble-induced ego boosts; and Annie Jenney, Evi Van Italie, Marika Ziesack, Isaac Plant and Steph Hays and for bringing the party when I needed it. In addition to SysBio, I've found the PGG to be a wonderful resource as I've grown into a budding geneticist. I owe a great deal to Fred Winston, who has acted as the leader of that program as well as the chair of my dissertation and qualifying exam committees. Fred is a true educator who always pushed me to excel by giving me precise, helpful and critical feedback and never letting me cut corners; I hope I can pay his service forward in the future. In addition to Fred, Jesse Gray and Martha Bulyk were indispensable members of my DAC, and I'd like to thank Karen Adelman, Allon Klein and Juan Fuxman-Bass for their service on my defense committee. Finally, the BBS program has been my home and my family during my years in Boston. Kate Hodgins, Maria Bollinger, Danny Gonzales and Anne O'Shea helped me navigate the impenetrable Harvard bureaucracy with a refreshing side of advice and free mints; I'll miss them dearly. My heart truly lies with my BBS classmates, who gave me some of the best years of my life as we travelled the winding paths of graduate school and adulthood together. I'll remember emergency PGG meetings with Alex Meeske, Danielle Heller and Stephen Hinshaw; dance parties in the garage with Laura Smith, Stéphane Ricoult, Kal Tsanov, Silvia Escudero, Brian Pepe-Mooney and Tyler Ford; and

gay/straight movie nights with Kostadin Petrov and Mary Haas. From the beginning, Kaitlin Samocha and Niroshi Senaratne were at my side, my Pumbaa and my Mufasa, both fembots and fireworks, and I thank them most of all.

My mentors and friends outside of Harvard have enriched my life in countless ways. Caroline Goutte made me fall in love with genetics, and that love persists almost 10 years later. I worked through my awkward scientific adolescence with Dominic Poccia, who saw my academic potential despite my predilection for extreme procrastination. I'll always cherish my time in paradise with Dave Wadden and Heather Leonard, and while 21R Dane may collapse one day (soon), it will never be forgotten. I faced graduate school together (but separately) with Leigh Harris, Julian Damashek and Tracy Montgomery, who constantly remind me that the future of biological research is bright. I'm grateful to Geoff Giller, Mwanzaa Brown and Anna Meyendorff for keeping the varsity YGS team together in the offseason. It has given me great comfort to know that Deidra Montgomery is just on the other side of the big blue wall, and she has been there for me in the rare times I've faced the void in Boston. Finally, Mikey Yang has been a constant source of love and support in a time that, without him, could have been a constant source of stress and dread.

To close, I'd like to thank my family, and my oldest friends who have become family. I've stolen the best aspects of my own personality from Tracy Markowitz, as anyone who has met her can confirm. Amanda Korman has taught me to infuse my work with creativity and eloquence. Jen Birnkrant stuck with me when it was hardest to do so, and I am so lucky to have benefitted from her deep perception and wisdom. Paul Kunnath has been a brother to me for years, and I can't wait to watch him grow as a father to Ella. Finally, my brother-in-law Matthew, my stepfather Dan, my nephew Mac, my sister Rachel and my mother Doreen have taught me that, ironically, family has little to do with genetics and everything to do with love. I may not be able to draw my own complicated pedigree from memory, but I know who I love most in this world.

This dissertation is dedicated to the memory of my father, Harry Giles Vincent III, who taught me that I should stick to my dreams when the road seems long and lonely, and that I can cover a tremendous amount of ground going at my own speed.

CHAPTER 1: INTRODUCTION

The majority of this chapter was published as Vincent, B.J., Estrada, J. & DePace, A.H., 2016. The appeasement of Doug: a synthetic approach to enhancer biology. Integrative biology, 8(4), pp.475–484. Published by The Royal Society of Chemistry.

Author Contributions for Vincent et al. 2016:
B.J.V., J.E. and A.H.D. designed the experiments. B.J.V. performed experiments. J.E. processed raw image files for further analysis. B.J.V. analyzed the data. B.J.V. and A.H.D. wrote the manuscript.

Meet Bill and Doug

The beloved parable of Bill, Doug, and their quest to understand the inner workings of the automobile is a scientific fable in two parts, first from the perspective of a geneticist, and then from the perspective of a biochemist (Sullivan 1993). In "The Salvation of Doug," geneticist Bill investigates the function of individual automotive components by randomly selecting factory workers to incapacitate and observing the effects on the cars rolling off the lot. Biochemist Doug is forced to concede that Bill's method is more successful than his own attempts to reconstitute a car from its component fractions of glass, plastic and steel. However, in "The Demise of Bill," Doug finds his own success through painstaking efforts to uncover how individual 'pathways' function in his car, such as the fuel injection and electrical systems. Doug's hard won expertise helps him fix Bill's car each time it breaks because he understands how the parts work together. While written in a spirit of friendly competition and good fun, these parables teach serious lessons about the limitations of genetic and biochemical approaches. Genetics relies on a simple principle: break individual parts and catalog the resulting phenotypes. This approach is analogous to building a parts list and defining their interrelationships in relatively abstract epistatic terms that may or may not imply direct physical interactions. In contrast, biochemistry is about building systems from their component parts. This approach makes these abstract relationships mechanistically concrete. Biochemist Doug sums up this contrast quite nicely when he challenges geneticist Bill: "Now that you have learned so much, tell me how the car works."

Like automobiles, developmental enhancers are complex machines built of many parts. Enhancers are thought to be the main determinants of cell-type specific gene expression in space and time and play critical roles in evolution and human disease (Levine 2010; Carroll 2008; Zeitlinger & Stark 2010; Maurano et al. 2012). Geneticists like Bill have dominated the study of developmental enhancers for years. Genetic approaches are incredibly effective at identifying enhancer components – binding sites for the transcription factors (TFs) that regulate them. However, we cannot yet interpret regulatory sequence variation within enhancers because

it remains complicated to determine the contribution of each TF binding site. Unlike variation in coding regions, where a given change will generally have the same effect on protein sequence no matter the cell type or organism, disrupting or adding a TF binding site within an enhancer will only affect cells where the TF is expressed. Even in cells expressing a given TF, the phenotypic effects of binding site changes may be buffered or augmented by interactions within or between enhancers (Ludwig et al. 2000; Moses et al. 2006; Doniger & Fay 2007; Barolo 2012; Arnold et al. 2014). In the context of interpreting variation within regulatory DNA, Doug's challenge rings true in the ears of many experimentalists – we know a great deal about developmental enhancers, but we still can't tell you how they work.

How can we satisfy Doug? We argue that we need to take a "biochemical" approach and build enhancers from their component parts. The field has taken promising steps in this direction. Synthetic approaches have been successful in cell lines, where arrays of TF binding sites have been used to decipher their interactions and mechanisms of action (Gertz et al. 2009; Gertz & Cohen 2009; Keung et al. 2014; Stampfel et al. 2015). Arrays of binding sites can also drive patterned expression in embryos (Liberman & Stathopoulos 2009; Erceg et al. 2014). This approach forms the basis of the Gal4/Upstream Activating Sequence (UAS) system, which has become an indispensable part of the genetic toolkit in animal systems (Brand & Perrimon 1993; Ornitz et al. 1991; Halpern et al. 2008). Systematic characterization of binding site arrays in embryos is also an effective way to interrogate the grammatical rules of TF binding site arrangement (Fakhouri et al. 2010; Erceg et al. 2014).

To appease Doug, however, we must take a specific enhancer and reconstitute it from scratch. This has not yet been achieved. How can we design this experiment? At one extreme, we could use an unbiased empirical approach, in which we synthesize large libraries of random sequence and screen them for the activity we want. While eventually this approach would yield a new sequence that generates a target pattern, the combinatorial space is vast, the experiment expensive, and we would need many successes to derive any general rules that govern enhancer function. Instead, Doug would use rational design to reconstitute a developmental enhancer

from its component binding sites. This approach requires that we determine an enhancer's complete regulatory logic – the identity and function of all TFs that bind it. Once this criterion is met, we can identify and parameterize interactions between binding sites by altering their number, affinity and spacing. Armed with this information, we should be able to generate any number of functional sequences that produce the expression pattern of interest. This level of understanding may also allow us to design sequences that generate any expression pattern we want, which may be useful as experimental or therapeutic reagents (Goverdhana et al. 2005). Because our goal is reconstituting a particular enhancer, the gold standard would be to test whether a reconstituted enhancer can rescue deletions of its endogenous counterpart (Ludwig et al. 2005). This type of rescue experiment is now tractable in animal systems due to the recent explosion of genome editing technology, and would provide the best argument for convincing Doug that we understand how developmental enhancers work (Wright et al. 2016).

In this introduction, I discuss how close we are to appeasing Doug using illustrative examples primarily from Drosophila. I outline how regulatory logic is experimentally determined and the limitations inherent in these approaches. I also examine rules governing the interactions between TF binding sites. Finally, I discuss what we have learned from recent attempts to build developmental enhancers and how these results motivated the studies in this dissertation.

Classic Bill: Identifying regulators of *eve* stripe 2

Identifying the TFs that regulate an enhancer takes years of assiduous work in experimental genetics. While genetic methods have proven successful in many systems, they are fundamentally limited in their ability to characterize an enhancer's complete regulatory logic. To reveal these successes and limitations, we consider a single case study – the enhancer that controls the second stripe of the *even-skipped* (*eve*) expression pattern in Drosophila blastoderm embryos (*eve2*) (Figure 1.1). Even after 25 years of research on this 484 base pair (bp) sequence, we are still discovering putative *eve2* regulators using a combination of

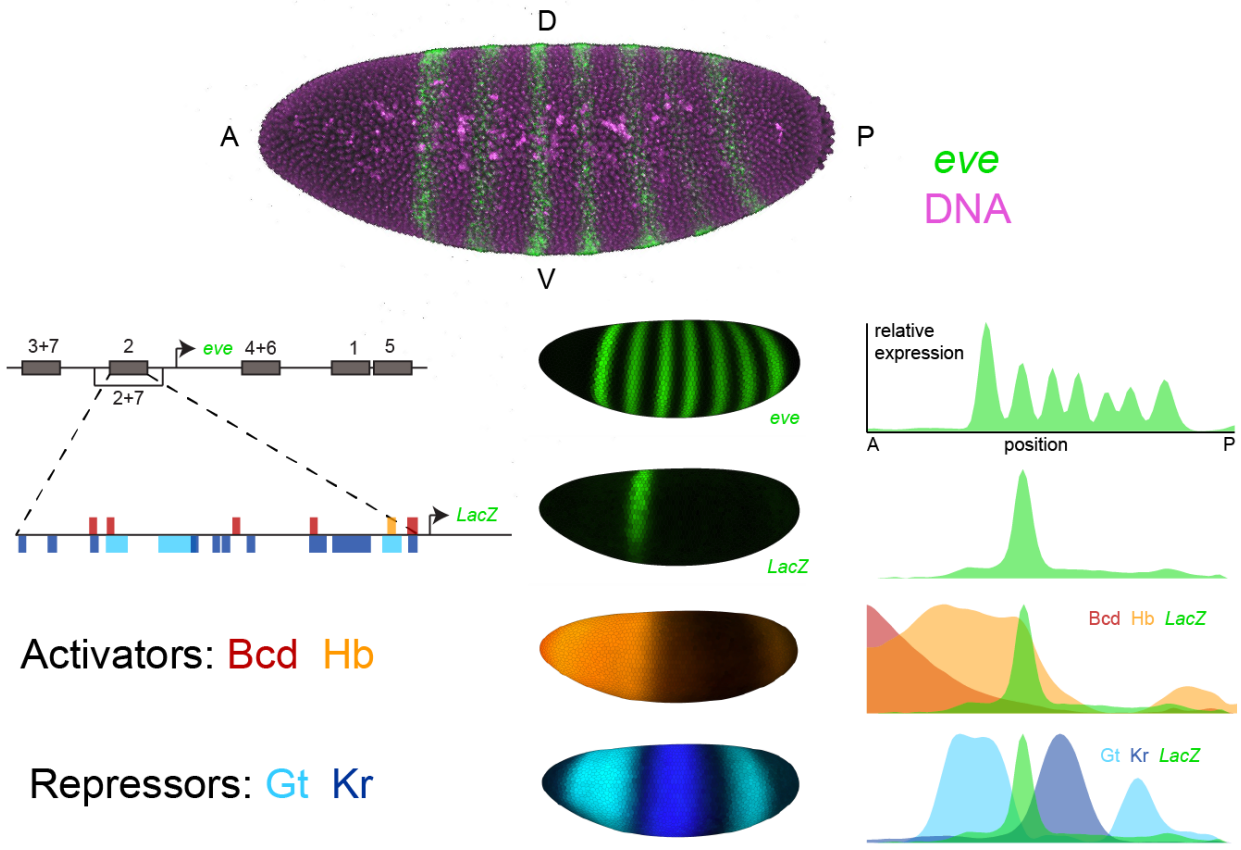


Figure 1.1: The *even-skipped* stripe 2 enhancer in *Drosophila melanogaster*. *even-skipped* (*eve*) is expressed in seven transverse stripes along the anterior-posterior (AP) axis of the blastoderm embryo. The *eve* locus contains five enhancers that generate expression during this stage, each of which generates one or two individual stripes. The *eve* stripe 2 enhancer (*eve2*) is a 484bp sequence that drives expression of the second *eve* stripe in transgenic animals when cloned into a *LacZ* reporter construct. *eve2* contains binding sites for the activators Bicoid (Bcd) and Hunchback (Hb) as well as the repressors Giant (Gt) and Krüppel (Kr). Top: Maximum intensity projection of a *Drosophila melanogaster* blastoderm embryo imaged with 2-photon confocal microscopy where *eve* mRNA has been stained using *in situ* hybridization and outer nuclei and yolk cells (DNA) have been labelled with SYTOX Green (Luengo Hendriks et al. 2006). Dorsal side (D); ventral side (V). Left: a schematic of the *eve* locus and annotated binding sites within *eve2* (Gallo et al. 2011) represented as bars. Color indicates TF identity. Middle: The 3D expression patterns of *eve*, *eve2*, and its regulators. Shown are visual renderings of quantitative expression data from gene expression atlases (Fowlkes et al. 2008; Staller et al. 2015) generated using PointCloudXplore (Rübel et al. 2006). Right: expression patterns of *eve*, *eve2*, and its regulators are displayed as line traces where relative expression along the lateral side of the embryo is plotted as a function of AP position. Plots were generated in MATLAB using the PointCloudToolbox (http://bdtnp.lbl.gov/Fly-Net/) and normalized by the maximum of each individual line trace.

developmental genetics, functional genomics and quantitative biology.

The story of *eve2* begins with the seminal screens that identified genes controlling embryonic patterning in *Drosophila melanogaster* embryos. Larval cuticles derived from *eve* mutants lacked every other abdominal segment, which placed *eve* in the 'pair rule' class of mutants affecting segmentation (Nüsslein-Volhard & Wieschaus 1980). Genetic mapping, *in situ* hybridization and immunostaining revealed the characteristic seven-stripe *eve* expression pattern, which beautifully mirrored the mutant phenotype (Harding et al. 1986; Frasch et al. 1987). Painstaking molecular dissection of the *eve* regulatory region with transgenic reporter constructs identified the *eve2* enhancer region as the sequence necessary (Goto et al. 1989) and sufficient (Small et al. 1991) to drive expression of the second *eve* stripe in embryos. Candidate regulators were identified by examining mutants that altered the *eve2* expression pattern (Frasch & Levine 1987), and binding site mutations in *eve2* confirmed that these interactions were direct (Small et al. 1991; Stanojevic et al. 1991; Small et al. 1992). These early experiments produced a working model of enhancer function that persists in cell and developmental biology classrooms to this day (Alberts et al. 2014; Gilbert 2013). In this model, *eve2* is regulated by direct binding of the anterior activators Bicoid and Hunchback as well as the gap gene repressors Giant and Krüppel (Figure 1.1).

Aficionados know that the story of *eve2* did not end there. First, the *eve2* cartoon shown in Figure 1.1 fails to explain the absence of expression in the region anterior to *giant*, where both *eve2* activators are present in high concentrations without a repressor. Genetic experiments identified additional direct and indirect regulators that account for this discrepancy (Andrioli et al. 2002). The ubiquitously expressed activator *zelda* (*zld*) also directly binds *eve2* (Struffi et al. 2011), and computational work has proposed additional roles for the activator *caudal* and the repressors *knirps* and *tailless* (Janssens et al. 2006; Kim et al. 2013). Finally, while the minimal *eve2* element is sufficient to drive expression of stripe 2, surrounding sequences are critical for robustness to genetic and environmental perturbation (Ludwig et al. 2011) as well as the response of *eve* stripe 7 to TF misexpression (Staller et al. 2015).

The *eve2* story demonstrates the power of genetic approaches for identifying and validating enhancer regulators. However, these methods are candidate-driven: they can only determine whether a given regulator has a role or not. They cannot tell us whether we know all relevant regulators; this question can only be resolved by reconstitution. In other words, while we have come far in characterizing the regulatory logic of *eve2*, genetic approaches cannot tell us how much further we have to go.

New Age Bill: Leveraging big data to define regulatory logic

The methods we use to define an enhancer's regulatory logic have transformed in the last decade. Classical genetics and molecular biology are now complemented by high-throughput assessment of chromatin state and TF binding. These functional genomic methods can complement classic genetic approaches in two ways. First, we can systematically identify candidate regulators by measuring binding genome-wide for all known TFs (Slattery et al. 2014; Araya et al. 2014; ENCODE Project Consortium 2012; Yue et al. 2014; Boyle et al. 2014). As with the classic approach, we must then validate these candidates by mutating sites using experimentally-derived TF binding preferences (Noyes et al. 2008; Newburger & Bulyk 2009; Portales-Casamar et al. 2009). However, we can also circumvent a candidate approach altogether by screening large libraries of mutated enhancer reporter constructs for expression and analyzing them retrospectively (Kwasnieski et al. 2012; Kheradpour et al. 2013; Canver et al. 2015).

These new methods carry their own limitations. First, it is difficult to capture spatial information in high-throughput. This limitation applies to measurements of TFs (inputs) as well as enhancer-driven expression patterns (outputs). For inputs, we can measure gene expression and chromatin accessibility in single cells (Buenrostro et al. 2013; Deng et al. 2014), but *in vivo* measurements of TF binding must be performed in homogenized tissue or cell populations. New single-cell methods may soon address this shortcoming (Rotem et al. 2015). For outputs, measurements of enhancer reporter libraries are typically performed in cell populations or in

disaggregated tissue (Gisselbrecht et al. 2013), although high-throughput imaging of whole animals after transfection is an emerging technology (Farley et al. 2015). Reporter libraries can only incorporate DNA fragments of a certain size, which is currently less than 200bp (Levo & Segal 2014). Therefore, an unbiased approach cannot be applied to larger enhancers. Many endogenous developmental enhancers exceed this size limit (Yue et al. 2014; Ho et al. 2014), including some that have been experimentally 'minimized,' such as our example, *eve2* (Small et al. 1992; Small et al. 1996; Fujioka et al. 1999). In addition, this experimental design requires reporters to drive transcription and thus does not detect fragments that repress or modulate the activity of other sequences (Swanson et al. 2010; Ludwig et al. 2011). This is part of the larger issue of assaying regulatory DNA in reporter constructs outside of its native context. Finally, making transgenic animals is slow compared to transfection of isolated cells. This constraint limits throughput to tens or hundreds of constructs by an individual, and thousands of constructs by consortia or larger projects (Jenett et al. 2012; Kvon et al. 2014). In the long-term, genomic technologies may advance to the point where we can determine the regulatory logic of developmental enhancers in an unbiased, data-centric way. For the moment, however, enhancer reconstitution through rational design provides a critical reciprocal test of an enhancer's complete regulatory logic.

Some general rules governing binding site interactions

A complete understanding of enhancer logic requires not only identifying all of its TF binding sites, but also characterizing their interactions. Here, we review some rules governing binding site interactions within enhancers, with an emphasis on case studies in Drosophila. Many other flagship systems have revealed mechanistically similar interactions between TFs (Lee et al. 1987; Dolan & Fields 1991; Shaulian & Karin 2002; Yamamoto & Gaynor 2004), and these have been reviewed elsewhere (Latchman 2010; Weingarten-Gabbay & Segal 2014).

Cooperativity, competition and context

Interactions between TFs are common in animal enhancers. TFs can exhibit cooperativity by influencing one another's binding directly through protein-protein interactions, or indirectly through effects on nucleosome occupancy or other cofactors (Spitz & Furlong 2012). In the simplest case, cooperativity occurs between binding sites for the same TF, as has been proposed for the activator Bicoid binding to the *hunchback* P2 enhancer (Driever et al. 1989). Cooperativity can also occur between different TFs, which has been suggested for Dorsal and Twist in Drosophila as well as their homologs in mice (Shirokawa & Courey 1997; Šošić et al. 2003). In the extreme, cooperative physical interactions between many TFs underlie the formation of a complex structure such as the enhanceosome that regulates the human interferon-beta gene during viral infection (Carey 1998). TFs can also physically interact with each other to interfere with binding. Repressors can directly compete with activators for overlapping sites within enhancers, as was originally proposed for regulation of *eve2* (Stanojevic et al. 1991). However, this does not appear to be the dominant mode of repressor function, as others mechanisms operate within animal enhancers (Spitz & Furlong 2012). Finally, bifunctional TFs can act as activators or repressors depending on context (Shore & Nasmyth 1987; Deng et al. 2010; Staller et al. 2015). Context-dependent function may be controlled by dimerization on DNA (Papatsenko & Levine 2008), interactions with neighboring TFs (Kim et al. 2013), or inputs from signalling cascades (Barolo & Posakony 2002).

Long-range interactions allow design flexibility

Cooperative binding and TF competition occur at short DNA length scales (tens of base pairs). However, TFs bound to enhancers also interact over longer length scales through other mechanisms. TFs can collaborate to expel nucleosomes, leading to indirect "collaborative" cooperativity on the length-scale of ~150bp (Mirny 2010). Activators have been proposed to functionally interact due to effects on nucleosomes or the core transcriptional machinery (Blau et al. 1996; Keung et al. 2014). Repressors can be classed by their range of influence on

neighboring activators, with short-range repressors operating at a scale of ~100bp and long-range repressors operating at a scale of more than 1 kb (Courey & Jia 2001); these different classes of repressors induce distinct chromatin changes when bound (Li & Arnosti 2011). Finally, cofactors and architectural proteins can mediate interactions between pieces of regulatory DNA at various length-scales ranging from intergenic to interchromosomal contacts (Nguyen & Bosco 2015).

Given that there are so many different types and length scales of TF interactions, it may seem hopeless to design an enhancer from scratch. However, the wide variety of mechanisms allows for flexible enhancer design with many ways to be successful (Arnosti & Kulkarni 2005). Indeed, orthologous enhancers drive similar expression patterns and are functionally conserved despite extensive turnover in TF binding sites (Ludwig et al. 2005; Hare et al. 2008; Liberman & Stathopoulos 2009). These general categories of interactions have been known for quite some time, and have been explored computationally, as we discuss below. However, we have not yet effectively leveraged these tools to rationally design enhancers that function *in vivo*.

Tools for rational design

Mathematical models can be useful tools for reasoning about complex systems. They formalize our assumptions and can generate testable predictions (Gunawardena 2014). In tackling the problem of building developmental enhancers, mathematical models can be useful in multiple ways. Regression models on imaging or sequencing data can suggest candidate regulators for a given expression pattern (Ilsley et al. 2013; Segal et al. 2003; Karlebach & Shamir 2008; Amit et al. 2011). More complex models based on fractional occupancy of TFs on enhancers have been used to analyze how expression patterns are generated (Kim et al. 2009). This type of model has been widely used for bacterial sequences (von Hippel et al. 1974; Ackers et al. 1982; Shea & Ackers 1985; Garcia et al. 2012; Brewster et al. 2014) and has been extended to developmental enhancers (Janssens et al. 2006; Zinzen et al. 2006; Segal et al. 2008; He et

al. 2010; White et al. 2012; Kim et al. 2013; Martinez et al. 2013; Samee & Sinha 2013; Samee & Sinha 2014).

Rules governing TF binding site arrangement have been uncovered by fitting mathematical models with quantitative measurements of the patterns driven by synthetic enhancers. The Arnosti lab used this approach to study the function of short-range repressors (Fakhouri et al. 2010). They constructed a set of synthetic enhancers where binding sites for short-range repressors were placed in configurations that altered their number, affinity and spacing relative to a common set of activator binding sites. By fitting parameters for each binding site within a fractional occupancy model, they uncovered a non-monotonic relationship between repressor efficiency and distance from the activator sites. They then used their results to dissect the relative contributions of individual binding sites within an endogenous enhancer. More recently, the Furlong lab used an analogous approach to study activator function in the *Drosophila melanogaster* mesoderm (Erceg et al. 2014). In this case, the authors measured expression patterns driven by homotypic and heterotypic binding site arrays using computationally designed spacer sequences that prevented the creation of additional sites. They found that some arrays of single activators were sufficient to drive expression in embryos; these arrays recapitulated the expression pattern of the activator. Some heterotypic arrays also drove expression; these arrays varied in their sensitivity to changes in the spacing and orientation of the sites. Finally, the authors used fractional site occupancy modeling to explore the potential impact of 'higher-order' cooperativity between binding site clusters. Both case studies demonstrate how TF interactions can be interrogated by combining careful measurements, highly controlled synthetic reporter constructs and computational modeling.

Even these systematic efforts will not satisfy Doug, who remains perhaps unreasonably demanding. Both studies employ the common practice of post hoc analysis, where models are used to analyze existing data to explain how binding site arrangement dictates expression patterns. We advocate for a complementary approach, where models are used at the outset to predict new, functional arrangements of binding sites, and these predictions are explicitly tested

by embedding these binding site configurations in 'neutral' DNA sequence. This 'model-first' approach would test the predictive power of models for synthetic applications, as well as the completeness of *cis*-regulatory rules governing enhancer function (Phillips 2015).

Instructive failures in enhancer reconstitution

Surprisingly few studies have attempted to reconstitute enhancer function from their component binding sites. One notable exception from the Barolo lab merits a detailed discussion (Johnson et al. 2008). For two candidate enhancers – the proneural enhancer of *Enhancer of split m4* and the *sparkling* enhancer of *shaven* (*dPax2*) – arrays of high affinity binding sites for key activators were insufficient to drive expression *in vivo*. These results contrast with the studies discussed above, where binding site arrays drive patterned expression in the Drosophila blastoderm, mesoderm and neurogenic ectoderm (Liberman & Stathopoulos 2009; Erceg et al. 2014). Furthermore, reconstituted enhancers from *dPax2* and *decapentaplegic* (*dpp*) that maintained the endogenous arrangement, affinity and spacing of known binding sites also failed to drive expression *in vivo*. Finally, an in-depth study of the *dppVM* enhancer revealed that regulatory sequence between known binding sites is essential for function. By combining mapped binding sites with evolutionarily conserved sequences, the authors generated an enhancer that drove expression in the correct location, but not to the correct level. This study represents one of the most comprehensive attempts to build a specific enhancer from its component parts, but even after the creation and measurement of over 25 constructs in transgenic animals, we still do not know the complete regulatory logic of the *dppVM* enhancer.

Legend has it that Steve Small attempted to build *eve2* from scratch following its initial characterization. This legend is true. Small fragments of *eve2* recapitulate TF-specific activation and repression in cell culture reporter constructs (Small et al. 1991). However, when these fragments were introduced into embryos, they failed to drive expression on their own (Figure 1.2A). Additional constructs tested whether function could be restored by multimerization of the elements or the introduction of a sequence that restored endogenous spacing between the them.

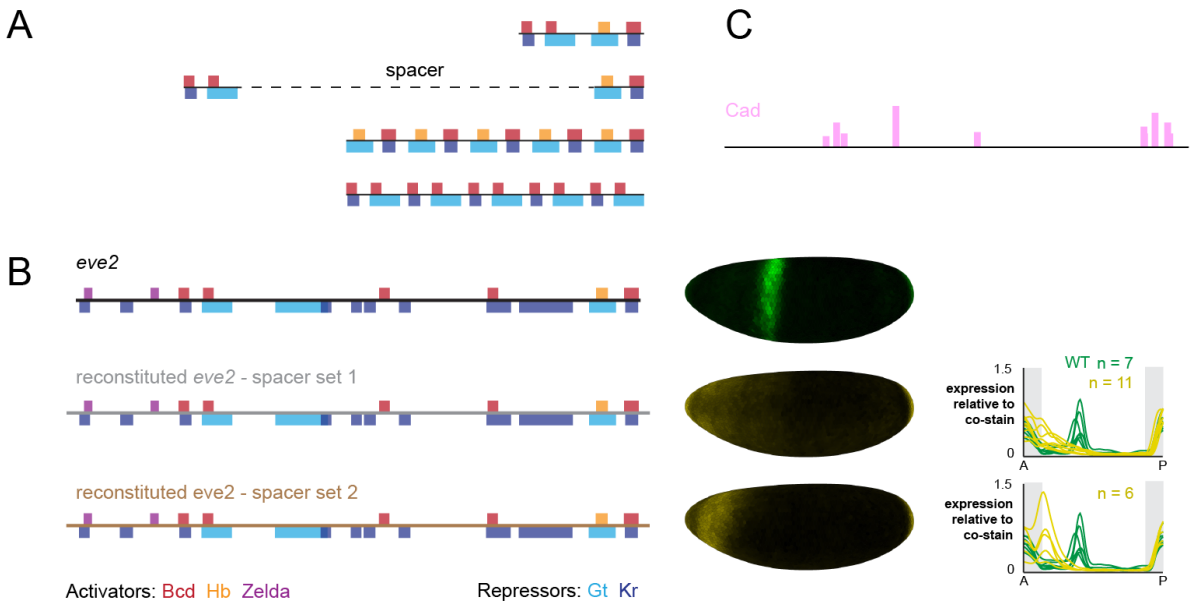


Figure 1.2: Reconstituted *eve2* enhancers do not drive correct expression patterns in embryos. (A) Experiments by Steve Small to reconstitute *eve2* function from enhancer fragments. A fusion of two ~50bp fragments which drives regulated expression in cell lines fails to drive expression of stripe 2 in embryos. Other constructs that tested the impact of maintaining endogenous spacing or multimerization also failed to drive expression of stripe 2 in embryos (Steve Small, personal communication). (B) Left: Bars indicate annotated TF binding sites in *eve2*, where color indicates TF identity (Gallo et al. 2011). Red: Bicoid; tangerine: Hunchback; violet: Zelda; light blue: Giant; dark blue: Krüppel. We used the computational tool SiteOut to design two sets of spacer sequences that did not create binding sites for these TFs or other known *eve* regulators (Estrada et al. 2016). We note that reconstituted enhancers disrupted a footprinted binding site for the repressor *sloppy paired 1* (*slp*), which was not contained in Redfly, as well as a single base pair within a footprinted Bicoid site (see Enhancer Sequences in Appendix A). However, this single base pair change does not disrupt any Bicoid sites predicted using binding preferences derived from bacterial 1-hybrid experiments (Noyes et al. 2008). Middle: The 3D expression patterns generated by wild-type *eve2* (green) and two reconstituted enhancers (yellow). Shown here are visual representations of gene expression atlases from embryos stained for *LacZ* mRNA and co-stained for *huckebein* (*hkb*) mRNA to normalize for *LacZ* levels between lines. *hkb* is expressed in the poles of the embryos, as indicated by grey boxes. Right: Expression patterns for all three constructs are displayed as line traces normalized to the *hkb* co-stain. (C) Binding sites for Caudal were predicted using PATSER software and displayed as pink vertical bars, where height is proportional to the PATSER score of the site (see Appendix A for Supplemental Methods).

These additional constructs also failed to recapitulate *eve2* expression in embryos, and the results were never published (Figure 1.2A, Steve Small, personal communication).

We recently attempted to reconstitute *eve2* again using some newly available resources. *eve2* contains many binding sites for known regulators including Zelda, a recently discovered activator that is critical for function of *bicoid*-dependent enhancers in *Drosophila melanogaster* embryos (Liang et al. 2008; Struffi et al. 2011; Xu et al. 2014, Gallo et al. 2011). We were able to design our constructs while rigorously controlling for the inadvertent creation of other binding sites using a computational tool, SiteOut (Estrada et al. 2016). We commercially synthesized two synthetic *eve* enhancers that included all annotated binding sites from the RedFly database at their endogenous locations relative to one another separated two different sets of spacer sequences (Figure 1.2B). We integrated both constructs into the same genomic location (attP2) using the phiC31 system (Groth et al. 2004) and measured their expression patterns at cellular resolution relative to a control construct containing wild-type (WT) *eve2* (Figure 1.2B). We have deposited quantitative data in individual embryos as well as an averaged gene expression atlas (Fowlkes et al. 2008) for these three constructs on figshare.com (see Supplemental Methods in Appendix A).

As Figure 1.2B shows, we were unsuccessful in reconstituting *eve2* function from its annotated sites in RedFly. However, our experiment 'failed well' in that it suggested some specific ways forward. Both synthetic constructs drive expression in the anterior of the embryo, outside the area of *eve* stripe 2. We hypothesize this ectopic expression is due to disruption of a footprinted site for the repressor *sloppy-paired 1* (*slp*), which was not included in the RedFly annotation (Andrioli et al. 2002). We also hypothesize that our synthetic construct is missing sites for one or more additional activators which are expressed in the region of eve stripe 2. Caudal (Cad) is an attractive candidate: Cad sites have been included in computational models of *eve2* flanking sequences (Janssens et al. 2006) and we can find predicted Cad sites within the boundaries of *eve2* using position weight matrices derived from bacterial 1-hybrid experiments (Figure 1.2C, Noyes et al. 2008). Our results also suggest that constructs including two different

sets of spacer sequences drive slightly different levels of expression, despite our attempts to control binding site content with SiteOut. These observations may help identify other regulating TFs or relevant sequence features in future experiments. Other explanations for our results are also possible. For example, the predicted binding sites that we used may not be accurate enough for this purpose. We included binding sites previously identified using DNA footprinting (Pollard et al. 2004; Gallo et al. 2011); but other data sources are available (Noyes et al. 2008; MacArthur et al. 2009), and not all binding preferences are the same across methods. Finally, we may have disrupted local sequence features outside annotated sites that affect binding, such as DNA shape or nucleosome positioning (Rohs et al. 2009; White et al. 2013; Barozzi et al. 2014; Weingarten-Gabbay & Segal 2014; Levo et al. 2015). All of these possibilities can be tested with additional constructs, and we hope that the quantitative data we provide here will be useful in designing additional experiments and testing computational models.

One thing is clear from all of these reconstitution experiments: we still have much to learn, even for the most well-characterized animal enhancers.

Doug needs us to share our negative results

We have argued that our understanding of developmental enhancers is incomplete, and experiments that attempt to reconstitute enhancer function from individual binding sites are an important test of how much we know thus far. In enhancer reconstitution, both successes and failures are informative. Successes confirm what we know, but do not always identify the sources of our success. Failures illuminate what we do not know, and converting a failure into a success pinpoints a new critical feature of enhancer design. Due to the explosion of functional genomic data generated from large consortia, the low cost of DNA synthesis and the availability of computational tools for rational design, we believe the time is right to perform reconstitution experiments for endogenous enhancers in many systems. We advocate for an iterative cycle for enhancer reconstitution, wherein computational models are used to predict active enhancer sequences, which are then synthesized and tested using precise quantitative techniques (Luengo

Hendriks et al. 2006; Raj & van Oudenaarden 2009; Little et al. 2013; Garcia et al. 2013; Crosetto et al. 2015). The results can be fed back into computational models to improve design for the next round. For a given enhancer, the first rounds of this cycle are likely to fail, as we have shown with our own constructs. However, in order for this approach to succeed, failures must be shared publicly so that the greater community can learn from the mistakes. Despite the known difficulties of publishing negative data (Fanelli 2012; Matosin et al. 2014), new online resources such as bioRXiv.org and figshare.com make this type of open data sharing possible. We believe that with a communal investment in synthetic approaches, we can finally meet Doug's challenge and learn the basic principles governing how enhancers work.

Motivation for current studies

The *eve2* story teaches us that we cannot characterize the complete regulatory logic of an enhancer with classical genetic techniques alone. Developmental enhancers are complex – they bind many TFs that may functionally interact with themselves and each other. Reconstitution experiments are an important test of regulatory logic, but the results are often negative and may not suggest specific ways forward. Computational models of enhancer function provide an important intermediate between these experimental extremes – they can incorporate many putative aspects of enhancer logic and make testable predictions. The ability of a model to fit experimental data does not necessarily imply that the model is correct – many different models can often fit the same dataset. However, models may be distinguished by follow-up experiments that might not have been obvious without quantitative predictions. In this way, it is useful to "model first" when designing experiments to investigate the regulatory logic of developmental enhancers (Phillips 2015).

In this dissertation, I describe how my co-authors and I used this "model first" approach to investigate the regulatory logic and evolution of *eve* enhancers. In Chapter 2, we distinguished between two models for the enhancer that controls *eve* stripes 3 and 7 (*eve3+7*). We found that the regulatory logic of the enhancer differs from the regulatory logic that controls

the endogenous stripes, and that this discrepancy is due to two enhancers in the *eve* locus that generate the same pattern in different ways. In Chapter 3, we distinguished between two models of Hb bifunctionality and found that Cad acts as a Hb counter-repressor in regulation of *eve* stripe 2. This interaction appears conserved in orthologous sequences and elaborates the textbook version of *eve* stripe 2 regulation. In Chapter 4, we used modeling to explore the evolution of the *eve* locus as a whole by predicting different ways to generate the entire seven stripe expression pattern. We validated these predictions by engineering endogenous enhancers, and we and developed methods to screen for diverged enhancer function in other insect species. In Chapter 5, I discuss on how we can build on these results to investigate the function of individuals TFs, the biochemical mechanisms underlying counter-repression, relationship between regulatory logic and phenotypic robustness, and compensatory evolution at the level of the locus.

References

Ackers, G.K., Johnson, A.D. & Shea, M.A., 1982. Quantitative model for gene regulation by lambda phage repressor. Proceedings of the National Academy of Sciences of the United States of America, 79(4), pp.1129–1133.

Alberts, B. et al., 2014. Molecular Biology of the Cell, Sixth Edition, Garland Science.

Amit, I., Regev, A. & Hacohen, N., 2011. Strategies to discover regulatory circuits of the mammalian immune system. Nature reviews immunology, 11(12), pp.873–880.

Andrioli, L.P.M. et al., 2002. Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. Development, 129(21), pp.4931–4940.

Araya, C.L. et al., 2014. Regulatory analysis of the C. elegans genome with spatiotemporal resolution. Nature, 512(7515), pp.400–405.

Arnold, C.D. et al., 2014. Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. Nature genetics, 46(7), pp.685–692.

Arnosti, D.N. & Kulkarni, M.M., 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? Journal of cellular biochemistry, 94(5), pp.890–898.

Barolo, S., 2012. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. BioEssays: news and reviews in molecular, cellular and developmental biology, 34(2), pp.135–141.

Barolo, S. & Posakony, J.W., 2002. Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. Genes & development, 16(10), pp.1167–1181.

Barozzi, I. et al., 2014. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. Molecular cell, 54(5), pp.844–857.

Blau, J. et al., 1996. Three functional classes of transcriptional activation domains. Molecular and cellular biology, 16(5), pp.2044–2055.

Boyle, A.P. et al., 2014. Comparative analysis of regulatory information and circuits across distant species. Nature, 512(7515), pp.453–456.

Brand, A.H. & Perrimon, N., 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development, 118(2), pp.401–415.

Brewster, R.C. et al., 2014. The transcription factor titration effect dictates level of gene expression. Cell, 156(6), pp.1312–1323.

Buenrostro, J.D. et al., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods, 10(12), pp.1213–1218.

Canver, M.C. et al., 2015. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature, 527(7577), pp.192–197.

Carey, M., 1998. The enhanceosome and transcriptional synergy. Cell, 92(1), pp.5–8.

Carroll, S.B., 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell, 134(1), pp.25–36.

Courey, A.J. & Jia, S., 2001. Transcriptional repression: the long and the short of it. Genes & development, 15(21), pp.2786–2796.

Crosetto, N., Bienko, M. & van Oudenaarden, A., 2015. Spatially resolved transcriptomics and beyond. Nature reviews genetics, 16(1), pp.57–66.

Deng, Q. et al., 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science, 343(6167), pp.193–196.

Deng, Z. et al., 2010. Yin Yang 1: a multifaceted protein beyond a transcription factor. Transcription, 1(2), pp.81–84.

Dolan, J.W. & Fields, S., 1991. Cell-type-specific transcription in yeast. Biochimica et biophysica acta, 1088(2), pp.155–169.

Doniger, S.W. & Fay, J.C., 2007. Frequent gain and loss of functional transcription factor binding sites. PLoS computational biology, 3(5), p.e99.

Driever, W., Thoma, G. & 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(6232), pp.363–367.

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), pp.57–74.

Erceg, J. et al., 2014. Subtle changes in motif positioning cause tissue-specific effects on robustness of an enhancer's activity. PLoS genetics, 10(1), p.e1004060.

Estrada, J. et al., 2016. SiteOut: An Online Tool to Design Binding Site-Free DNA Sequences. PloS one, 11(3), p.e0151740.

Fakhouri, W.D. et al., 2010. Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo. Molecular systems biology, 6, p.341.

Fanelli, D., 2012. Negative results are disappearing from most disciplines and countries. Scientometrics, 90(3), pp.891–904.

Farley, E.K. et al., 2015. Suboptimization of developmental enhancers. Science, 350(6258), pp. 325–328.

Fowlkes, C.C. et al., 2008. A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell, 133(2), pp.364–374.

Frasch, M. et al., 1987. Characterization and localization of the even-skipped protein of Drosophila. The EMBO journal, 6(3), pp.749–759.

Frasch, M. & Levine, M., 1987. Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in Drosophila. Genes & development, 1(9), pp.981–995.

Fujioka, M. et al., 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(11), pp.2527–2538.

Gallo, S.M. et al., 2011. REDfly v3. 0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic acids research, 39(suppl 1), pp.D118–D123.

Garcia, H.G. et al., 2012. Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. Cell reports, 2(1), pp.150–161.

Garcia, H.G. et al., 2013. Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Current biology, 23(21), pp.2140–2145.

Gertz, J. & Cohen, B.A., 2009. Environment-specific combinatorial cis-regulation in synthetic promoters. Molecular systems biology, 5, p.244.

Gertz, J., Siggia, E.D. & Cohen, B.A., 2009. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature, 457(7226), pp.215–218.

Gilbert, S.F., 2013. Developmental Biology, Tenth Edition 10 edition., Sinauer Associates, Inc.

Gisselbrecht, S.S. et al., 2013. Highly parallel assays of tissue-specific enhancers in whole Drosophila embryos. Nature methods, 10(8), pp.774–780.

Goto, T., Macdonald, P. & Maniatis, T., 1989. Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell, 57(3), pp.413–422.

Goverdhana, S. et al., 2005. Regulatable gene expression systems for gene therapy applications: progress and future challenges. Molecular therapy: the journal of the American Society of Gene Therapy, 12(2), pp.189–211.

Groth, A.C. et al., 2004. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics, 166(4), pp.1775–1782.

Gunawardena, J., 2014. Models in biology: “accurate descriptions of our pathetic thinking.” BMC biology, 12(1), p.29.

Halpern, M.E. et al., 2008. Gal4/UAS transgenic tools and their application to zebrafish. Zebrafish, 5(2), pp.97–110.

Harding, K. et al., 1986. Cross-regulatory interactions among pair-rule genes in Drosophila. Science, 233(4767), pp.953–959.

Hare, E.E. et al., 2008. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS genetics, 4(6), p.e1000106.

He, X. et al., 2010. Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. PLoS computational biology, 6(9), p.e1000935.

von Hippel, P.H. et al., 1974. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. Proceedings of the National Academy of Sciences of the United States of America, 71(12), pp.4808–4812.

Ho, J.W.K. et al., 2014. Comparative analysis of metazoan chromatin organization. Nature, 512(7515), pp.449–452.

Ilsley, G.R. et al., 2013. Cellular resolution models for even skipped regulation in the entire Drosophila embryo. eLife, 2, p.e00522.

Janssens, H. et al., 2006. Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. Nature genetics, 38(10), pp.1159–1165.

Jenett, A. et al., 2012. A GAL4-driver line resource for Drosophila neurobiology. Cell reports, 2(4), pp.991–1001.

Johnson, L.A. et al., 2008. Reverse-Engineering a Transcriptional Enhancer: A Case Study in Drosophila. Tissue engineering. Part A, 14(9), pp.1549–1559.

Karlebach, G. & Shamir, R., 2008. Modelling and analysis of gene regulatory networks. Nature reviews. Molecular cell biology, 9(10), pp.770–780.

Keung, A.J. et al., 2014. Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. Cell, 158(1), pp.110–120.

Kheradpour, P. et al., 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. Genome research, 23(5), pp.800–811.

Kim, A.-R. et al., 2013. Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic. PLoS genetics, 9(2), p.e1003243.

Kim, H.D. et al., 2009. Transcriptional regulatory circuits: predicting numbers from alphabets. Science, 325(5939), pp.429–432.

Kvon, E.Z. et al., 2014. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature, 512(7512), pp.91–95.

Kwasnieski, J.C. et al., 2012. Complex effects of nucleotide variants in a mammalian cis-regulatory element. Proceedings of the National Academy of Sciences, 109(47), pp. 19498–19503.

Latchman, D.S., 2010. Eukaryotic Transcription Factors 5 edition., Academic Press.

Lee, W. et al., 1987. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature, 325(6102), pp.368–372.

Levine, M., 2010. Transcriptional enhancers in animal development and evolution. Current biology, 20(17), pp.R754–63.

Levo, M. et al., 2015. Unraveling determinants of transcription factor binding outside the core binding site. Genome research, 25(7), pp.1018–1029.

Levo, M. & Segal, E., 2014. In pursuit of design principles of regulatory sequences. Nature reviews. Genetics, 15(7), pp.453–468.

Liang, H.-L. et al., 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature, 456(7220), pp.400–403.

Liberman, L.M. & Stathopoulos, A., 2009. Design flexibility in cis-regulatory control of gene expression: synthetic and comparative evidence. Developmental biology, 327(2), pp. 578–589.

Li, L.M. & Arnosti, D.N., 2011. Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes. Current biology, 21(5), pp.406–412.

Little, S.C., Tikhonov, M. & Gregor, T., 2013. Precise developmental gene expression arises from globally stochastic transcriptional activity. Cell, 154(4), pp.789–800.

Ludwig, M.Z. et al., 2011. Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. PLoS genetics, 7(11), p.e1002364.

Ludwig, M.Z. et al., 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature, 403(6769), pp.564–567.

Ludwig, M.Z. et al., 2005. Functional evolution of a cis-regulatory module. PLoS biology, 3(4), p.e93.

Luengo Hendriks, C.L. et al., 2006. Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. Genome biology, 7(12), p.R123.

MacArthur, S. et al., 2009. Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome biology, 10(7), p.R80.

Martinez, C.A. et al., 2013. A synthetic biology approach to the development of transcriptional regulatory models and custom enhancer design. Methods, 62(1), pp.91–98.

Matosin, N. et al., 2014. Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. Disease models & mechanisms, 7(2), pp. 171–173.

Maurano, M.T. et al., 2012. Systematic localization of common disease-associated variation in regulatory DNA. Science, 337(6099), pp.1190–1195.

Mirny, L.A., 2010. Nucleosome-mediated cooperativity between transcription factors. Proceedings of the National Academy of Sciences of the United States of America, 107(52), pp.22534–22539.

Moses, A.M. et al., 2006. Large-scale turnover of functional transcription factor binding sites in Drosophila. PLoS computational biology, 2(10), p.e130.

Newburger, D.E. & Bulyk, M.L., 2009. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. Nucleic acids research, 37(Database issue), pp.D77–82.

Nguyen, H.Q. & Bosco, G., 2015. Gene Positioning Effects on Expression in Eukaryotes. Annual review of genetics, 49, pp.627–646.

Noyes, M.B. et al., 2008. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nucleic acids research, 36(8), pp.2547–2560.

Nüsslein-Volhard, C. & Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature, 287(5785), pp.795–801.

Ornitz, D.M., Moreadith, R.W. & Leder, P., 1991. Binary system for regulating transgene expression in mice: targeting int-2 gene expression with yeast GAL4/UAS control elements. Proceedings of the National Academy of Sciences of the United States of America, 88(3), pp.698–702.

Papatsenko, D. & Levine, M.S., 2008. Dual regulation by the Hunchback gradient in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 105(8), pp.2901–2906.

Phillips, R., 2015. Theory in Biology: Figure 1 or Figure 7? Trends in cell biology, 25(12), pp. 723–729.

Pollard, D.A. et al., 2004. Benchmarking tools for the alignment of functional noncoding DNA. BMC bioinformatics, 5, p.6.

Portales-Casamar, E. et al., 2010. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic acids research, 38(Database issue), pp.D105–10.

Raj, A. & van Oudenaarden, A., 2009. Single-molecule approaches to stochastic gene expression. Annual review of biophysics, 38, pp.255–270.

Rohs, R. et al., 2009. The role of DNA shape in protein-DNA recognition. Nature, 461(7268), pp. 1248–1253.

Rotem, A. et al., 2015. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nature biotechnology, 33(11), pp.1165–1172.

Rübel, O. et al., 2006. PointCloudXplore: Visual Analysis of 3D Gene Expression Data Using Physical Views and Parallel Coordinates. In EuroVis. pp. 203–210.

Samee, A.H. & Sinha, S., 2013. Evaluating thermodynamic models of enhancer activity on cellular resolution gene expression data. Methods, 62(1), pp.79–90.

Samee, M.A.H. & Sinha, S., 2014. Quantitative modeling of a gene's expression from its intergenic sequence. PLoS computational biology, 10(3), p.e1003467.

Segal, E. et al., 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nature genetics, 34(2), pp.166–176.

Segal, E. et al., 2008. Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature, 451(7178), pp.535–540.

Shaulian, E. & Karin, M., 2002. AP-1 as a regulator of cell life and death. Nature cell biology, 4(5), pp.E131–6.

Shea, M.A. & Ackers, G.K., 1985. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. Journal of molecular biology, 181(2), pp.211–230.

Shirokawa, J.M. & Courey, A.J., 1997. A direct contact between the dorsal rel homology domain and Twist may mediate transcriptional synergy. Molecular and cellular biology, 17(6), pp.3345–3355.

Shore, D. & Nasmyth, K., 1987. Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell, 51(5), pp.721–732.

Slattery, M. et al., 2014. Diverse patterns of genomic targeting by transcriptional regulators in Drosophila melanogaster. Genome research, 24(7), pp.1224–1235.

Small, S. et al., 1991. Transcriptional regulation of a pair-rule stripe in Drosophila. Genes & development, 5(5), pp.827–839.

Small, S., Blair, A. & Levine, M., 1992. Regulation of even-skipped stripe 2 in the Drosophila embryo. The EMBO journal, 11(11), pp.4047–4057.

Small, S., Blair, A. & Levine, M., 1996. Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. Developmental biology, 175(2), pp.314–324.

Šošić, D. et al., 2003. Twist regulates cytokine gene expression through a negative feedback loop that represses NF-kappaB activity. Cell, 112(2), pp.169–180.

Spitz, F. & Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. Nature reviews genetics, 13(9), pp.613–626.

Staller, M.V., Fowlkes, C.C., et al., 2015. A gene expression atlas of a bicoid-depleted Drosophila embryo reveals early canalization of cell fate. Development, 142(3), pp.587–596.

Staller, M.V., Vincent, B.J., et al., 2015. Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 112(3), pp.785–790.

Stampfel, G. et al., 2015. Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature, 528(7580), pp.147–151.

Stanojevic, D., Small, S. & Levine, M., 1991. Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science, 254(5036), pp.1385–1387.

Struffi, P. et al., 2011. Combinatorial activation and concentration-dependent repression of the Drosophila even skipped stripe 3+7 enhancer. Development, 138(19), pp.4291–4299.

Sullivan, W.T., 1993. The salvation of Doug. Generations, 1(3), pp.1439–1443.

Swanson, C.I., Evans, N.C. & Barolo, S., 2010. Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Developmental cell, 18(3), pp.359–370.

Weingarten-Gabbay, S. & Segal, E., 2014. The grammar of transcriptional regulation. Human genetics, 133(6), pp.701–711.

White, M.A. et al., 2012. A model of spatially restricted transcription in opposing gradients of activators and repressors. Molecular systems biology, 8, p.614.

White, M.A. et al., 2013. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. Proceedings of the National Academy of Sciences, 110(29), pp.11952–11957.

Wright, A.V., Nuñez, J.K. & Doudna, J.A., 2016. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell, 164(1-2), pp.29–44.

Xu, Z. et al., 2014. Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. Genes & development, 28(6), pp.608–621.

Yamamoto, Y. & Gaynor, R.B., 2004. IκB kinases: key regulators of the NF-κB pathway. Trends in biochemical sciences, 29(2), pp.72–79.

Yue, F. et al., 2014. A comparative encyclopedia of DNA elements in the mouse genome. Nature, 515(7527), pp.355–364.

Zeitlinger, J. & Stark, A., 2010. Developmental gene regulation in the era of genomics. Developmental biology, 339(2), pp.230–239.

Zinzen, R.P. et al., 2006. Computational models for neurogenic gene expression in the Drosophila embryo. Current biology, 16(13), pp.1358–1365.

CHAPTER 2: SHADOW ENHANCERS ENABLE HUNCHBACK BIFUNCTIONALITY IN THE DROSOPHILA EMBRYO

Max V. Staller, Ben J. Vincent, Zeba Wunderlich, Meghan D.J. Bragdon, Tara Lydiard-Martin, Jonathan A. White, Javier Estrada and Angela H. DePace

The majority of this chapter was published in Staller, M.V. et al., 2015. Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 112(3), pp.785–790.

Author Contributions for Staller, et al. 2015:
M.V.S., B.J.V., and A.H.D. designed research; M.V.S., B.J.V., M.D.J.B., and T.L.-M. performed research; M.V.S., B.J.V., Z.W., and J.E. analyzed data; and M.V.S., B.J.V., and A.H.D. wrote the paper.

A portion of this chapter was published in Wunderlich, Z. et al. 2015. Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. Cell reports, 12(11), pp. 1740–1747.

Author Contributions for Wunderlich et al. 2016:
Z.W., B.J.V., and A.H.D. designed the study; Z.W., M.D.J.B., B.J.V., and J.A.W. collected the data; Z.W. and J.E. processed the data; Z.W. analyzed the data; and Z.W. and A.H.D. wrote the paper, with input from all authors.

Abstract

Hunchback (Hb) is a bifunctional transcription factor that activates and represses distinct enhancers. Here, we investigate the hypothesis that Hb can activate and repress the same enhancer. Computational models predicted that Hb bifunctionally regulates the *even-skipped* (*eve*) stripe 3+7 enhancer (*eve3+7*) in Drosophila blastoderm embryos. We measured and modeled *eve* expression at cellular resolution under multiple genetic perturbations and found that the *eve3+7* enhancer could not explain endogenous *eve* stripe 7 behavior. Instead, we found that *eve* stripe 7 is controlled by two enhancers: the canonical *eve3+7* enhancer and a sequence encompassing the minimal *eve* stripe 2 enhancer (*eve2+7*). Hb bifunctionally regulates *eve* stripe 7, but it executes these two activities on different pieces of regulatory DNA – it activates the *eve2+7* enhancer and represses the *eve3+7* enhancer. We also found that Hb bifunctionally regulates the gap gene *Krüppel* by activating the proximal enhancer and repressing the distal enhancer. These examples demonstrate that "shadow enhancers" can use different regulatory logic to create the same pattern.

Introduction

Transcription factors (TFs) are typically categorized as activators or repressors, but many TFs can act bifunctionally by both activating and repressing target genes (Struhl et al. 1992; Deng et al. 2010; Di Stefano et al. 2014; Shore & Nasmyth 1987). Changes in TF activity can result from post-translational modifications, protein cleavage, or translocation of cofactors into the nucleus (Lynch et al. 2011; Briscoe & Thérond 2013; Hori et al. 2013). However, in cases where a TF activates and represses genes in the same cells, bifunctionality is controlled by enhancer sequences, which are responsible for tissue-specific gene expression (Levine et al. 2014). For example, in Drosophila, Dorsal activates genes when it binds to enhancers alone or near Twist (Shirokawa & Courey 1997; Jiang & Levine 1993) but represses genes when it binds near other TFs (Dubnicoff et al. 1997; Flores-Saaib et al. 2001; Ratnaparkhi et al. 2006). The

DNA sequence of a TF's binding site can also alter TF activity [e.g., the glucocorticoid receptor (Meijsing et al. 2009; Latchman 2001)]. Identifying how the activity of bifunctional TFs is controlled will be critical for inferring accurate gene regulatory networks from genomic data (Kim et al. 2009).

Here, we investigate how TF bifunctionality is controlled using a classic example: the Drosophila gene *hunchback* (*hb*) (Struhl et al. 1992; Schulz & Tautz 1994; Zuo et al. 1991). Hb both activates and represses *even-skipped* (*eve*) by acting on multiple enhancers. Hb activates *eve* stripes 1 and 2 and represses stripes 4, 5, and 6 (Small et al. 1991; Arnosti et al. 1996; Fujioka et al. 1999; Clyde et al. 2003). Computational models from us and others support the hypothesis that Hb both activates and represses the enhancer that controls *eve* stripes 3 and 7 (*eve3+7*) (Figure 2.1; Clyde et al. 2003; Small et al. 1996; Ilsley et al. 2013; Papatsenko & Levine 2008).

In contrast to others, our computational models of *eve3+7* activity do not include regulatory DNA sequence (He et al. 2010; Kazemian et al. 2010; Janssens et al. 2006; Sherman & Cohen 2012; Kim et al. 2013). Instead, our modeling approach uses regression to identify the activators and repressors that control a given pattern; we refer to the identity and role of the regulators as "regulatory logic." Modeling regulatory logic without including DNA sequence enables a powerful strategy to dissect gene regulation in a complex locus. We can compare the regulatory logic of an enhancer reporter pattern to that of the corresponding portion of the endogenous pattern to determine whether the annotated enhancer contains all relevant regulatory DNA.

Here, we tested the hypothesis that Hb bifunctionally regulates *eve3+7*. We measured the endogenous *eve* expression pattern and the expression pattern driven by an *eve3+7* enhancer reporter at cellular resolution under multiple genetic perturbations. We then used these data to challenge two computational models of *eve3+7* activity. Hb acts only as a repressor in one model, but acts as both an activator and a repressor in the other (Figure 2.1; Ilsley et al. 2013). Our modeling indicated that *eve3+7* and the endogenous locus use different regulatory

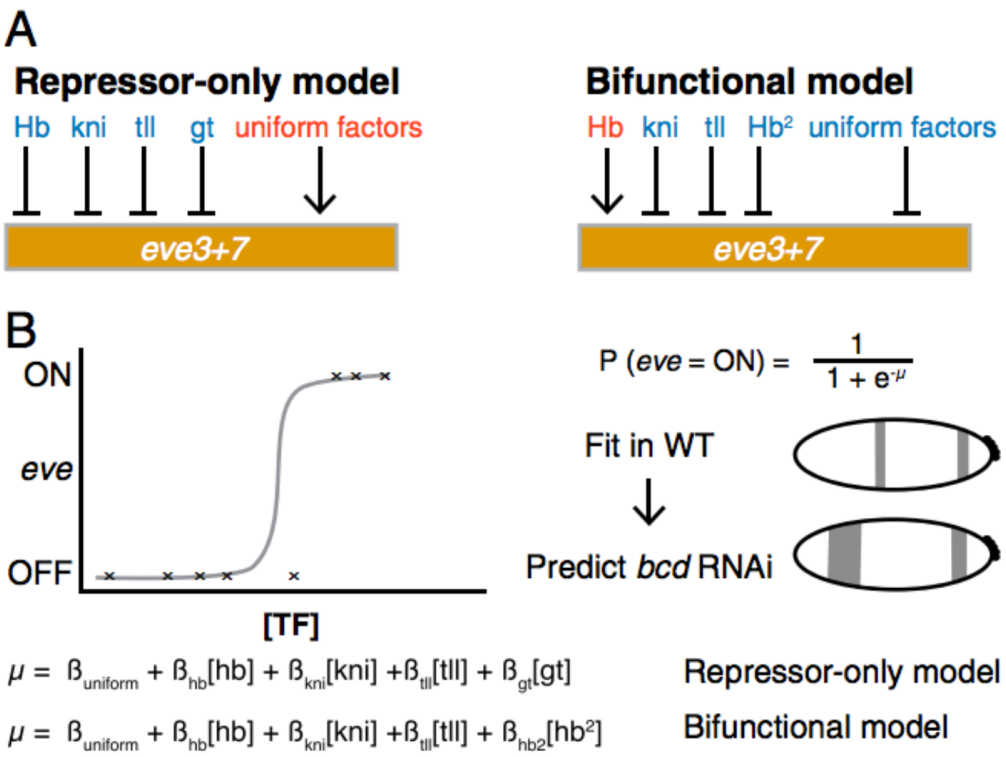


Figure 2.1: The repressor-only and bifunctional models formalize two alternative regulator sets for *eve* stripes 3 and 7. (A) The repressor-only model includes repression (blue) by Hb, *knirps* (*kni*), *giant* (*gt*), and *tailless* (*tll*) and activation (red) by a constant term that represents spatially uniform factors. The bifunctional model includes activation by a linear Hb term and repression by a quadratic Hb term, *kni*, *tll*, and uniform factors. (B) A schematic of the logistic regression framework. Logistic regression calculates the probability the target will be ON based on a linear combination of the concentrations of regulators (μ). We fit models in WT and use the perturbed regulator gene expression patterns to predict the perturbed *eve* patterns in embryos depleted of the maternal gene *bicoid* (*bcd* RNAi embryos).

logic to position stripe 7. Specifically, *eve3+7* is only repressed by Hb, whereas the endogenous stripe 7 is both activated and repressed. We show that an additional sequence is activated by Hb and contributes to the regulation of *eve* stripe 7 (Small et al. 1991; Small et al. 1996; Janssens et al. 2006; Goto et al. 1989; Harding et al. 1989; Hare et al. 2008). Thus, *eve* stripe 7 is controlled by a pair of shadow enhancers – separate sequences in a locus that drive overlapping spatiotemporal patterns (Barolo 2012). These shadow enhancers respond to Hb in opposite ways and use different regulatory logic.

Results

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

To determine whether Hb bifunctionally regulates *eve3+7*, we compared the endogenous *eve* pattern to the pattern driven by a β-galactosidase (*lacZ*) reporter construct in two genetic backgrounds (Figure 2.2A; Supplemental Figure 2.1; Supplemental Figure 2.2). We refer to these data throughout the manuscript as "the *eve3+7* reporter pattern" and "the endogenous pattern." We examined both wild-type (WT) embryos and embryos laid by females expressing short hairpin RNAs (shRNAs) against *bicoid* (*bcd* RNAi embryos), where expression of all of the regulators, especially Hb, is perturbed (Supplemental Figure 2.3; Supplemental Figure 2.4; Staller et al. 2015). We measured expression patterns quantitatively at cellular resolution using *in situ* hybridization, two-photon microscopy, and an automated image processing toolkit (Luengo Hendriks et al. 2006; Keränen et al. 2006). We averaged data from many embryos into gene expression atlases (Fowlkes et al. 2008). Importantly, the *eve3+7* reporter pattern results from the activity of *eve3+7* alone, whereas the endogenous pattern integrates the whole locus.

Our high-resolution measurements revealed discrepancies between the endogenous pattern and the *eve3+7* reporter pattern. In WT embryos, the *eve3+7* reporter pattern overlaps the corresponding endogenous *eve* stripes, but these stripes are broader, have uneven levels, and the peaks lie posterior to the endogenous peaks (Figure 2.2). These discrepancies were more pronounced in *bcd* RNAi embryos than in WT embryos, especially for the anterior stripe (Figure 2.2 D–F). When we tested reporters for other *eve* enhancers, we also found discrepancies between reporter patterns and the endogenous pattern (Supplemental Figures 2.1 and 2.2).

To test whether the discrepancies between the *eve3+7* reporter pattern and the endogenous pattern resulted from differences in the *eve* and *lacZ* transcripts, we measured the expression driven by a reporter encompassing the entire *eve* locus where the coding sequence had been replaced with *lacZ* (*eve* locus reporter, a gift from Miki Fujioka, Thomas Jefferson University, Philadelphia). In both WT and *bcd* RNAi embryos, the locus reporter pattern was

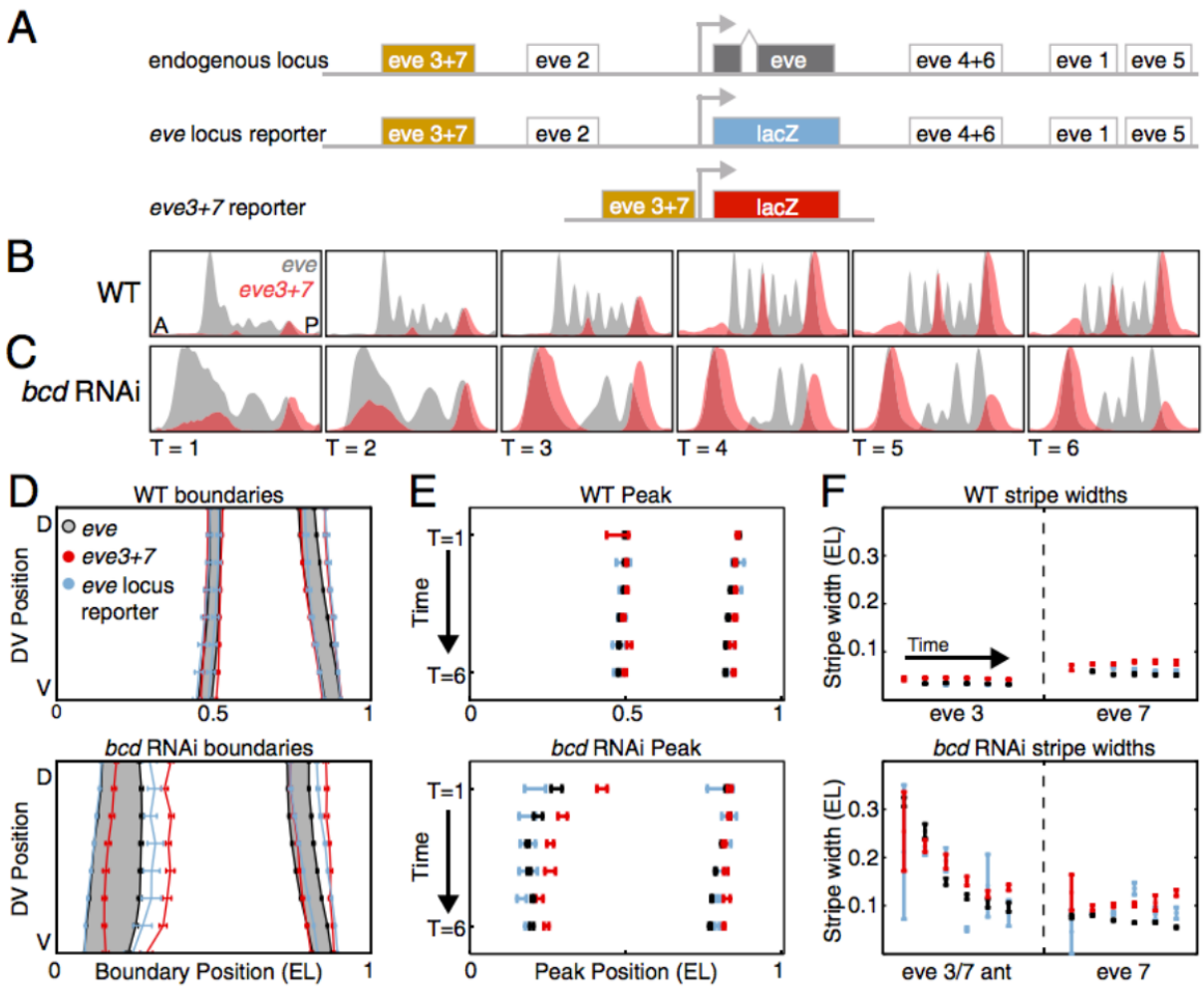


Figure 2.2: The *eve3+7* reporter pattern differs from the endogenous pattern. (A) The *eve* locus contains five annotated primary stripe enhancers. The endogenous pattern integrates the whole locus. The *eve* locus reporter pattern also integrates the whole locus, but the transcript is the same as the *eve3+7* reporter construct. The *eve3+7* reporter construct isolates the activity of the annotated enhancer sequence. (B) WT expression patterns are represented as line traces for a lateral strip of the embryo where AP position is plotted on the x axis with expression level on the y axis. Endogenous *eve* pattern (gray), *eve3+7* reporter pattern (red). The reporter pattern was manually scaled to match the level of the endogenous pattern; this scaling highlights differences in the position of expression. (C) Line traces in *bcd* RNAi embryos. Data presented as in B. (D) The boundaries of the endogenous pattern (gray), the *eve3+7* reporter pattern (red), and the *eve* locus reporter pattern (blue) at time point 3. All error bars are the standard error of the mean. The *eve* locus reporter pattern is more faithful to the endogenous pattern than the *eve3+7* reporter pattern, especially in the anterior of *bcd* RNAi embryos. The endogenous pattern is shaded for visual clarity. (E) Peak positions of stripes 3 and 7, calculated from the line traces in B and C. The *eve3+7* reporter pattern shows better agreement to the endogenous pattern in WT than in *bcd* RNAi embryos. (F) Stripe widths were calculated from the inflection point of the line traces in B and C. The *eve3+7* reporter pattern is wider than the corresponding endogenous pattern.

more faithful to the endogenous pattern in terms of stripe peak positions and widths (Figure 2.2; Supplemental Figures 2.1 and 2.2). Remaining differences between the endogenous and locus reporter patterns arise from differences in the transcripts. Differences between the locus reporter and the *eve3+7* reporter patterns may arise from regulatory DNA outside of *eve3+7*. Together, these data suggest that the *eve3+7* reporter construct may not contain all of the regulatory DNA that controls the expression of *eve* stripes 3 and 7.

Different computational models capture the behavior of the endogenous locus and the enhancer reporter after Hb perturbation.

We used computational models to dissect discrepancies between the *eve3+7* reporter pattern and the endogenous pattern. With our collaborators, we previously modeled the regulation of the endogenous *eve* stripes 3 and 7 in WT embryos and simulated genetic perturbations that mimicked published experimental data (Ilsley et al. 2013). These models use logistic regression to directly relate the concentrations of input regulators to output expression in single cells. We constructed two models that together test the hypothesis that Hb both activates and represses *eve* stripes 3 and 7. In the "repressor-only" model (the linear logistic model in Ilsley et al. 2013), Hb has one parameter and only represses. In the "bifunctional" model (the quadratic logistic model in Ilsley et al. 2013) Hb has two parameters that allow it to both activate and repress (Figure 2.1). Both models performed equally well in WT embryos, but we favored the bifunctional model because it predicted the effect of a genetic perturbation. At that time, cellular resolution data for the *eve3+7* reporter pattern were not available, so we used a standard assumption to interpret the models: the endogenous expression of *eve* stripes 3 and 7 could be attributed to the activity of the annotated *eve3+7* enhancer.

Here, we tested this assumption explicitly by modeling the *eve3+7* reporter pattern and the endogenous pattern separately. Importantly, it is difficult to interpret the success or failure of a single model. It is much more powerful to compare the performance of two models that together formalize a hypothesis. We compared the performance of the repressor-only and

bifunctional models in WT and *bcd* RNAi embryos. We used Hb protein and *giant* (*gt*), *tailless* (*tll*), and *knirps* (*kni*) mRNA as input regulators and thresholded the endogenous pattern and the reporter pattern for model fitting (Figure 2.1; see Materials and Methods). We report our modeling of the third time point, which is representative of results for other time points (Supplemental Figures 2.5 and 2.6) and evaluated model performance by computing the area under the receiver operating characteristic curve (AUC, Swets 1988).

We first analyzed the endogenous pattern: we fit our models in WT embryos and used the resulting parameters to predict expression in *bcd* RNAi embryos. Both models correctly predicted the positional shifts of stripe 7 and a wide anterior stripe, but the bifunctional model performed better than the repressor-only model ($AUC_{repressor}$ = 0.93, $AUC_{bifunctional}$ = 0.98; Figure 2.3F and Supplemental Figure 2.4).

We next analyzed the *eve3+7* reporter pattern: again, we fit both models in WT embryos and used the resulting parameters to predict expression in *bcd* RNAi embryos. In this case, the repressor-only model was more accurate than the bifunctional model ($AUC_{repressor}$ = 0.90, $AUC_{bifunctional}$ = 0.87; Figure 2.3L). We controlled for several factors that may confound prediction accuracy: we assessed sensitivity to changes in regulator concentrations, refit the models with *bcd* RNAi data, and refit the models on all of the data, none of which changed our conclusions (Supplemental Figures 2.5 and 2.6; see Materials and Methods).

These results suggest that Hb bifunctionally regulates the endogenous pattern but only represses the reporter pattern. Although the differences in relative model performance were subtle, the results supported our hypothesis that the *eve3+7* reporter pattern is regulated differently from the endogenous pattern. However, these differences in model performance were not conclusive of their own accord and prompted us to return to the perturbation that previously distinguished the repressor-only and bifunctional models, ventral misexpression of *hb* (Ilsley et al. 2013; Clyde et al 2003).

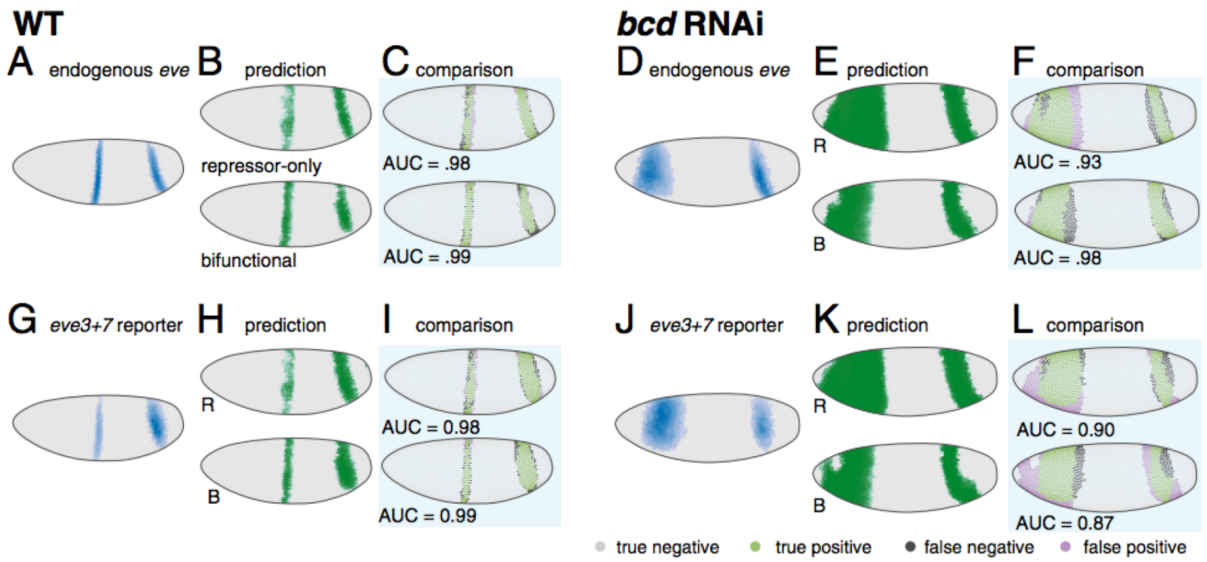


Figure 2.3: In *bcd* RNAi embryos, the bifunctional model more accurately predicts the endogenous pattern and the repressor-only model more accurately predicts the *eve3+7* reporter pattern. (A) The endogenous *eve* pattern in WT embryos is shown as a rendering of a gene expression atlas. Cells with expression below an ON/OFF threshold (Materials and Methods) are plotted in gray. For cells above this threshold, darker color indicates higher relative amounts. (B) The predictions of the repressor-only (R) and bifunctional (B) models in WT embryos. (C) Comparison of model predictions to the endogenous pattern in WT embryos. Green cells are true positives, purple cells are false positives, dark gray cells are false negatives, and light gray cells are true negatives. For visualization, the threshold is set to 80% sensitivity, but the AUC metric quantifies performance over all thresholds. (D) The endogenous *eve* pattern in *bcd* RNAi embryos. (E) The predictions of the repressor-only (R) and bifunctional (B) models in *bcd* RNAi embryos. (F) Comparison of model predictions to the endogenous pattern in *bcd* RNAi embryos. The bifunctional model more accurately predicts the endogenous pattern in *bcd* RNAi embryos. (G–L) Same as A–F, respectively, for the *eve3+7* reporter pattern. The repressor-only model predicts the *eve3+7* reporter pattern more accurately in *bcd* RNAi embryos. Model parameters and AUC scores are in Supplemental Tables 2.1 and 2.2, respectively.

hb *misexpression confirms that the endogenous* eve *pattern and the* eve3+7 *reporter pattern respond to Hb differently*.

In Ilsley et al. 2013, we preferred the bifunctional model because it qualitatively predicted the behavior of a classic genetic perturbation. Misexpressing *hb* along the ventral surface of the embryo from the *snail* promoter (*sna::hb* embryos) causes *eve* stripe 3 to retreat and bend and stripe 7 to bend and bulge (Clyde et al. 2003; Figure 2.4 A and B). In simulations of this perturbation, the bifunctional model predicted this behavior, whereas the repressor-only

model predicted retreat of both stripes (Figure 2.4 E and F, adapted with permission from Ilsley et al. 2013). We hypothesized that the endogenous and reporter patterns would respond differently to *hb* misexpression if Hb bifunctionally regulates the endogenous pattern but only represses the reporter pattern.

We measured both patterns quantitatively at cellular resolution in *sna::hb* embryos (Figure 2.4). As previously observed, the endogenous *eve* stripe 3 retreated from the ventral Hb domain and bent posteriorly, whereas the endogenous stripe 7 expanded and bent anteriorly, consistent with the bifunctional model (Figure 2.4 B and E). By contrast, in the *eve3+7* reporter pattern, both stripes retreated from the ventral Hb domain, consistent with the repressor-only model (Figure 2.4 C and F).

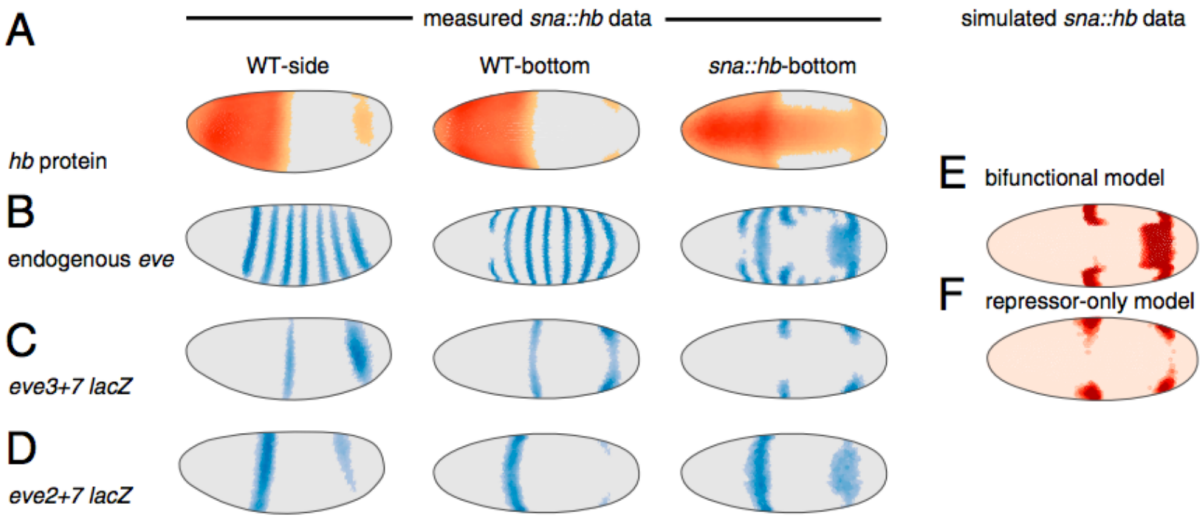


Figure 2.4: In *hb* ventral misexpression (*sna::hb*) embryos, the bifunctional model predicts the endogenous pattern whereas the repressor-only model predicts the *eve3+7* reporter pattern. (A) Hb protein in WT and *sna::hb* embryos (Left, lateral view; Right, ventral view). These data are renderings of gene expression atlases that average together data from multiple embryos (see Supplemental Figure 2.8 for number of embryos per time point). Relative expression level of each gene is plotted in individual cells; cells with expression below an ON/OFF threshold (Materials and Methods) are plotted in gray. For ON cells, darker colors indicate higher relative amounts. (B) Endogenous *eve* pattern. (C) The *eve3+7* reporter pattern. Both stripes retreat from ectopic Hb. (D) The *eve2+7* reporter pattern. Stripe 7 expands into the ectopic Hb domain. (E and F) Bottom (ventral) view of predictions of the bifunctional model (E) and repressor-only model (F) based on simulated *sna::hb* data (adapted with permission from Ilsley et al. 2013). OFF cells are light pink and ON cells are red. All data and modeling are from time point 3 (other time points are shown in Supplemental Figure 2.8).

Two shadow enhancers enable bifunctional Hb regulation.

We hypothesized that additional regulatory DNA in the locus is activated by Hb to produce the *eve* stripe 7 bulge in *sna::hb* embryos. We tested an extended version of the minimal *eve* stripe 2 (*eve2)* enhancer for this activity based on several previous observations. Hb is known to activate the *eve2* enhancer (Small et al. 1991; Arnosti et al. 1996; Small et al. 1992; Stanojevic et al. 1991); longer versions of *eve2* drive stripe 7 in some embryos (Janssens et al. 2006; Goto et al. 1989; Harding et al. 1989; Small et al. 1992); orthologous *eve2* enhancers from other species sometimes drive stripe 7 expression (Hare et al. 2008; Peterson et al. 2009); and, finally, in *sna::hb* embryos, the border of the expanded stripe 7 appeared to be set by *Krüppel* (*Kr*), an *eve2* regulator (Supplemental Figure 2.7; Small et al. 1991; Stanojevic et al. 1991). The fragment we chose drives both stripes 2 and 7 (Supplemental Figure 2.8); we call this enhancer reporter construct *eve2+7*.

In *sna::hb* embryos, the stripe 7 region of the *eve2+7* reporter pattern expanded, recapitulating the bulge observed in the endogenous *eve* pattern (Figure 2.4 B and D). We conclude that Hb activates endogenous *eve* stripe 7 through the *eve2+7* enhancer. Taken together, our results indicate that *eve* stripe 7 expression is controlled by at least two enhancers with different regulatory logic.

Inspired by this example, we decided to examine Hb bifunctionality in a different pair of shadow enhancers. *Krüppel* is a gap gene that had previously been the subject of experimental and computational studies on Hb bifunctionality (Struhl et al. 1992; Schulz & Tautz 1994; Papatsenko & Levine 2008). The *Krüppel* locus also contains a pair of shadow enhancers known as the proximal and distal enhancers (Hoch et al. 1991; Jacob et al. 1991; Perry et al. 2011). To determine how Hb regulates *Krüppel* shadow enhancers, we measured reporter patterns for these sequences in *sna:hb* embryos. In the region of misexpressed *hb*, the proximal enhancer's expression pattern expanded to the posterior, while the distal enhancer's pattern retreated (Figure 2.5). These results suggest that Hb activates the proximal enhancer and represses the

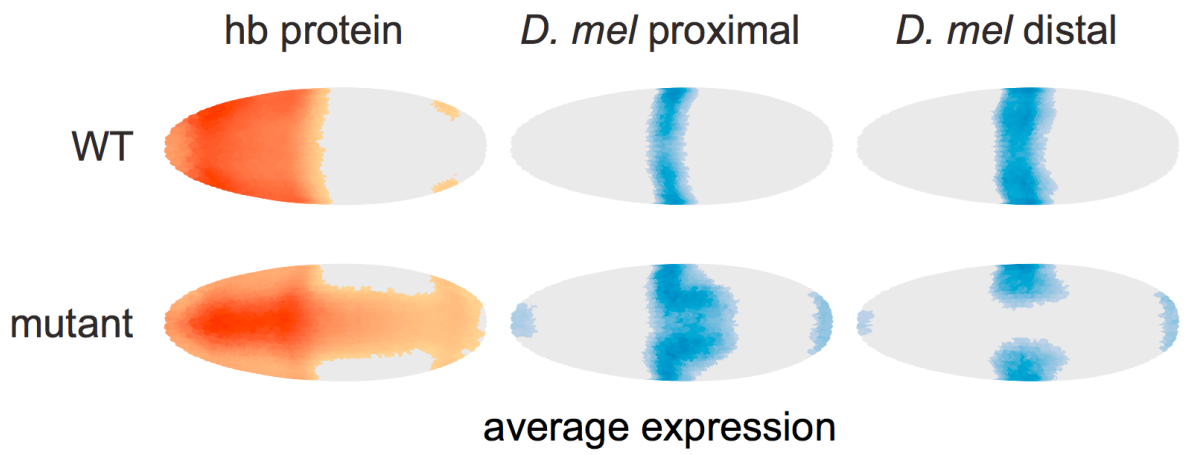


Figure 2.5: Hunchback activates and represses distinct *Krüppel* shadow enhancers. In the left column, we show the average pattern of Hb protein expression in WT and *sna::hb* embryos. We thresholded cells with expression levels greater than the mode + 0.5 standard deviation (SD) as ON and colored OFF cells grey. Deeper colors indicate greater expression. The middle and right columns show the average *lacZ* expression pattern in each genetic background. The data were thresholded at the mode + 1 SD. The expansion of the proximal enhancer's pattern is consistent with Hb activation, while the retreat of the distal enhancer's pattern is consistent with Hb repression. The expression in the poles of the embryos is due to an unused *huckebein* co-stain (Wunderlich et al. 2014).

distal, and demonstrate that shadow enhancers enable Hb bifunctionality in at least two loci in the *Drosophila melanogaster* embryo.

eve3+7 *and* eve2+7 *respond differently to* giant *misexpression.*

By measuring the response of the *eve3+7* and the *eve2+7* patterns to *hb* misexpression, we demonstrated that the *eve* stripe 7 shadow enhancers build the same expression pattern using different regulatory logic. We hypothesized that the regulatory logic of these enhancers differed in other ways. Specifically, the anterior border of *eve2* is defined by the repressor *giant* (Small et al. 1992; Arnosti et al. 1996). In accordance with previous computational work, we hypothesized that *giant* also defines the anterior border of the stripe 7 pattern driven by *eve2+7*, since *giant* is also expressed anterior to *eve* stripe 7 (Janssens et al. 2006). Previous studies using targeted mutagenesis in *eve3+7* demonstrated that its anterior stripe 7 border is defined by *knirps* (Struffi et al. 2011), although the repressor-only model suggests that *eve3+7* may also

be regulated by *giant* (Ilsley et al. 2013). To determine how *giant* regulates *eve3+7* and *eve2+7*, we measured expression driven by *eve3+7* and *eve2+7* reporter constructs in embryos that misexpress *giant* along the lateral surface of the embryo from the *short-gastrulation* (*sog*) promoter (*sog::gt* embryos, Figure 2.6).

With regard to the endogenous pattern, *giant* misexpression repressed stripe 2 and abolished stripe 5, but stripe 7 remained unaffected (Figure 2.6). We observed no effect on the *eve3+7* reporter pattern in *sog::gt* embryos, but in the same background, we observed repression of both stripes in the *eve2+7* reporter pattern (Figure 2.6). These results suggest that the anterior border of stripe 7 is specified by different TFs in *eve* stripe 7 shadow enhancers: *giant* specifies the border in *eve2+7*, while *knirps* specifies the border in *eve3+7*.

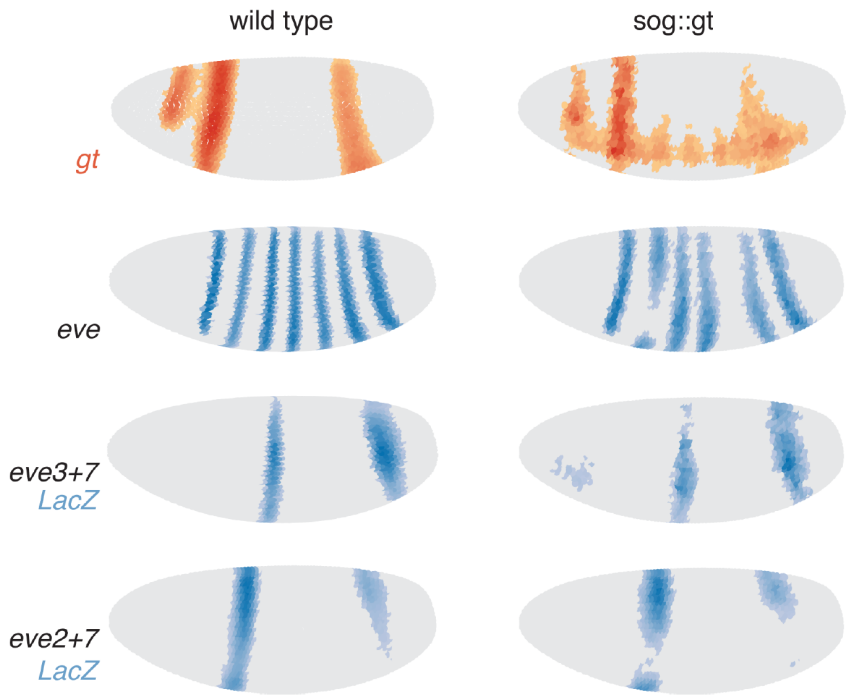


Figure 2.6: *giant* represses *eve2+7* but not *eve3+7*. In the left column, we show the expression pattern of *giant*, *eve, eve3+7 lacZ* and *eve2+7 lacZ* in WT embryos, while in the right column we show the same expression patterns in *sog::gt* embryos. The data were compiled into gene expression atlases and thresholded at the mode + 1 SD. OFF cells are indicated in grey, while ON cells are colored. Deeper colors indicate higher expression levels.

Discussion

To test whether Hb can activate and repress the same enhancer, we used quantitative data to challenge two computational models that together formalize different roles for Hb. We measured expression of endogenous *eve* and transgenic reporter constructs at cellular resolution under multiple genetic perturbations. By comparing the regulatory logic of the endogenous pattern and the *eve3+7* reporter pattern, we uncovered two enhancers that both direct expression of *eve* stripe 7. These shadow enhancers direct the same pattern in different ways: one is activated by Hb whereas the other is repressed. This form of regulatory redundancy enables Hb to "drive with the brakes on" to control *eve* stripe 7.

Two shadow enhancers control eve *stripe 7 expression.*

Early studies suggested that control of *eve* stripe 7 expression was distributed over DNA encompassing both the minimal *eve3+7* and *eve2* enhancers (Small et al. 1991; Small et al. 1996; Goto et al. 1989; Harding et al. 1989; Small et al. 1992). We find that there are at least two pieces of regulatory DNA in this region that position stripe 7. The minimal *eve3+7* enhancer is repressed by Hb (Clyde et al. 2003; Small et al. 1996; Struffi et al. 2011), whereas the *eve2+7* enhancer, which encompasses the minimal *eve2* enhancer, is activated by Hb. This activation may be direct or indirect. Based on the results presented here, we cannot rule out the possibility that the bulge of *eve2+7* in *sna::hb* embryos is due to indirect Hb activity – the result of activation by other TFs and retreat of Gt and Kni (Clyde et al. 2003; Yu & Small 2008). However, we hypothesize that Hb activation of *eve2+7* is direct. If Hb activation of *eve2+7* is indirect, Hb binding to *eve2+7* in these cells would have to have little or no effect on stripe 7 expression (Li et al. 2008). Moreover, Hb binds to and activates the minimal *eve2* enhancer (Small et al. 1991; Arnosti et al. 1996; Small et al. 1992; Stanojević et al. 1989).

In addition to responding to Hb in opposite ways, the *eve2+7* and *eve3+7* enhancers are likely differentially sensitive to additional TFs. *eve3+7* is activated by *Stat92E* and Zelda (Struffi et al. 2011; Yan et al. 1996). The anterior border of stripe 7 is set by Kni repression, and the

posterior border is set by Hb repression (Clyde et al. 2003; Small et al. 1996; Struffi et al. 2011). The minimal *eve2* enhancer is activated by Bcd and Hb, its anterior boundary is set by Gt, and its posterior boundary is set by Kr (Small et al. 1991; Arnosti et al. 1996; Small et al. 1992; Stanojevic et al. 1991). In agreement with others, the anterior boundary of *eve* stripe 7 in eve2+7 appears to be set by Gt (Janssens et al. 2006). Taken together, this evidence argues that *eve3+7* and *eve2+7* position stripe 7 using different regulatory logic.

The molecular mechanism by which Hb represses and activates separate enhancers remains unclear. One hypothesis is that other TFs bound nearby convert Hb from a repressor into an activator, as is the case for Dorsal (Shirokawa & Courey 1997; Jiang & Levine 1993; Dubnicoff et al. 1997; Flores-Saaib et al. 2001; Ratnaparkhi et al. 2006). There is genetic evidence for activator synergy between Bcd and Hb (Small et al. 1991; Simpson-Brose et al. 1994), and activator synergy between Hb and Caudal has been proposed by computational work (Kim et al. 2013). Another hypothesis is that Hb monomers are activators but DNA-bound Hb dimers are repressors (Papatsenko & Levine 2008; Bieler et al. 2011). Testing these hypotheses will require quantitative data in additional genetic backgrounds and mutagenesis of individual binding sites in the two enhancers.

Comparing the regulatory logic of reporter patterns and endogenous patterns can help map regulatory DNA.

"Veteran enhancer-bashers, and those who carefully read the papers, know that 'minimal' enhancer fragments do not always perfectly replicate the precise spatial boundaries of expression of the native gene…" (Barolo 2012). Our data clearly support this often neglected aspect of enhancer reporter constructs. One explanation offered for such discrepancies is different transcript properties. We controlled for this possibility and conclude that transcript properties contribute to the differences between reporter and endogenous patterns but are not the only source. Here, we find that additional regulatory DNA in the locus plays a role.

Finding all of the active regulatory DNA in a locus is challenging. Enhancer reporter constructs are powerful but can only determine whether a piece of DNA is sufficient to drive a particular pattern in isolation when placed next to the promoter. By comparing the regulatory logic of the *eve3+7* reporter pattern and the endogenous pattern, we uncovered an additional feature of *eve* regulation. However, *eve3+7* and *eve2+7* may not contain all of the DNA that contributes to stripe 7 expression *in vivo*. Emerging technologies for manipulating the endogenous locus and larger reporter constructs will be helpful for comprehensively mapping regulatory DNA (Venken & Bellen 2012; Ren et al. 2013; Crocker & Stern 2013).

The bifunctional model is a superposition of two computations.

Models are not ends, but merely means to formalize assumptions and develop falsifiable hypotheses (Gunawardena 2014; Wunderlich & DePace 2011). The bifunctional model accurately predicts the behavior of the endogenous *eve* stripe 7 pattern in WT and perturbed embryos, but it does not predict the behavior of either *eve3+7* or *eve2+7*. The interpretation in (Ilsley et al. 2013) that Hb bifunctionally regulated *eve3+7* was based on a common assumption: that the endogenous pattern could be attributed to the annotated enhancer. Here we show that Hb bifunctionality is due to separate enhancers. Ilsley et al. interpreted the success of the bifunctional model as evidence for concentration-dependent control of Hb activation and repression, as has been proposed for Hb and other TFs (Schulz & Tautz 1994; Papatsenko & Levine 2008; Kelley et al. 2003). This interpretation cannot be true because both Hb activates and represses in the same stripe 7 cells. Our favored hypothesis is that Hb bifunctionality is controlled by sequence features in each enhancer.

The bifunctional model effectively behaves as a superposition of the *eve3+7* and *eve2+7* enhancer activities to accurately predict the behavior of the endogenous locus. It is currently unclear how multiple active enhancers impinge on the same promoter, which makes it challenging to predict their combined behavior. The promoter may integrate information from multiple enhancers in various ways, ranging from independent addition to dominance of one

enhancer owing to a long-range repressor (Barolo 2012; Dunipace et al. 2011; Perry et al. 2011). The behavior of stripe 7 is not consistent with dominant repression by Hb, but we cannot rule out any other mechanisms. Elucidating how promoters integrate information will be critical for predicting the behavior of complex developmental loci where shadow enhancers are prevalent.

Conclusions

By combining computational modeling and directed experiments, we uncovered a previously unidentified feature of a highly studied locus, long held up as a textbook example of modular enhancer organization (Maeda & Karch 2011). We tested the hypothesis that Hb bifunctionally regulates the *eve3+7* enhancer and discovered that bifunctionality is due to two enhancers that respond to Hb in opposite ways. This example provides an opportunity to uncover how Hb bifunctionality is controlled, which will improve our ability to interpret regulatory DNA and infer connections in gene regulatory networks.

Regulatory redundancy in control of *eve* stripe 7 expression may have functional consequences. Shadow enhancers may arise from genetic drift (Lynch 2007); however, shadow enhancers in other developmental loci confer robustness to genetic or environmental stresses (Dunipace et al. 2011; Perry et al. 2010; Frankel et al. 2010), facilitate temporal refinement of patterns (Dunipace et al. 2013), and/or increase expression synchrony and precision (Boettiger & Levine 2009). This example demonstrates that shadow enhancers can use different regulatory logic to position the same pattern, which may have useful properties for the embryo.

Acknowledgements

We thank Miki Fujioka for sharing the *eve* locus reporter flies ahead of publication; Kelly Eckenrode for staining the WT reporter embryos; Steve Small for the *sna::hb* flies; Garth Ilsley for developing the initial models, help with figures, and stimulating discussions; John Reinitz for the Hb antibody; and Steve Small, Becky Ward, Garth Ilsley, Peter Combs, Alistair Boettiger, two anonymous reviewers, and members of the A.H.D. laboratory for comments on the

manuscript. This work was supported by the Harvard Herchel Smith Graduate Student Fellowship (to M.V.S.), Jane Coffin Childs Memorial Fund for Medical Research (to Z.W.), National Institutes of Health (NIH) Grant K99HD073191 (to Z.W.), and NIH Grant U01 GM103804-01A1 (to A.H.D.).

Materials and Methods

Fly work

The *bcd* RNAi gene expression atlas is described in (Staller et al. 2015). Briefly, we combined short hairpin RNA knockdown of *bcd* with *in situ* hybridization and two-photon imaging and automated image segmentation (Fowlkes et al. 2008; Fowlkes et al. 2011; Wunderlich et al. 2012; Staller et al. 2013). Hb protein stains used a guinea pig anti-Hb from John Reinitz, University of Chicago, Chicago, IL. Embryos were partitioned into six time points using the degree of membrane invagination (0–3, 4–8, 9–25, 26–50, 51–75, and 76–100%), which evenly divide the ~60-min blastoderm stage (Keränen et al. 2006). All enhancer reporters are in pBΦY and integrated at attP2 (see Enhancer Sequences in Appendix B; Hare et al. 2008; Groth et al. 2004). The *eve* locus *lacZ* reporter was a gift from Miki Fujioka, Thomas Jefferson University, Philadelphia. *hb* ventral misexpression was performed as described in Clyde et al. using two copies of the *sna::hb* transgene on chromosome 2. *gt* lateral misimpression was perform similarly using two copies of a s*og::gt* transgene on chromosome 2.

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

We determined the genotype of the *sna::hb* embryos by examining the *eve* or *fushi-tarazu* (*ftz*) mRNA patterns. Embryos were aligned morphologically to create a coarsely registered gene expression atlas (Fowlkes et al. 2008). Data are available at dx.doi.org/10.6084/m9.figshare.1270915 and https://depace.med.harvard.edu.

Logistic modeling of enhancer gene regulatory functions

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

To threshold the endogenous WT *eve* pattern into ON and OFF cells, we used 0.2 for all time points (Ilsley et al. 2013). To threshold the endogenous *eve* patterns in the *bcd* RNAi atlas, we used the lowest threshold that would separate the stripes: 0.1, 0.15, 0.15, 0.2, and 0.21 for T = 2 through T = 6, respectively. To compare the modeling of the reporter and the endogenous patterns, the ON set included all cells in the endogenous *eve* stripes 3 and 7 and the OFF set included all other cells. This OFF set is different from that in Ilsley et al. (Ilsley et al. 2013), but this change does not have a large effect on the AUC scores in *bcd* RNAi embryos (Supplemental Tables 2.1 and 2.2).

We controlled for several factors that may confound model performance when models are fit to WT data and used to predict patterns in *bcd* RNAi. When building gene expression atlases, gene expression levels are normalized separately in each atlas, so changes in TF levels between genotypes are obscured. In WT, the gap genes are expressed at similar levels (Little et al. 2013), but their relative levels may be different in *bcd* RNAi. We controlled for this possibility in several ways. First, to simulate changes in level between genotypes, we systematically scaled the levels of each gene in *bcd* RNAi, one at a time, and measured model performance. For the pattern driven by the *eve3+7* reporter, the repressor-only model was more accurate than the bifunctional model for all tested scalings of Hb and a very large range of scalings of the other

regulators (Supplemental Figure 2.4). Analogously, for the endogenous pattern, the bifunctional model always outperformed the repressor-only model as we scaled *kni* and *tll* levels (Supplemental Figure 2.4). For Hb, the bifunctional model was more accurate than the repressor-only model so long as Hb levels in *bcd* RNAi were less than 135% of WT (Supplemental Figure 2.4). Because *hb* levels are reduced in *bcd* mutant embryos (Driever and Nüsslein-Volhard 1989), Hb levels in *bcd* RNAi embryos are almost certainly below this level. Second, we refit the models in *bcd* RNAi and predicted *bcd* RNAi output patterns, testing whether an optimal set of parameters could change the relative predictive performance of the models, but it did not for time points 3–6 (Supplemental Figure 2.5). Third, when we refit the models on both datasets simultaneously and predicted *bcd* RNAi patterns the relative order of model performance did not change for time points 3–6 (Supplemental Figure 2.5).

Sensitivity analysis

For the sensitivity analysis, for each TF we scaled the concentration of the bcd RNAi atlas *in silico* and recomputed the model AUC scores (Supplemental Figure 2.5).

Binding site predictions

For the Kr binding site analysis in Supplemental Figure 2.7, we predicted binding sites using PATSER (stormo.wustl.edu) with a position weight matrix derived from bacterial one-hybrid data (Noyes et al. 2008). Binding sites were visualized using InSite (cs.utah.edu/ miriah/projects).

Quantifying concordance between reporters and endogenous patterns

For each embryo we used the PointCloudToolBox in MATLAB to find pattern boundaries by creating 16 anterior–posterior line traces and finding the inflection point of each trace (adapted from DEMO_EVESTRIPES in the PointCloudToolBox, bdtnp.lbl.gov). Finding the boundary by using half the maximum value of the stripe peak identifies a very similar boundary

to the inflection point. To find the peaks of the endogenous and reporter stripes we took one lateral line trace (extractpattern function) and calculated local maxima.

References

Arnosti, D.N. et al., 1996. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development, 122(1), pp.205–214.

Barolo, S., 2012. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. BioEssays: news and reviews in molecular, cellular and developmental biology, 34(2), pp.135–141.

Bieler, J., Pozzorini, C. & Naef, F., 2011. Whole-embryo modeling of early segmentation in Drosophila identifies robust and fragile expression domains. Biophysical journal, 101(2), pp.287–296.

Boettiger, A.N. & Levine, M., 2009. Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science, 325(5939), pp.471–473.

Briscoe, J. & Thérond, P.P., 2013. The mechanisms of Hedgehog signalling and its roles in development and disease. Nature reviews molecular cell biology, 14(7), pp.416–429.

Clyde, D.E. et al., 2003. A self-organizing system of repressor gradients establishes segmental complexity in Drosophila. Nature, 426(6968), pp.849–853.

Crocker, J. & Stern, D.L., 2013. TALE-mediated modulation of transcriptional enhancers in vivo. Nature methods, 10(8), pp.762–767.

Deng, Z. et al., 2010. Yin Yang 1: a multifaceted protein beyond a transcription factor. Transcription, 1(2), pp.81–84.

Di Stefano, B. et al., 2014. C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature, 506(7487), pp.235–239.

Dubnicoff, T. et al., 1997. Conversion of dorsal from an activator to a repressor by the global corepressor Groucho. Genes & development, 11(22), pp.2952–2957.

Dunipace, L. et al., 2013. Autoregulatory feedback controls sequential action of cis-regulatory modules at the brinker locus. Developmental cell, 26(5), pp.536–543.

Dunipace, L., Ozdemir, A. & Stathopoulos, A., 2011. Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. Development, 138(18), pp.4075–4084.

Flores-Saaib, R.D., Jia, S. & Courey, A.J., 2001. Activation and repression by the C-terminal domain of Dorsal. Development, 128(10), pp.1869–1879.

Fowlkes, C.C. et al., 2011. A conserved developmental patterning network produces quantitatively different output in multiple species of Drosophila. PLoS genetics, 7(10), p.e1002346.

Fowlkes, C.C. et al., 2008. A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell, 133(2), pp.364–374.

Frankel, N. et al., 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature, 466(7305), pp.490–493.

Fujioka, M. et al., 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(11), pp.2527–2538.

Goto, T., Macdonald, P. & Maniatis, T., 1989. Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell, 57(3), pp.413–422.

Groth, A.C. et al., 2004. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics, 166(4), pp.1775–1782.

Gunawardena, J., 2014. Models in biology: “accurate descriptions of our pathetic thinking.” BMC biology, 12(1), p.29.

Harding, K. et al., 1989. Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. The EMBO journal, 8(4), pp.1205–1212.

Hare, E.E. et al., 2008. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS genetics, 4(6), p.e1000106.

He, X. et al., 2010. Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. PLoS computational biology, 6(9), p.e1000935.

Hoch, M., Seifert, E. & Jäckle, H., 1991. Gene expression mediated by cis-acting sequences of the Krüppel gene in response to the Drosophila morphogens bicoid and hunchback. The EMBO journal, 10(8), pp.2267–2278.

Hori, K., Sen, A. & Artavanis-Tsakonas, S., 2013. Notch signaling at a glance. Journal of cell science, 126(Pt 10), pp.2135–2140.

Ilsley, G.R. et al., 2013. Cellular resolution models for even skipped regulation in the entire Drosophila embryo. eLife, 2, p.e00522.

Jacob, Y. et al., 1991. Analysis of Krüppel control elements reveals that localized expression results from the interaction of multiple subelements. Proceedings of the National Academy of Sciences of the United States of America, 88(13), pp.5912–5916.

Janssens, H. et al., 2006. Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. Nature genetics, 38(10), pp.1159–1165.

Jiang, J. & Levine, M., 1993. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. Cell, 72(5), pp. 741–752.

Kazemian, M. et al., 2010. Quantitative analysis of the Drosophila segmentation regulatory network using pattern generating potentials. PLoS biology, 8(8). Available at: http://dx.doi.org/10.1371/journal.pbio.1000456.

Kelley, K.M.M., Wang, H. & Ratnam, M., 2003. Dual regulation of ets-activated gene expression by SP1. Gene, 307, pp.87–97.

Keränen, S.V.E. et al., 2006. Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution II: dynamics. Genome biology, 7(12), p.R124.

Kim, A.-R. et al., 2013. Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic. PLoS genetics, 9(2), p.e1003243.

Kim, H.D. et al., 2009. Transcriptional regulatory circuits: predicting numbers from alphabets. Science, 325(5939), pp.429–432.

Latchman, D.S., 2001. Transcription factors: bound to activate or repress. Trends in biochemical sciences, 26(4), pp.211–213.

Levine, M., Cattoglio, C. & Tjian, R., 2014. Looping back to leap forward: transcription enters a new era. Cell, 157(1), pp.13–25.

Li, X.-Y. et al., 2008. Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS biology, 6(2), p.e27.

Luengo Hendriks, C.L. et al., 2006. Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. Genome biology, 7(12), p.R123.

Lynch, V.J., May, G. & Wagner, G.P., 2011. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. Nature, 480(7377), pp.383–386.

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

Maeda, R.K. & Karch, F., 2011. Gene expression in time and space: additive vs hierarchical organization of cis-regulatory regions. Current opinion in genetics & development, 21(2), pp.187–193.

Meijsing, S.H. et al., 2009. DNA binding site sequence directs glucocorticoid receptor structure and activity. Science, 324(5925), pp.407–410.

Noyes, M.B. et al., 2008. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nucleic acids research, 36(8), pp.2547–2560.

Papatsenko, D. & Levine, M.S., 2008. Dual regulation by the Hunchback gradient in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 105(8), pp.2901–2906.

Perry, M.W. et al., 2010. Shadow enhancers foster robustness of Drosophila gastrulation. Current biology, 20(17), pp.1562–1567.

Perry, M.W., Boettiger, A.N. & Levine, M., 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 108(33), pp.13570–13575.

Peterson, B.K. et al., 2009. Big genomes facilitate the comparative identification of regulatory elements. PloS one, 4(3), p.e4688.

Pisarev, A. et al., 2009. FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. Nucleic acids research, 37(Database issue), pp.D560–6.

Ratnaparkhi, G.S., Jia, S. & Courey, A.J., 2006. Uncoupling dorsal-mediated activation from dorsal-mediated repression in the Drosophila embryo. Development, 133(22), pp.4409–4414.

Ren, X. et al., 2013. Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. Proceedings of the National Academy of Sciences of the United States of America, 110(47), pp.19012–19017.

Schulz, C. & Tautz, D., 1994. Autonomous concentration-dependent activation and repression of Krüppel by hunchback in the Drosophila embryo. Development, 120(10), pp.3043–3049.

Sherman, M.S. & Cohen, B.A., 2012. Thermodynamic state ensemble models of cis-regulation. PLoS computational biology, 8(3), p.e1002407.

Shirokawa, J.M. & Courey, A.J., 1997. A direct contact between the dorsal rel homology domain and Twist may mediate transcriptional synergy. Molecular and cellular biology, 17(6), pp.3345–3355.

Shore, D. & Nasmyth, K., 1987. Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell, 51(5), pp.721–732.

Simpson-Brose, M., Treisman, J. & Desplan, C., 1994. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. Cell, 78(5), pp. 855–865.

Small, S. et al., 1991. Transcriptional regulation of a pair-rule stripe in Drosophila. Genes & development, 5(5), pp.827–839.

Small, S., Blair, A. & Levine, M., 1992. Regulation of even-skipped stripe 2 in the Drosophila embryo. The EMBO journal, 11(11), pp.4047–4057.

Small, S., Blair, A. & Levine, M., 1996. Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. Developmental biology, 175(2), pp.314–324.

Staller, M.V. et al., 2015. A gene expression atlas of a bicoid-depleted Drosophila embryo reveals early canalization of cell fate. Development, 142(3), pp.587–596.

Staller, M.V. et al., 2013. Depleting gene activities in early Drosophila embryos with the “maternal-Gal4-shRNA” system. Genetics, 193(1), pp.51–61.

Stanojević, D., Hoey, T. & Levine, M., 1989. Sequence-specific DNA-binding activities of the gap proteins encoded by hunchback and Krüppel in Drosophila. Nature, 341(6240), pp.331–335.

Stanojevic, D., Small, S. & Levine, M., 1991. Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science, 254(5036), pp.1385–1387.

Struffi, P. et al., 2011. Combinatorial activation and concentration-dependent repression of the Drosophila even skipped stripe 3+7 enhancer. Development, 138(19), pp.4291–4299.

Struhl, G., Johnston, P. & Lawrence, P.A., 1992. Control of Drosophila body pattern by the hunchback morphogen gradient. Cell, 69(2), pp.237–249.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science, 240(4857), pp.1285–1293.

Venken, K.J.T. & Bellen, H.J., 2012. Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and ΦC31 integrase. Methods in molecular biology, 859, pp.203–228.

Wunderlich, Z. et al., 2012. Dissecting sources of quantitative gene expression pattern divergence between Drosophila species. Molecular systems biology, 8, p.604.

Wunderlich, Z. & DePace, A.H., 2011. Modeling transcriptional networks in Drosophila development at multiple scales. Current opinion in genetics & development, 21(6), pp. 711–718.

Yan, R. et al., 1996. Identification of a Stat gene that functions in Drosophila development. Cell, 84(3), pp.421–430.

Yu, D. & Small, S., 2008. Precise registration of gene expression boundaries by a repressive morphogen in Drosophila. Current biology, 18(12), pp.868–876.

Zuo, P. et al., 1991. Activation and repression of transcription by the gap proteins hunchback and Krüppel in cultured Drosophila cells. Genes & development, 5(2), pp.254–264.

CHAPTER 3: CAUDAL COUNTER-REPRESSES HUNCHBACK TO REGULATE *EVEN-SKIPPED* STRIPE 2 EXPRESSION IN DROSOPHILA EMBRYOS

Ben J. Vincent, Max V. Staller, Francheska Lopez-Rivera, Meghan D.J. Bragdon, Zeba Wunderlich, Javier Estrada, and Angela H. DePace.

Author Contributions

B.J.V., M.V.S., Z.W., F.L.-R. and A.H.D. designed the study; B.J.V., F.L.-R. and M.D.J.B. collected the data; Z.W. and J.E. processed raw images for further analysis; B.J.V. analyzed the data; and B.J.V. and A.H.D. wrote the manuscript.

Abstract

Hunchback is a bifunctional transcription factor that can activate and repress separate enhancers in the same cells. We investigated the local sequence features controlling Hunchback function by perturbing its expression in Drosophila embryos and mutating enhancer sequences for its target gene, *even-skipped*. We found that while Hunchback directly represses the *eve* stripe 3+7 enhancer, it is counter-repressed by the transcription factor Caudal in the *eve* stripe 2+7 enhancer. A comparative analysis suggests that the interaction between Hunchback and Caudal is conserved in orthologous enhancers. These results alter the textbook view of *eve* stripe 2 regulation: in order to generate stripe 2, Hunchback repression must be counteracted by Caudal binding.

Introduction

Bifunctional transcription factors (TFs) that can activate or repress their target genes are critical in animal development (Hong et al. 2008; Deng et al. 2010) and are associated with human disease (Aradhya & Nelson 2001; Hui & Angers 2011). In certain cases, the function of these transcription factors depends on the context of the enhancer sequences they bind (Jiang et al. 1992; Jiang & Levine 1993; Dubnicoff et al. 1997; Meijsing et al. 2009). To infer accurate regulatory networks in developmental and disease systems, we must define the sequence features that control bifunctional TFs (Kim et al. 2009).

Hunchback (Hb) is a bifunctional TF that was first identified due to its role in patterning the *Drosophila melanogaster* embryo (Nüsslein-Volhard & Wieschaus 1980). *hb* is a gap gene with many targets throughout Drosophila development including other gap genes (Jaeger 2011), homeotic genes (White & Lehmann 1986; Zhang & Bienz 1992; Wu et al. 2001) and neuronal genes (Tran et al. 2010). The Hb homolog Ikaros is also critical in human hematopoiesis (John & Ward 2011). Hb regulation of the pair-rule gene *even-skipped* (*eve*) has been particularly well-studied. Classic studies which measured the expression of wild-type (WT) and mutated versions of *eve* enhancers in reporter constructs found that Hb directly activates the minimal *eve* stripe 2 enhancer (*eve2min*) (Stanojevic et al. 1991; Small et al. 1991; Arnosti et al. 1996). Other studies that examined endogenous *eve* expression in embryos misexpressing *hb* along the ventral surface of the embryo found that Hb represses *eve* stripes 3 and 7 (Clyde et al. 2003). Qualitative measurements of mutated versions of the *eve* stripe 3+7 enhancer (*eve3+7*) suggest that Hb repression is direct (Stanojević et al. 1989; Small et al. 1996; Struffi et al. 2011).

Even after decades of study, Hb bifunctionality remains controversial in terms of the sequence features that control its function. One hypothesis suggests that binding of a nearby Hb molecule converts Hb from an activator to a repressor (dimerization hypothesis). The dimerization hypothesis is supported by computational work demonstrating that mathematical models with dimerization features included more accurately predict expression of *eve3+7* and the gap gene *Krüppel* (Papatsenko & Levine 2008; Ilsley et al. 2013). *In vitro* experiments

demonstrate that Hb and Ikaros contain zinc finger domains that allow for dimerization (McCarty et al. 2003). Another hypothesis suggests that binding of a different TF converts Hb from a repressor to an activator (co-activation hypothesis). The co-activation hypothesis is supported by *in vivo* measurements of binding site arrays that show that nearby sites for the TF Bicoid (Bcd) stimulate Hb activation (Simpson-Brose et al. 1994). Hb co-activation by Bcd has been incorporated into thermodynamic models of *eve* enhancer function (Janssens et al. 2006; He et al. 2010; Samee & Sinha 2014) and recent models have also incorporated Hb co-activation by another TF, Caudal (Cad) (Kim et al. 2013). Cad activates gap and pair-rule gene expression in the posterior of both long and short-band insects (Rivera-Pomar et al. 1995; Häder et al. 1998; Copf et al. 2004; Olesnicky et al. 2006), and Cad homologs are critical in vertebrate development and human disease (Guo et al. 2004; van Nes et al. 2006; Strumpf et al. 2005; Chawengsaksophak et al. 2004; Beck 2004; Wingert et al. 2007; Freund et al. 1998). Despite the inclusion of both hypothesis in computational models of enhancer function (Ilsley et al. 2013; Kim et al. 2013), neither has been experimentally tested within *eve* enhancers.

We recently found that *eve* stripe 7 is generated by two shadow enhancers that respond differently to *hb* misexpression – Hb represses *eve3+7* and activates an extended version of *eve2min* that also generates stripe 7 (*eve2+7*) (Staller et al. 2015). We observed similar results in the shadow enhancers that control the gap gene *Krüppel* (Wunderlich et al. 2015). The *eve* stripe 7 shadow enhancers are an ideal system to test the dimerization and co-activation hypotheses. Both enhancers are active in the same nuclei, which suggests that Hb function is controlled by enhancer sequence and not modification of Hb protein. Binding sites for Hb and other regulators have been experimentally mapped in both sequences (Gallo et al. 2011) and Cad binding sites can be predicted using position weight matrices (PWMs) in the literature (Noyes et al. 2008; Zhu et al. 2011). Finally, we can measure the effects of genetically perturbing both enhancers quantitatively and at cellular resolution (Figure 3.1; Luengo Hendriks et al. 2006; Fowlkes et al. 2008; Wunderlich et al. 2014).

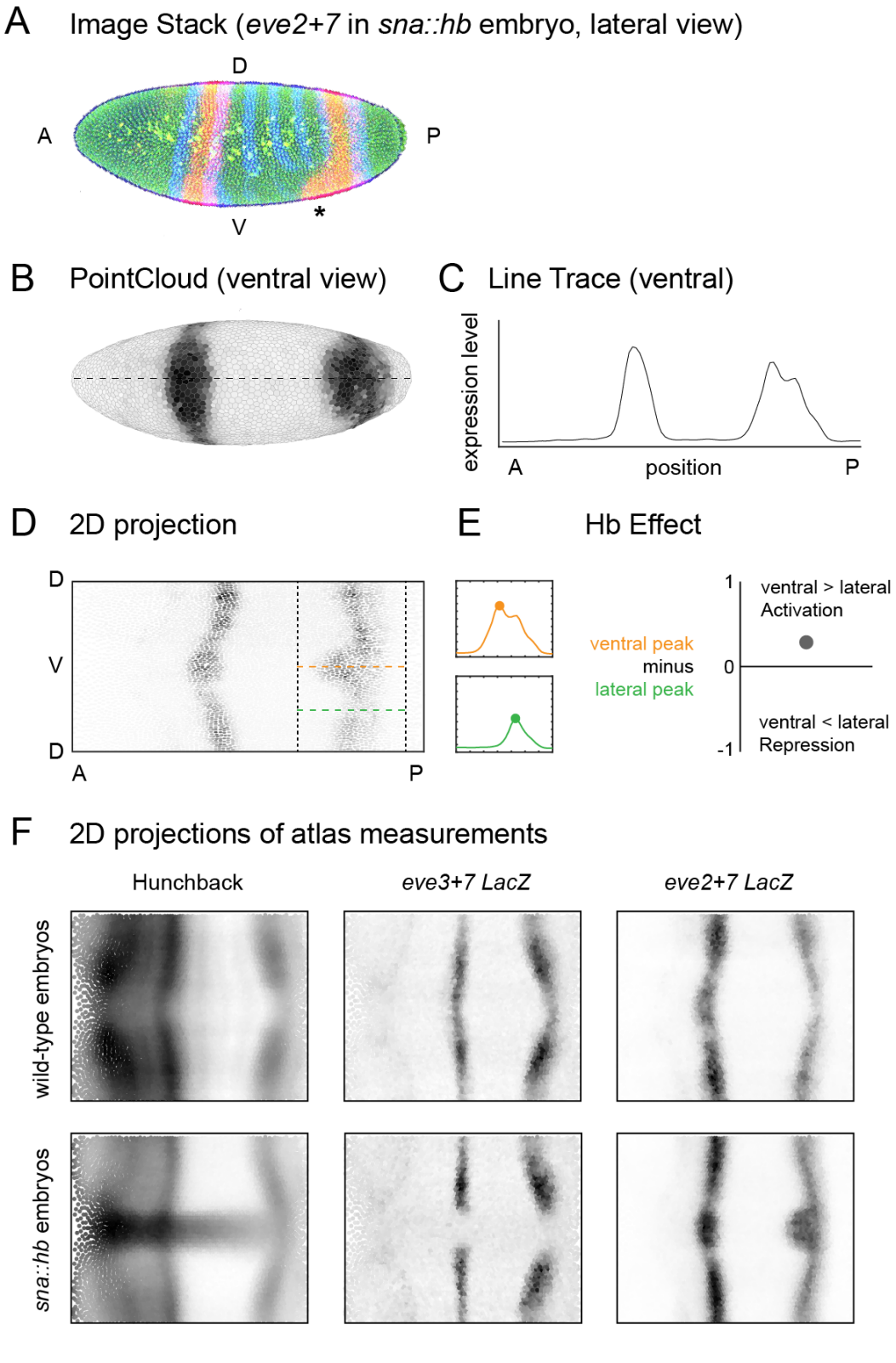


Figure 3.1: Quantitative expression data in embryos expressing ventral *hunchback*. Maximum intensity projection of a *Drosophila melanogaster* embryo expressing

Figure 3.1 (Continued). ventral *hb* with an *eve2+7 LacZ* reporter construct. Individual nuclei are stained using SYTOX green (green), and mRNA for *fushi-tarazu* (blue) and *LacZ* (red) are labeled using fluorescent *in situ* hybridization. Anterior expansion of the ventral stripe 7 pattern is indicated with an asterisk. (B) Visual rendering of expression data for the embryo in (A) following image processing. Individual nuclei are outlined, darker coloring indicates higher expression level. Rendering was generated using PointCloudXplore (Rübel et al. 2006); the raw image for the rendering was color inverted and adjusted for levels using Adobe Photoshop. (C) A line trace plots expression level as a function of position along the anterior-posterior (AP) axis for an individual embryo. (D) Expression data projected as a 2-dimensional plot. Positions of individual nuclei along the dorsal-ventral (DV) axis are plotted as a function of position along the AP axis; darker color indicates higher expression for each nucleus. Relative expression values are normalized to the maximum value and range from 0 to 1. (E) The effect of *hb* misexpression is measured for individual embryos by subtracting the peak level of a lateral line trace (green) from the peak level of a ventral line trace (orange). Negative values indicate repression by ventral Hb; positive values indicate activation. (F) 2-dimensional projections of atlas data from WT embryos (top) and *sna::hb* embryos (bottom) in time point 4. Darker colors indicate higher levels of protein or mRNA. Measurements were made for for Hb protein (left), *eve3+7 LacZ* mRNA (middle) and *eve2+7 LacZ* mRNA (right) using immunofluorescence and *in situ* hybridization.

In this work, we combine perturbations in enhancer sequence and the Hb expression pattern to dissect Hb bifunctionality in *eve3+7* and *eve2+7*. We find that Hb directly represses *eve3+7* and indirectly activates *eve2+7*. We also find that Cad binding sites prevent Hb repression in *eve2+7*, and this interaction appears to be evolutionarily conserved. Our work alters the textbook model that states that Hb is a direct activator of *eve2min* (Alberts et al. 2014) – Hb represses *eve* stripe 2, and Cad is a direct regulator that counter-represses Hb. The co-evolutionary signature in *eve2* is notable since this enhancer has been a foundation for comparative studies on gene regulation for years (Ludwig et al. 1998; Ludwig et al. 2005; Hare et al. 2008; Peterson et al. 2009; Crocker and Erives 2008; Hare et al. 2008). We discuss the possibility that other activators may function as counter-repressors in development.

Results

Our approach was to predict and mutate binding sites for Hb and Cad in *eve3+7* and *eve2+7* and measure the consequences quantitatively in WT embryos and embryos expressing ventral *hunchback* (*sna::hb* embryos). We display our quantitative data in multiple ways: 1) a

line trace of expression level versus anterior-posterior (AP) position along the lateral or ventral surface of the embryo (Figure 3.1C); 2) a 2-dimensional rendering of expression level in every cell in a single embryo or a gene expression atlas (Figure 3.1 D and F); or 3) a plot of the differences in peak expression level between a lateral and a ventral line trace in individual embryos (Figure 3.1E). Mutating enhancer sequence (*cis* perturbations) and misexpressing *hb* (*trans* perturbations) can have unintended consequences. Mutating binding sites for a single TF may affect binding sites for other TFs (Stanojevic et al. 1991), and misexpressing *hb* alters the expression patterns of other regulators (Staller et al. 2015). Combining *cis* and *trans* perturbations helps interpret the results from each experiment alone. We apply this approach to the question of Hb bifunctionality: how is it controlled within *eve* enhancers?

Hb directly represses eve3+7.

Previous work suggested that Hb defines the anterior border of stripe 3, but due to the qualitative nature of the *in situ* hybridization method, the effect of Hb on stripe 7 was unclear (Struffi et al. 2011). Furthermore, the stripe 7 pattern driven by *eve3+7* retreats in response to ectopic Hb, but this effect could be indirect (Staller et al. 2015). For this study, we designed an *eve3+7* sequence containing mutations in Hb binding sites (e*ve3+7 mut Hb*, Figure 3.2A) predicted from bacterial 1-hybrid methods (Noyes et al. 2008). We compared the expression pattern driven by this construct with the pattern driven by *eve3+7* (Figure 3.2) as well as a mutant construct from Struffi et al. 2011 (Supplemental Figure 3.1). All enhancers were cloned into the same pBΦY-*LacZ* reporter backbone (Hare et al. 2008) and integrated into the attP2 landing site using phiC31 transgenesis (Groth et al. 2004). We measured expression driven by these reporter constructs in WT and *sna::hb* embryos.

In accordance with previous results, the eve3+7 *mut Hb* construct drove expression patterns that expanded anteriorly from stripe 3 (Figure 3.2B). In addition, the expression pattern driven by *eve3+7 mut Hb* did not retreat from ventral Hb in *sna::hb* embryos (Figure 3.2 C and D). These results suggest that Hb directly represses *eve3+7*.

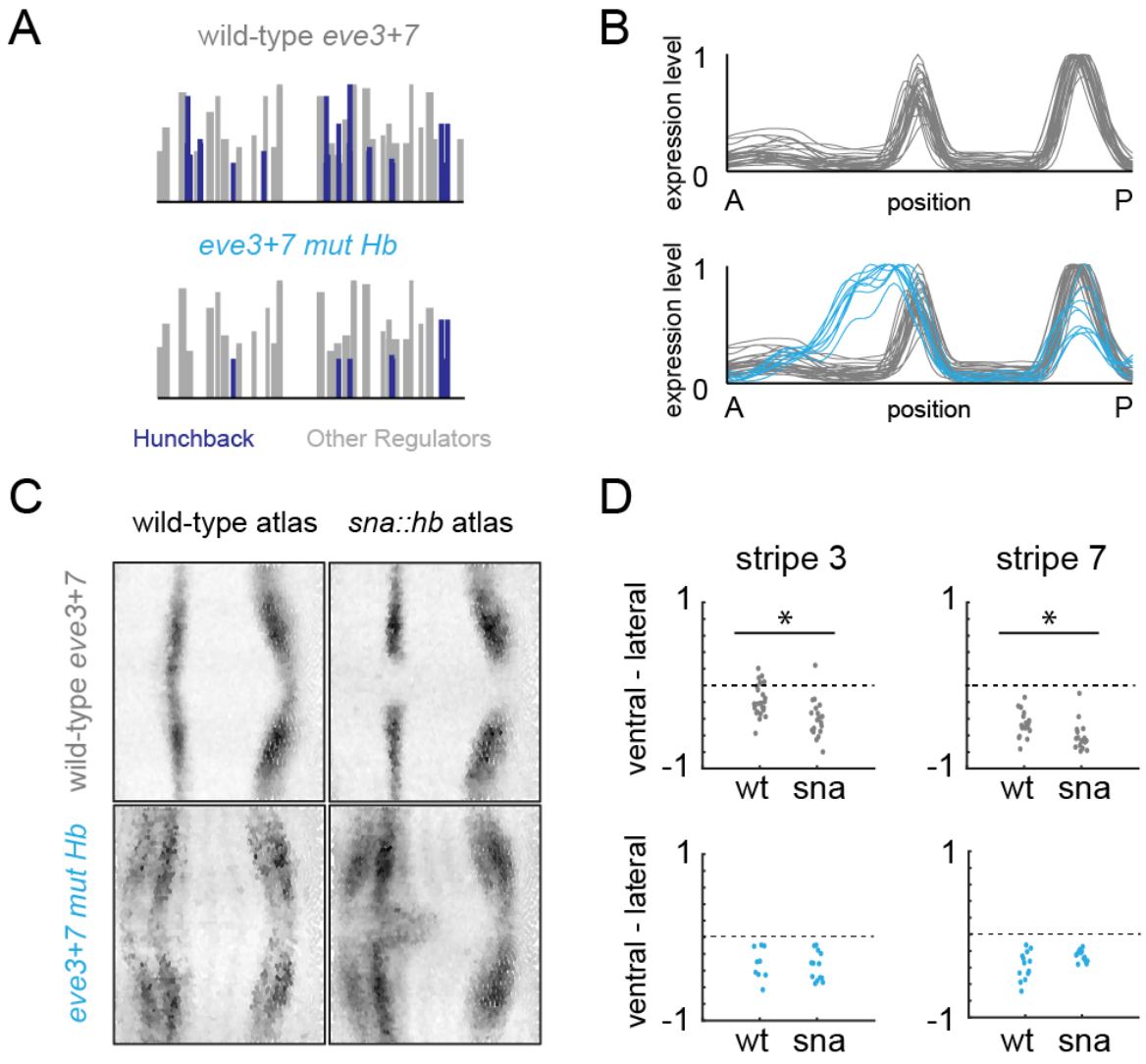


Figure 3.2: Hunchback directly represses *eve3+7*. (A) Predicted Hb binding sites in *eve3+7* and *eve3+7 mut Hb*. Vertical bars represent individual predicted sites along the sequence; bar height is proportional to PATSER score. We used a position weight matrix derived from bacterial 1-hybrid experiments for prediction (Noyes et al. 2008). Binding sites were visualized using inSite software. (B) Lateral line traces of *LacZ* mRNA from individual WT embryos containing *eve3+7* reporter constructs (WT: grey, n= 26; m*ut Hb*: blue, n = 9). Each trace is normalized to its maximum value. Embryos are from all six time points in stage 5. (C) 2D projections of atlas data for reporter constructs expressed in WT or *sna::hb* embryos. Data is taken from time point 4. (D) Differences in the maximum values of ventral and lateral line traces are plotted for individual WT and *sna::hb* embryos in all 6 time points in stage 5 (see Materials and Methods). Top: *eve3+7* (wt: n = 26; *sna::hb*: n = 19); Bottom: *eve3+7 mut Hb* (wt: n = 9; *sna::hb*: n = 13). Asterisks indicate significant differences between WT and *sna::hb* embryos (Mann-Whitney U test, p-value < 0.001,). Differences between WT and *sna::hb* embryos containing *eve3+7 mut Hb* were not significant (Mann-Whitney U test, p-value > 0.4 for both stripes).

Hb indirectly activates stripe 7 in eve2+7.

Previous experimental and computational work suggest that Hb directly activates the minimal *eve* stripe 2 enhancer (Arnosti et al. 1996; Janssens et al. 2006), and the stripe 7 pattern driven by *eve2+7* bulges in response to ectopic Hb (Staller et al. 2015). However, whether Hb directly activates stripe 7 through *eve2+7* remained untested. We synthesized an *eve2+7* construct that contained mutations in predicted Hb sites (*eve2+7 mut Hb*, Figure 3.3A), and we compared expression driven by this construct with *eve2+7* in both WT and *sna::hb* embryos. In WT embryos, we used a co-stain to determine differences in stripe level between *eve2+7* and *eve2+7 mut Hb*. We found that *eve2+7 mut Hb* drives slighty lower expression levels in stripe 2 ($p < 0.05$, Mann-Whitney U test), but stripe 7 levels were indistinguishable between *eve2+7* and *eve2+7 mut Hb* (Figure 3.3B). In *sna::hb* embryos, stripe 2 expression levels driven by both *eve2+7* and *eve2+7 mut Hb* are unaffected by ventral Hb (Figure 3.3 C and D). Finally, the stripe 7 pattern driven by both reporter constructs expanded in identical ways in *sna::hb* embryos (Figure 3.3 E and F). These results suggest that while Hb may be a weak activator of stripe 2, it does not influence stripe 7 expression at endogenous Hb levels in either *eve3+7* or *eve2+7*.

Given the subtle effects that Hb binding site mutations had in *eve2+7*, we examined the behavior of other *eve* stripe 2 enhancer fragments in the literature. When we mutated all three annotated binding sites from the sequence analyzed in Ludwig et al. 1998, we observed a slight but significant decrease in stripe 2 expression level (Supplemental Figure 3.2). However, we observed no significant decrease in expression level as a result of mutating the annotated Hb binding site in the context of the minimal *eve* stripe 2 enhancer, despite contradictory evidence from previous qualitative work (Supplemental Figure 3.2; Small et al. 1992; Arnosti et al. 1996). These data suggest that Hb binding site mutations in *eve* stripe 2 enhancers have small, context-dependent effects on gene expression.

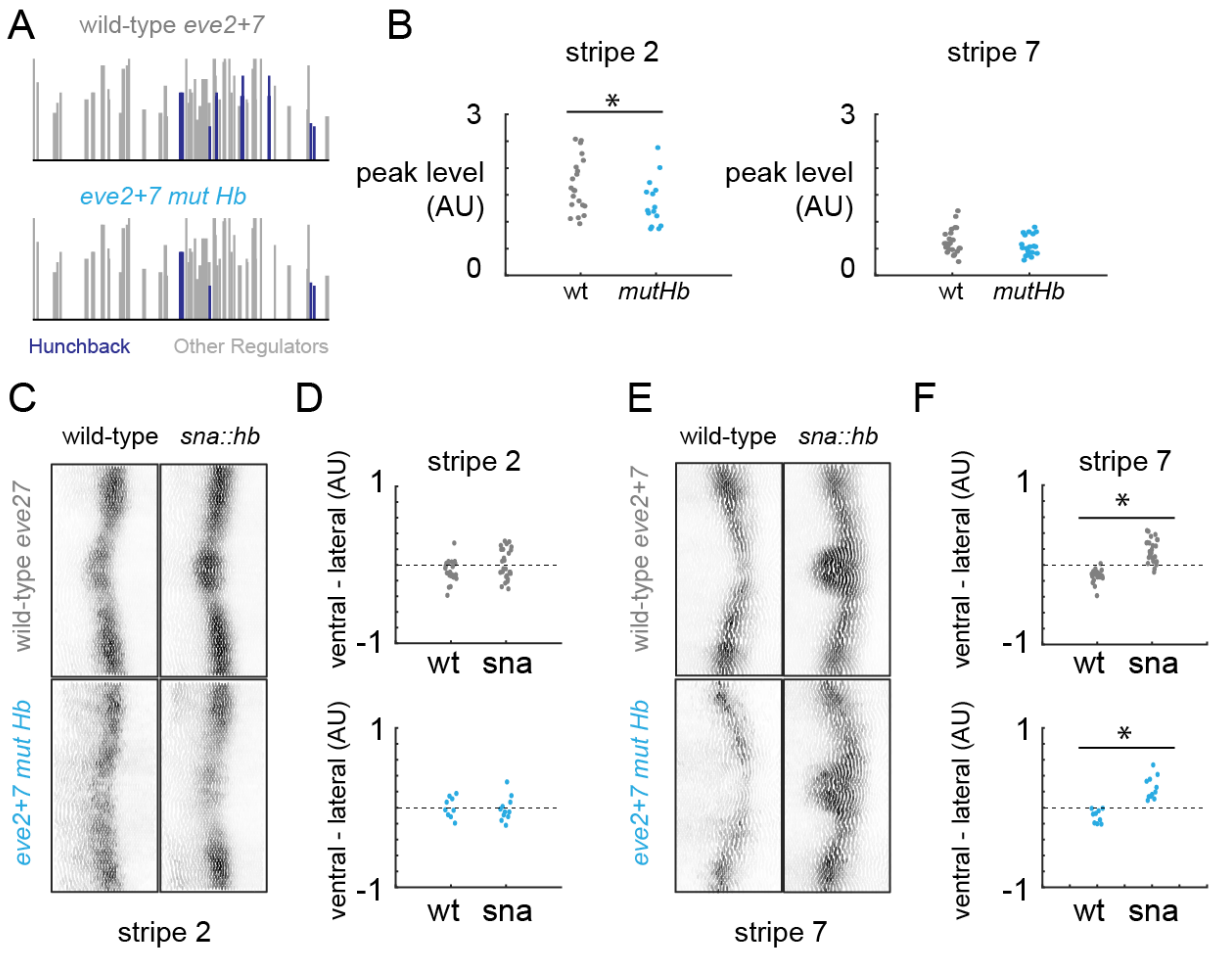


Figure 3.3: Hunchback indirectly activates *eve2+7*. (A) Predicted Hb binding sites in WT *eve2+7* and *eve2+7 mut Hb*. Binding sites predicted and represented as in Figure 3.2. (B) Peak stripe expression levels for individual embryos from time points 2-4 containing reporter constructs for *eve2+7* (grey, n = 23) and *eve2+7 mut Hb* (blue, n = 20). In this analysis, we normalized *LacZ* expression levels using a *huckebein* co-stain (Wunderlich et al. 2014) and extracting local maxima from lateral line traces. Differences in stripe 2 peaks were significant (asterisk; p-value < 0.05, Mann-Whitney U test) but differences in stripe 7 peaks were not (p-value > 0.25, Mann-Whitney U test). (C) 2D projections of stripe 2 atlas data for reporter constructs expressed in wild-type or *sna::hb* embryos. Data is taken from time point 4. (D) Differences in the maximum value of ventral and lateral stripe 2 line traces are plotted for individual WT and *sna::hb* embryos in all 6 time points in stage 5. Top: *eve2+7* (wt: n = 22; *sna::hb*: n = 26); Bottom: *eve2+7 mut Hb* (wt: n = 9; *sna::hb*: n = 11). (E and F) Same as (C) and (D), respectively, for stripe 7. Asterisks indicate significant differences between wild-type and *sna::hb* embryos (p-value < 0.001, Mann-Whitney U test). Differences in stripe 7 modulation between *sna::hb* embryos containing reporter constructs for eve2+7 and *eve2+7 mut Hb* were not significant (p-value > 0.26, Mann-Whitney U test).

Caudal counter-represses Hunchback in eve2+7.

Our data suggest that Hb acts as a direct repressor in *eve3+7* and a weak activator in *eve2+7*. These results could be explained by either the dimerization hypothesis or the co-activation hypothesis. These hypotheses are not easily distinguished, since the parameters governing the proposed Hb binding site interactions are unknown. However, the co-activation hypothesis makes a strong prediction – mutating Cad sites in *eve2+7* should allow for Hb repression within that sequence.

We tested this prediction by synthesizing an *eve2+7* construct with mutations in predicted Cad sites (*eve2+7 mut Cad*, Figure 3.4A). In WT embryos, these mutations abolished stripe 2 expression completely, and caused an anterior expansion of the stripe 7 pattern (Figure 3.4B). We believe this anterior expansion was caused by unintended mutations in binding sites for the repressor Giant (Gt). Gt is hypothesized to set the stripe 7 border generated by *eve2+7* (Janssens et al. 2006; Staller et al. 2015), and predicted Gt sites overlap with predicted Cad sites in this sequence (Supplemental Figure 3.3). One of our added Cad sites disrupts a predicted Gt site that overlaps an annotated Gt site in *eve2min* (Supplemental Figure 3.3; Gallo et al. 2011). While these effects make it difficult to determine whether Cad directly activates *eve2+7*, they do not affect our conclusions concerning the influence of Cad on Hb function.

We also measured the expression pattern driven by *eve2+7 mut Cad* in *sna::hb* embryos. We found that the stripe 7 pattern retreats from ventral Hb (Figure 3.4 C and D). To confirm that this effect was due to direct repression by Hb, we measured the expression pattern driven by a double mutant construct containing mutations in predicted sites for both Cad and Hb (*eve2+7 mut Cad and Hb*). Importantly, the Hb binding site mutations introduced in this construct were the same as those introduced in *eve2+7 mut Hb*. These additional Hb mutations restored expression in the area of stripe 2 (Figure 3.4B) and abolished repression by ventral Hb (Figure 3.4 C and D). *eve2+7 mut Cad and Hb* drives a stripe 2 pattern that expanded to the anterior compared to WT *eve2+7* (Figure 3.4B); this effect could also be due to mutations in Gt binding sites. Taken together, these results disfavor the dimerization hypothesis in favor of the

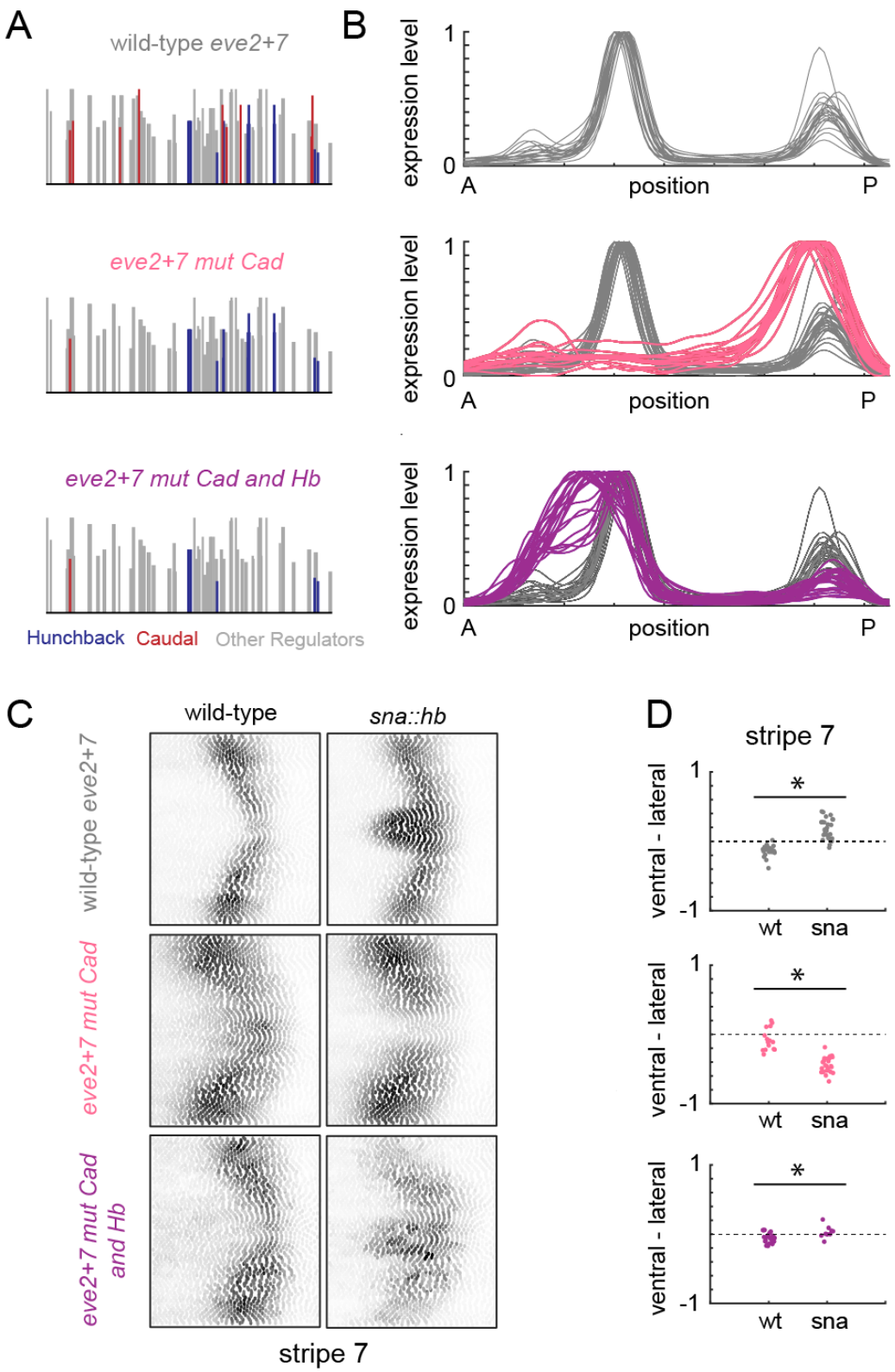


Figure 3.4: Caudal counter-represses Hunchback in *eve2+7*. (A) Predicted Cad (red) and Hb (blue) binding sites in *eve2+7*, *eve2+7 mut Cad* and *eve2+7 mut Cad and Hb*. Binding

Figure 3.4 (Continued). sites predicted and represented as in Figures 3.2 and 3.3. (B) Line traces from individual embryos containing *LacZ* reporter constructs for *eve2+7* (grey), *eve2+7 mut Cad (salmon)* and *eve2+7 mut Cad and Hb* (plum). Each trace is normalized to its maximum value. Embryos are from all six time points in stage 5. (C) 2D projections of stripe 7 atlas data for reporter constructs expressed in WT or *sna::hb* embryos. Data is taken from time point 4. (D) Differences in the maximum value of ventral and lateral line traces are plotted for individual WT and *sna::hb* embryos in all 6 time points in stage 5. Top: *eve2+7* (wt: n = 22; *sna::hb*: n = 26); Middle: *eve2+7 mut Cad* (wt: n = 15; *sna::hb*: n = 24); Bottom: *eve2+7 mut Cad and Hb* (wt : n = 22; *sna:hb*: n = 9). Asterisks indicate significant differences between wild-type and *sna::hb* embryos as determined by a p-value < 0.05 using the Mann-Whitney U test. Differences in stripe 7 modulation between *sna::hb* embryos containing *eve2+7* and *eve2+7 mut Cad* were significant (p-value < 10^{-8}, Mann-Whitney U test), as were differences in stripe 7 modulation between *eve2+7 mut Cad* and *eve2+7 mut Cad and Hb* (p-value < 10^{-4}, Mann-Whitney U test).

co-activation hypothesis: Cad binding sites prevent Hb from directly repressing *eve2+7*.

Cad binding sites co-evolve with Hb binding sites in eve2+7*, but not* eve3+7.

To determine whether Cad directly regulates *eve3+7*, we synthesized an *eve3+7* construct containing mutations in predicted Cad sites. This construct drove lower levels of expression in both stripes (Supplemental Figure 3.4). These data suggest that Cad directly activates *eve3+7*, although we cannot rule out the possibility that these mutations in Cad sites allow for additional repression from nearby Hb sites.

These results confirm previous hypotheses concerning the regulation of *eve3+7*, but illuminate a previously unverified aspect of *eve* stripe 2 regulation: Cad sites in *eve2+7* are necessary to counter-repress Hb in the area of stripe 2. With this new view of *eve* stripe 2 regulation in mind, we hypothesized that Hb and Cad sites would co-evolve in orthologous *eve* stripe 2 enhancers because the creation of new Hb sites could be tolerated if accompanied by the creation of new Cad sites. To test this hypothesis, we calculated binding site enrichment scores for Cad and Hb (Wunderlich et al. 2015) in orthologous *eve* stripe *2* enhancers previously identified in Drosophila and Sepsid genomes (Hare et al. 2008). Based on this analysis, the *eve2* enhancer from *D. melanogaster* was slightly depleted for both Cad and Hb (Figure 3.5, magenta). The *eve2* enhancer from *Dicranosepsis sp*. was most highly enriched for Cad and Hb,

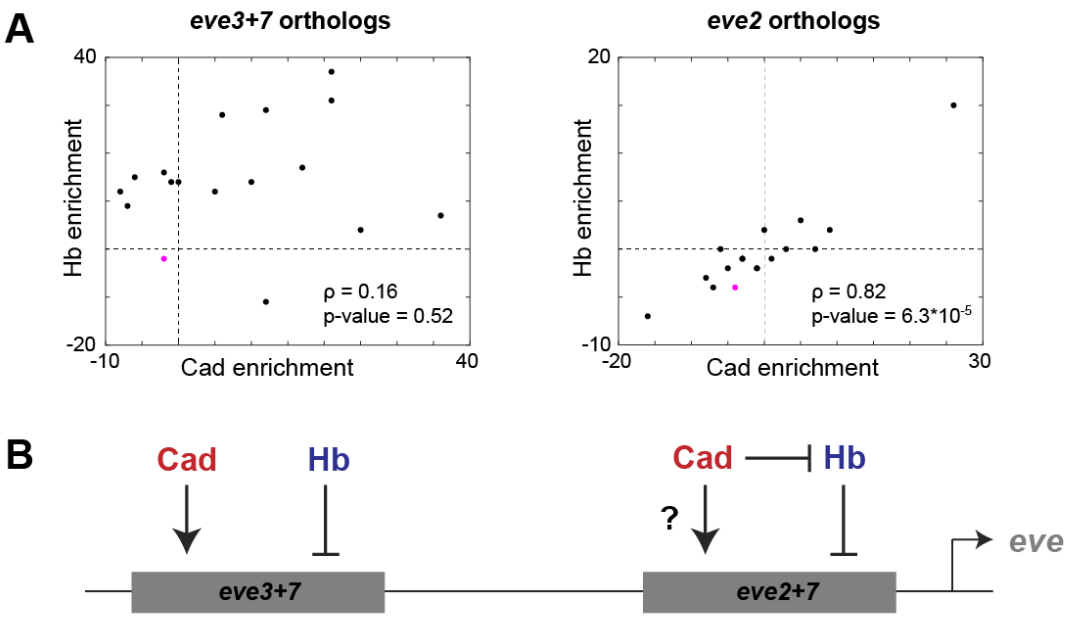


Figure 3.5: Caudal and Hunchback binding sites co-evolve in orthologous *eve2* enhancers. (A) Enrichment scores for predicted Hb and Cad binding sites were calculated using a comparison to the genomic background (Wunderlich et al. 2015) and plotted for previously identified *eve3+7* and *eve2* sequences in different Drosophila and Sepsid species (Hare et al. 2008) at a PATSER score threshold of 1.0 – the threshold that maximizes rho in *eve2* orthologs (see Supplemental Figure 3.5). Background levels of enrichment for both factors are indicated with dashed lines. Values for *Drosophila melanogaster* enhancers are plotted in magenta. We calculated the Spearman correlation and corresponding p-value for enrichment values in each of enhancers. (B) Summary of current findings. Cad and Hb directly activate and repress *eve3+7*, respectively. Cad prevents Hb repression in *eve2+7*. Whether Cad activates *eve2+7* independent of its effect on Hb repression is unknown.

while the *eve2* enhancer from *Sepsis punctum* was most highly depleted. Overall, we found that Cad and Hb enrichment scores were significantly correlated in orthologous *eve* stripe 2 enhancers (Figure 3.5A). However, we did not observe a similar correlation in orthologous *eve3+7* enhancers at the same binding site threshold, even though both Cad and Hb directly regulate *eve3+7* (Figure 3.5B). We also used the same method to test for co-evolution of Hb and Bcd in *eve2* and *eve3+7* orthologs. We observed no significant correlation in either case (Supplemental Figure 3.6).

We note that Hb and Cad enrichment correlations in *eve2* and *eve3+7* orthologs depend on the PATSER score threshold we use to count binding sites. Sensitivity analyses indicate that the correlation of Hb and Cad binding sites in *eve2* orthologs is highest when predicted sites with low PATSER scores are included (Supplemental Figure 3.5). This result suggests that low affinity binding sites for both factors may be important for their interaction. However, in *eve3+7* orthologs, we find that correlation values generally increase with PATSER threshold (Supplemental Figure 3.5). This result suggests that high affinity binding sites for both factors may co-evolve in *eve3+7* orthologs, although analyses for both sets of enhancers may suffer from multiple hypothesis testing. Binding site plots for Bcd, Cad and Hb in *eve2* orthologs can be found in Supplemental Figure 3.7. Nonetheless, when considered in light of the experimental data, the evolutionary signature in *eve2* orthologs suggests that the functional interaction between Cad and Hb sites may be a specific, conserved feature of *eve* stripe 2 regulation.

Discussion

In this work, we tested current hypotheses concerning the sequence features controlling Hb bifunctionality in *eve* enhancers. We found that Hb directly represses *eve3+7*, while mutations in Hb binding sites in *eve2+7* slightly decrease stripe 2 expression level and have no measurable effect on stripe 7. Cad binding sites in *eve2+7* prevent Hb repression, and this interaction is conserved in orthologous enhancers. Our results change our current understanding of *eve* stripe 2 regulatory logic and provide a tractable experimental foundation for uncovering the molecular mechanisms underlying context-dependent TF function.

New insights into eve *stripe 2 regulation*

The *eve* stripe 2 enhancer is a canonical example of how a complex regulatory architecture produces a specific gene expression pattern in development (Alberts et al. 2014). A product of classic studies in developmental genetics, the textbook picture of *eve* stripe 2 regulation includes direct activation by Bcd and Hb (Stanojevic et al. 1991; Arnosti et al. 1996)

and direct repression by Gt and Krüppel (Small et al. 1991). Later experiments showed that Sloppy-paired 1 (Andrioli et al. 2002) and Zelda (Struffi et al. 2011) also directly bind the sequence. However, the failure of reconstitution experiments that attempted to build this enhancer from annotated binding sites suggested that there is still more to learn about its regulation (Vincent et al. 2016).

We have shown that Cad binding sites in *eve2+7* are necessary to prevent Hb-mediated repression of the stripe 2 pattern. These results disfavor the dimerization hypothesis, which suggested that bound Hb monomers act as activators, while Hb molecules that can dimerize on DNA act as repressors (Papatsenko & Levine 2008). We find that the same Hb sites can mediate activation or repression depending on the presence of surrounding Cad sites. Our results favor a co-activation hypothesis that had previously been incorporated into quantitative models that predict *eve* expression patterns from enhancer sequence (Kim et al. 2013). However, these same models also predict that mutations in Hb binding sites dramatically decrease the expression level of stripes 2 and 7 (Janssens et al. 2006). Our data suggest more modest effects of Hb binding site mutations, in accordance with qualitative data in the literature (Small et al. 1992; Arnosti et al. 1996). In addition, these models incorporate Hb co-activation by Bcd (Kim et al. 2013), which has been experimentally validated in tissue culture (Small et al. 1991) and synthetic binding site arrays (Simpson-Brose et al. 1994). However, Bcd-mediated co-activation does not appear to be a feature of *eve2+7*, since mutations in predicted Caudal sites are sufficient to allow for Hb repression. While this result does not rigorously disprove Bcd co-activation in *eve2+7*, it appears that the Bcd sites remaining in *eve2+7 mut Cad* are not sufficient to prevent Hb repression. Additional experiments will be required to dissect how individual Cad and Bcd sites contribute to counter-repression of Hb.

We note that the endogenous *eve* stripe 2 is expressed in c*audal* mutant embryos (Olesnicky et al. 2006), which conflicts with our results suggesting that mutations in *eve2+7* Cad binding sites abolish stripe 2 expression. We can explain this discrepancy in a number of ways. First, other important regulatory DNA in the endogenous locus may contribute to stripe 2

expression. Indeed, the endogenous DNA that lies upstream of *eve2+7* is devoid of predicted Hb binding sites and therefore may not be affected by Hb repression in *cad* mutants (see Supplemental Figure 2.8). Second, *hb* expression may be affected in a *cad* mutant background. While *hb* expression is detectable in this background by qualitative *in situ* hybridization (Olesnicky et al. 2006), expression levels may be lower, which would affect the capacity of Hb to repress *eve* stripe 2. Finally, Hb sites in *eve2+7* reporter constructs may be directly repressing the promoter due to their close proximity. This possibility would not affect our conclusions regarding Cad counter-repression. It would, however, change our understanding of the target of Hb repression, which may be sufficient to affect the promoter, but not the activity of nearby activators in the endogenous context. The *cis*-regulatory rules governing Hb repression are not as well-characterized compared to other repressors in the blastoderm (Kulkarni and Arnosti 2005), but it will be important to test whether the effects of binding site mutations are consistent in *eve2+7* when placed at its endogenous distance from the promoter. As discrepancies between *eve3+7* and the endogenous stripes revealed the influence of *eve2+7* as a stripe 7 shadow enhancer, further dissection of stripe behavior in *cad* mutant embryos may yield additional insights into *eve* regulatory logic (Staller et al. 2015).

New insights into eve *stripe 2 evolution*

Despite turnover in mapped binding sites, orthologous *eve2* enhancers drive similar expression patterns (Ludwig et al. 1998; Hare, Peterson, Iyer, et al. 2008) and can functionally rescue an *eve2* deletion in *Drosophila melanogaster* (Ludwig et al. 2005). Measurements of chimeric enhancers suggest that compensatory evolution occurs in *eve2* (Ludwig et al. 2000), and a recent study suggests that regulatory evolution occurs outside the boundaries of the minimal stripe element (Crocker & Stern 2017). The search for conserved sequence features in *eve2* has focused primarily on small groups of binding sites (Crocker & Erives 2008; Hare, Peterson & Eisen 2008) or a global balance of activation and repression as parameterized by computational modeling (Martinez et al. 2014). Fitting computational models on orthologous

enhancers can also shed light on interactions between transcription factors, such as cooperative DNA binding by Cad (Duque et al. 2014).

Here, we show that a functional interaction between Hb and Cad has an evolutionary signature – Hb and Cad enrichment scores are correlated in orthologous enhancers. Overall, this result supports the conclusions of the targeted genetic experiments in *eve2+7* and represents a specific interaction that may underlie compensatory evolution in the sequences controlling *eve* stripe 2. Like previous work, our analysis suggests that a balance between transcription factors is a conserved feature of *eve2* (Martinez et al. 2014); however, our method for calculating binding site enrichment does not require computational modeling or parameter fitting. It can therefore be readily applied to any sequence of interest given recent investment in resources for binding site predictions (Zhu et al. 2011) and genome sequences for multiple species (Drosophila 12 Genomes Consortium et al. 2007).

Control of bifunctional transcription factors in development

While many bifunctional transcription factors are regulated in *trans*, relatively few have been characterized that are regulated by local sequence features within enhancers (Latchman 2010). One of the most well-studied is *dorsal* (*dl*), which acts in the Drosophila dorsal-ventral patterning network (Hong et al. 2008). Comparing Dorsal and Hunchback regulation is useful, as both are bifunctional transcription factors that act as master regulators in orthogonal transcriptional cascades. Dl is an intrinsic activator that can synergize with a co-activator, Twist, to regulate its targets (Shirokawa and Courey 1997). However, Dl also functions as a repressor in the context of the *zerknült* ventral repression element by interacting with bound molecules of the transcription factor Capicua (Shin & Hong 2014). In contrast, we have found that Hb is an intrinsic repressor that can activate *eve2+7* by interacting with bound Cad molecules. Binding site affinity and distance from neighboring sites are important in mediating interactions with Dl (Shirokawa and Courey 1997). However, in both systems, we lack a predictive *cis*-regulatory grammar underlying TF bifunctionality. We anticipate that the quantitative expression data

collected here will be useful in training computational models to learn these rules. This general approach has been successful in understanding other types of TF interactions, such as short-range repression (Fakhouri et al. 2010), although applying this approach to *eve2+7* would require many additional sequence variants. Because counter-repression appears to be a conserved feature in *eve2* orthologs, a bioinformatic analysis of these sequences may yield discrete rules dictating the required spacing or relative affinity of Cad and Hb binding sites.

Recent computational models of enhancer function have pinpointed other potential bifunctional TFs in the gap-gene network (Crocker et al. 2016; Hoermann et al. 2016). As shown by this case study, verifying these modeling predictions requires quantitative expression measurements as well as genetic perturbations in *cis* and *trans*. Indeed, predictions from computational modeling can appear correct for the wrong biological reasons, and many different models can often fit the same quantitative dataset (Ilsley et al. 2013; Staller et al. 2015).

Counter-repression in development

We have shown that Cad acts as a Hb counter-repressor in *eve2+7* regulation. Counter-repression is the converse of short-range repression: the former describes inhibition of repressor activity by an activator, while the latter describes inhibition of activator activity by a repressor (Gray et al. 1994; Kulkarni and Arnosti 2005). We currently do not know how counter-repression is controlled in *eve2+7*, either at the level of binding site grammar or at the level of TF function.

Cad may counter-repress Hb through any number of molecular mechanisms. For example, Cad may interfere with Hb via direct protein-protein interactions. Cad may directly compete with overlapping Hb binding sites, or may occlude Hb protein domains that execute its repressive function (Tran et al. 2010). Cad may also affect Hb activity via more flexible mechanisms analogous to the effects of short-range repressors on nearby activators (Gray et al. 1995). The effect of Cad on Hb binding in *eve2+7* could be directly interrogated using chromatin immunoprecipitation, although genome-wide measurements of Hb binding suggests

that it binds to the region surrounding *eve2+7* (Bradley et al. 2010; Paris et al. 2013). These data suggest that Hb has the capacity to bind *eve2+7* when Cad is also bound, although we cannot rule out quantitative effects on Hb binding without further investigation.

Cad is not the only known counter-repressor in Drosophila development. Stat92E functions as a counter repressor in the formation of the posterior spiracle, but its target repressor remains unknown (Pinto et al. 2015). The genetic signature for counter-repression is complex because it involves mutating binding sites for multiple factors. Specifically, mutations in counter-repressor binding sites can be rescued by mutations in repressor binding sites that have little or no effect when assayed alone. Because binding site mutations are not typically measured in combination, it is possible that other transcription factors that are currently considered necessary activators instead act as counter-repressors. If widespread, counter-repression may explain some amount of context-dependent repressor function in recent genome-wide measurements (Stampfel et al. 2015).

Conclusion

By combining precise genetic perturbations in *cis* and *trans* with quantitative expression measurements, we uncovered a new feature of *eve* stripe 2 regulation that appears conserved in orthologous sequences. We show that Hunchback "bifunctionality" in stripe 7 shadow enhancers is a consequence of direct repression in *eve3+7* and counter-repression in *eve2+7*. This example demonstrates how counter-repression can be detected genetically and bioinformatically, which may be useful for studying TF interactions in development.

Materials and Methods

Fly work

All reporter constructs were cloned into the pBΦY reporter plasmid (Hare et al. 2008) and integrated into the attP2 landing site (Groth et al. 2004). Transformants were

homozygosed, and embryos were fixed, stained and imaged as described in Chapter 2. *sna::hb* embryos were generated as described in Chapter 2.

Binding site predictions for construct design

We used PATSER to predict binding sites in *eve3+7* and *eve2+7* for the TFs expressed in the blastoderm. We used position weight matrices (PWMs) derived from bacterial 1-hybrid experiments for the following factors: Bicoid, Caudal, Dichaete, Stat92E, Hunchback, Krüppel, and Nubbin (Noyes et al. 2008, http://labs.umassmed.edu/WolfeLab/data_files/all_wtmx.txt). We used other published PWMs for the following factors: Giant, Knirps, Tailless (Schroeder et al. 2011). Finally, we used a Zelda PWM whose origins have been lost to the sands of time but we believe was from a personal communication with Christine Rushlow. Count matrices were converted into frequency matrices for use in PATSER using a pseudocount of 1. In designing binding site mutations, we predicted the effects on predicted sites above a p-value cutoff of 0.003. We visualized predictions using InSite software (http://www.cs.utah.edu/~miriah/).

Data analysis

We compiled quantitative data from individual embryos into gene expression atlases as described in Chapter 2. Importantly, we used coarse alignments for both wild-type and *sna*::*hb* embryos. Line traces for individual embryos that were not compared to a *hkb* co-stain were generated using the extractpattern function in the PointCloud toolbox following rotation, alignment and normalization by egg length. Lateral line traces were calculated by averaging the strips 5 and 13, while values for strip 1 represented the ventral line trace. When comparing ventral and lateral line traces for individual embryos, we avoided any stretching of the data by strip (for those in the know, we included the string 'false' in the fourth argument of the extractpattern command). Line traces were corrected for background by subtracting the minimum value and were then normalized by the maximum of the remaining values.

To compare relative expression levels between lines, we used a co-staining method (Wunderlich et al. 2014). This method requires that all relevant embryos are stained in a single batch using the same probes. We ensured that *LacZ* and *hkb* expression values in single embryos are correlated within samples prior to further analysis (rho > 0.6). Once we ensured that correlation values were satisfactory, we calculated line traces using the extractpattern function and normalized these values to the 95 percentile value of of the posterior *hkb* pattern. Individual stripe peaks were determined from these normalized line traces by calculating local maxima.

Binding site overrepresentation analysis

Binding site enrichment was performed as described previously (Wunderlich et al. 2015). The expected number of sites in a sequence of length N was determined by predicting binding sites in all accessible sequences in stage 5 (Thomas et al. 2011), dividing by 9,625,652 (the total number of nucleotides in those sequences), and multiplying by N. Enrichment values were equal to the total number of binding sites observed in a sequence minus the number of binding sites expected in that sequence. To perform this analysis, we used PWMs from FlyFactorSurvey (Zhu et al. 2011) with sequencing data collected on the SOLEXA platform. In this case, we considered sites above a p-value cutoff of 0.05.

References

Alberts, B. et al., 2014. Molecular Biology of the Cell, Sixth Edition, Garland Science.

Andrioli, L.P.M. et al., 2002. Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. Development, 129(21), pp.4931–4940.

Aradhya, S. & Nelson, D.L., 2001. NF-κB signaling and human disease. Current opinion in genetics & development, 11(3), pp.300–306.

Arnosti, D.N. et al., 1996. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development, 122(1), pp.205–214.

Barolo, S., 2012. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. BioEssays: news and reviews in molecular, cellular and developmental biology, 34(2), pp.135–141.

Beck, F., 2004. The role of Cdx genes in the mammalian gut. Gut, 53(10), pp.1394–1396.

Bradley, R.K. et al., 2010. Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related Drosophila species. PLoS biology, 8(3), p.e1000343.

Chawengsaksophak, K. et al., 2004. Cdx2 is essential for axial elongation in mouse development. Proceedings of the National Academy of Sciences of the United States of America, 101(20), pp.7641–7645.

Clyde, D.E. et al., 2003. A self-organizing system of repressor gradients establishes segmental complexity in Drosophila. Nature, 426(6968), pp.849–853.

Copf, T., Schröder, R. & Averof, M., 2004. Ancestral role of caudal genes in axis elongation and segmentation. Proceedings of the National Academy of Sciences of the United States of America, 101(51), pp.17711–17715.

Crocker, J. & Erives, A., 2008. A closer look at the eve stripe 2 enhancers of Drosophila and Themira. PLoS genetics, 4(11), p.e1000276.

Crocker, J., Ilsley, G.R. & Stern, D.L., 2016. Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors. Nature genetics, 48(3), pp.292–298.

Crocker, J. & Stern, D.L., 2017. Functional regulatory evolution outside of the minimal even-skipped stripe 2 enhancer. bioRxiv, p.101311. Available at: http://www.biorxiv.org/content/early/2017/01/18/101311 [Accessed February 16, 2017].

Deng, Z. et al., 2010. Yin Yang 1: a multifaceted protein beyond a transcription factor. Transcription, 1(2), pp.81–84.

Drosophila 12 Genomes Consortium et al., 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature, 450(7167), pp.203–218.

Dubnicoff, T. et al., 1997. Conversion of dorsal from an activator to a repressor by the global corepressor Groucho. Genes & development, 11(22), pp.2952–2957.

Duque, T. et al., 2014. Simulations of enhancer evolution provide mechanistic insights into gene regulation. Molecular biology and evolution, 31(1), pp.184–200.

Fakhouri, W.D. et al., 2010. Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo. Molecular systems biology, 6, p.341.

Fowlkes, C.C. et al., 2008. A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell, 133(2), pp.364–374.

Freund, J.-N. et al., 1998. The Cdx-1 and Cdx-2 homeobox genes in the intestine. Biochemistry and cell biology, 76(6), pp.957–969.

Gallo, S.M. et al., 2011. REDfly v3. 0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic acids research, 39(suppl 1), pp.D118–D123.

Gray, S., Szymanski, P. & Levine, M., 1994. Short-range repression permits multiple enhancers to function autonomously within a complex promoter. Genes & development, 8(15), pp. 1829–1838.

Gray, S. et al., 1995. Transcriptional repression in the Drosophila embryo. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 349(1329), pp. 257–262.

Groth, A.C. et al., 2004. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics, 166(4), pp.1775–1782.

Guo, R.-J., Suh, E.R. & Lynch, J.P., 2004. The role of Cdx proteins in intestinal development and cancer. Cancer biology & therapy, 3(7), pp.593–601.

Häder, T. et al., 1998. Activation of posterior pair-rule stripe expression in response to maternal caudal and zygotic knirps activities. Mechanisms of development, 71(1-2), pp.177–186.

Hare, E.E., Peterson, B.K., Iyer, V.N., et al., 2008. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS genetics, 4(6), p.e1000106.

Hare, E.E., Peterson, B.K. & Eisen, M.B., 2008. A careful look at binding site reorganization in the even-skipped enhancers of Drosophila and sepsids. PLoS genetics, 4(11), p.e1000268.

He, X. et al., 2010. Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. PLoS computational biology, 6(9), p.e1000935.

Hoermann, A., Cicin-Sain, D. & Jaeger, J., 2016. A quantitative validated model reveals two phases of transcriptional regulation for the gap gene giant in Drosophila. Developmental biology, 411(2), pp.325–338.

Hong, J.-W. et al., 2008. How the Dorsal gradient works: insights from postgenome technologies. Proceedings of the National Academy of Sciences of the United States of America, 105(51), pp.20072–20076.

Hui, C.-C. & Angers, S., 2011. Gli proteins in development and disease. Annual review of cell and developmental biology, 27, pp.513–537.

Ilsley, G.R. et al., 2013. Cellular resolution models for even skipped regulation in the entire Drosophila embryo. eLife, 2, p.e00522.

Jaeger, J., 2011. The gap gene network. Cellular and molecular life sciences: CMLS, 68(2), pp. 243–274.

Janssens, H. et al., 2006. Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. Nature genetics, 38(10), pp.1159–1165.

Jiang, J. et al., 1992. Individual dorsal morphogen binding sites mediate activation and repression in the Drosophila embryo. The EMBO journal, 11(8), pp.3147–3154.

Jiang, J. & Levine, M., 1993. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. Cell, 72(5), pp. 741–752.

John, L.B. & Ward, A.C., 2011. The Ikaros gene family: transcriptional regulators of hematopoiesis and immunity. Molecular immunology, 48(9-10), pp.1272–1278.

Kim, A.-R. et al., 2013. Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic. PLoS genetics, 9(2), p.e1003243.

Kim, H.D. et al., 2009. Transcriptional regulatory circuits: predicting numbers from alphabets. Science, 325(5939), pp.429–432.

Kulkarni, M.M. & Arnosti, D.N., 2005. cis-regulatory logic of short-range transcriptional repression in Drosophila melanogaster. Molecular and cellular biology, 25(9), pp.3411–3420.

Arnosti, D.N. & Kulkarni, M.M., 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? Journal of cellular biochemistry, 94(5), pp.890–898.

Latchman, D.S., 2010. Eukaryotic Transcription Factors 5 edition., Academic Press.

Li, L.M. & Arnosti, D.N., 2011. Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes. Current biology, 21(5), pp.406–412.

Ludwig, M.Z. et al., 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature, 403(6769), pp.564–567.

Ludwig, M.Z. et al., 2005. Functional evolution of a cis-regulatory module. PLoS biology, 3(4), p.e93.

Ludwig, M.Z., Patel, N.H. & Kreitman, M., 1998. Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development, 125(5), pp.949–958.

Luengo Hendriks, C.L. et al., 2006. Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. Genome biology, 7(12), p.R123.

Martinez, C. et al., 2014. Ancestral resurrection of the Drosophila S2E enhancer reveals accessible evolutionary paths through compensatory change. Molecular biology and evolution, 31(4), pp.903–916.

McCarty, A.S. et al., 2003. Selective dimerization of a C2H2 zinc finger subfamily. Molecular cell, 11(2), pp.459–470.

Meijsing, S.H. et al., 2009. DNA binding site sequence directs glucocorticoid receptor structure and activity. Science, 324(5925), pp.407–410.

van Nes, J. et al., 2006. The Cdx4 mutation affects axial development and reveals an essential role of Cdx genes in the ontogenesis of the placental labyrinth in mice. Development, 133(3), pp.419–428.

Noyes, M.B. et al., 2008. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nucleic acids research, 36(8), pp.2547–2560.

Nüsslein-Volhard, C. & Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature, 287(5785), pp.795–801.

Olesnicky, E.C. et al., 2006. A caudal mRNA gradient controls posterior development in the wasp Nasonia. Development, 133(20), pp.3973–3982.

Papatsenko, D. & Levine, M.S., 2008. Dual regulation by the Hunchback gradient in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 105(8), pp.2901–2906.

Paris, M. et al., 2013. Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression. PLoS genetics, 9(9), p.e1003748.

Peterson, B.K. et al., 2009. Big genomes facilitate the comparative identification of regulatory elements. PloS one, 4(3), p.e4688.

Pinto, P.B. et al., 2015. JAK/STAT and Hox Dynamic Interactions in an Organogenetic Gene Cascade. PLoS genetics, 11(7), p.e1005412.

Rivera-Pomar, R. et al., 1995. Activation of posterior gap gene expression in the Drosophila blastoderm. Nature, 376(6537), pp.253–256.

Samee, M.A.H. & Sinha, S., 2014. Quantitative modeling of a gene's expression from its intergenic sequence. PLoS computational biology, 10(3), p.e1003467.

Shin, D.-H. & Hong, J.-W., 2014. Capicua is involved in Dorsal-mediated repression of zerknüllt expression in Drosophila embryo. BMB reports, 47(9), pp.518–523.

Shirokawa, J.M. & Courey, A.J., 1997. A direct contact between the dorsal rel homology domain and Twist may mediate transcriptional synergy. Molecular and cellular biology, 17(6), pp.3345–3355.

Simpson-Brose, M., Treisman, J. & Desplan, C., 1994. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. Cell, 78(5), pp. 855–865.

Small, S. et al., 1991. Transcriptional regulation of a pair-rule stripe in Drosophila. Genes & development, 5(5), pp.827–839.

Small, S., Blair, A. & Levine, M., 1992. Regulation of even-skipped stripe 2 in the Drosophila embryo. The EMBO journal, 11(11), pp.4047–4057.

Small, S., Blair, A. & Levine, M., 1996. Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. Developmental biology, 175(2), pp.314–324.

Staller, M.V. et al., 2015. Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 112(3), pp.785–790.

Stampfel, G. et al., 2015. Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature, 528(7580), pp.147–151.

Stanojević, D., Hoey, T. & Levine, M., 1989. Sequence-specific DNA-binding activities of the gap proteins encoded by hunchback and Krüppel in Drosophila. Nature, 341(6240), pp.331–335.

Stanojevic, D., Small, S. & Levine, M., 1991. Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science, 254(5036), pp.1385–1387.

Struffi, P. et al., 2011. Combinatorial activation and concentration-dependent repression of the Drosophila even skipped stripe 3+7 enhancer. Development, 138(19), pp.4291–4299.

Strumpf, D. et al., 2005. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development, 132(9), pp.2093–2102.

Tran, K.D., Miller, M.R. & Doe, C.Q., 2010. Recombineering Hunchback identifies two conserved domains required to maintain neuroblast competence and specify early-born neuronal identity. Development, 137(9), pp.1421–1430.

Thomas, S. et al., 2011. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. Genome biology, 12(5), p.R43.

Vincent, B.J., Estrada, J. & DePace, A.H., 2016. The appeasement of Doug: a synthetic approach to enhancer biology. Integrative biology, 8(4), pp.475–484.

White, R.A. & Lehmann, R., 1986. A gap gene, hunchback, regulates the spatial expression of Ultrabithorax. Cell, 47(2), pp.311–321.

Wingert, R.A. et al., 2007. The cdx genes and retinoic acid control the positioning and segmentation of the zebrafish pronephros. PLoS genetics, 3(10), pp.1922–1938.

Wunderlich, Z. et al., 2015. Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. Cell reports, 12(11), pp.1740–1747.

Wunderlich, Z., Bragdon, M.D. & DePace, A.H., 2014. Comparing mRNA levels using in situ hybridization of a target gene and co-stain. Methods, 68(1), pp.233–241.

Wu, X. et al., 2001. Thoracic patterning by the Drosophila gap gene hunchback. Developmental biology, 237(1), pp.79–92.

Zhang, C.C. & Bienz, M., 1992. Segmental determination in Drosophila conferred by hunchback (hb), a repressor of the homeotic gene Ultrabithorax (Ubx). Proceedings of the National Academy of Sciences of the United States of America, 89(16), pp.7511–7515.

Zhu, L.J. et al., 2011. FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system. Nucleic acids research, 39(Database issue), pp.D111–7.

CHAPTER 4: DEFINING THE SOLUTION SPACE FOR THE *EVEN-SKIPPED* EXPRESSION PATTERN SUGGESTS REGULATORY PLASTICITY IN DROSOPHILA

Ben J. Vincent, Meghan D.J. Bragdon, Garth Ilsley, Tara Lydiard-Martin, Clarissa Scholes, Zeba Wunderlich, Javier Estrada, and Angela H. DePace

Author Contributions:

B.J.V., G.I., T.L.-M. and A.H.D. designed the study; B.J.V. and G.I. performed computational modeling; B.J.V., M.D.J.B. and C.S. performed experiments; Z.W. and J.E. processed raw images for further analysis; B.J.V. analyzed the data; and B.J.V. and A.H.D. wrote the manuscript.

Abstract

Developmental genes are regulated by multiple enhancers, each of which generates a portion of the total expression pattern. Within enhancers, transcription factor binding sites can change over evolutionary time while maintaining enhancer function. Here, we investigate whether the function of individual enhancers can change while maintaining the expression pattern of the whole locus. As a case study, we focus on the Drosophila gene *even-skipped*, which is expressed in seven stripes along the anterior-posterior axis in the blastoderm embryo. By combining computational modeling with enhancer engineering, we predict alternative ways to partition the seven even-*skipped stripes* among a set of modular enhancers, and we validate these predictions by generating new expression patterns in embryos. Our work suggests that there may be many ways to build a complex developmental expression pattern, which may influence how developmental loci evolve.

Introduction

Changes in DNA sequence underlie the staggering diversity of animal form and function that we observe in the natural world. Many of the sequence changes that cause phenotypic differences occur in segments of regulatory DNA called enhancers, especially over short evolutionary timescales (Rebeiz et al. 2015). Enhancers bind cell type-specific transcription factors (TFs) to turn genes on and off in response to developmental or environmental cues (Long et al. 2016). Changes in enhancers, and resulting changes to gene expression, have been linked to phenotypic evolution in animal case studies (Gompel et al. 2005; Wittkopp et al. 2008; Prud'homme et al. 2006; Reed et al. 2011; Prud'homme et al. 2011; Glassford et al. 2015; Indjeian et al. 2016). Out of the many variants that occur within enhancers, how can we identify the ones that have phenotypic consequences? Progress on this question would be directly applicable to human health, as the majority of disease-causing variants between human individuals map to regulatory regions (Maurano et al. 2012).

Interpreting regulatory sequence variation is complicated by redundancy and compensation. Regulatory sequences contain redundancy at multiple scales. Enhancers can contain several binding sites for the same transcription factor, such that changes in single site may not affect expression (Arnosti et al. 1996; Jiang et al. 2015). In addition, regulatory information can be redundantly encoded by multiple enhancers (Hong et al. 2008; Barolo 2012). These "shadow enhancers" contribute to the robustness of gene expression in development (Frankel et al. 2010; Perry et al. 2010; Perry et al. 2011), possibly by generating the same expression patterns using different regulatory logic (Staller et al. 2015; Wunderlich et al. 2015; Cannavò et al. 2016). As another complicating factor, compensatory evolution can act both within and between enhancers (Kimura 1985). Within an enhancer, TF binding sites can change in position, strength and number while enhancer function remains conserved (Ludwig et al. 1998; Ludwig et al. 2000; Ludwig et al. 2005; Hare et al. 2008; Peterson et al. 2009). Compensatory evolution also occurs between shadow enhancers, which can diverge in function while the total expression pattern of the gene remains conserved (Wunderlich et al. 2015). These

examples illustrate the remarkable evolutionary plasticity of developmental gene regulation, a consequence of the fact that even highly specific expression patterns can be generated in multiple ways.

The central hypothesis of this work is that compensatory evolution can occur between modular enhancers within a developmental locus. To validate this hypothesis, we must identify a case where orthologous genes have a conserved expression pattern, but where the component enhancers have functionally diverged. We can search for enhancers with diverged activity either experimentally or computationally. Experimentally, we can measure enhancer activity in high throughput using cell culture assays (Arnold et al. 2014; Villar et al. 2015). While this approach allows for genome-wide screening, measurements in cell culture do not necessarily reflect enhancer behavior in intact animals (Barolo & Posakony 2002). Alternatively, we can screen candidate enhancers using reporter constructs in embryos, but this approach is limited by the low throughput step of creating transgenic animals. Indeed, this approach is likely to fail unless we have a reasonable expectation that enhancer divergence is possible at short evolutionary timescales. We take a different approach by computationally defining the solution space of a complex expression pattern based on the location and function of upstream TFs.

As a case study, we focus on the *even-skipped* (*eve*) locus in Drosophila (Figure 4.1). *eve* is pair-rule gene that is expressed in seven transverse stripes along the anterior-posterior (AP) axis of the Drosophila embryo (Nüsslein-Volhard & Wieschaus 1980; Harding et al. 1986; Frasch et al. 1987). This pattern is controlled by five enhancers, each of which generates one or two stripes when placed in *LacZ* reporter constructs (Small et al. 1991; Small et al. 1996; Fujioka et al. 1999; Staller et al. 2015). As textbook examples of enhancer modularity, the *eve* enhancers provide an excellent system to investigate regulatory evolution at the locus-level (Alberts et al. 2014). Are there different ways to partition the seven *eve* stripes among a set of modular enhancers? This question has been the subject of speculation since the *eve* locus was first dissected (Akam 1989; Goltsev et al. 2004). Here, we define computational and experimental methods to address it directly.

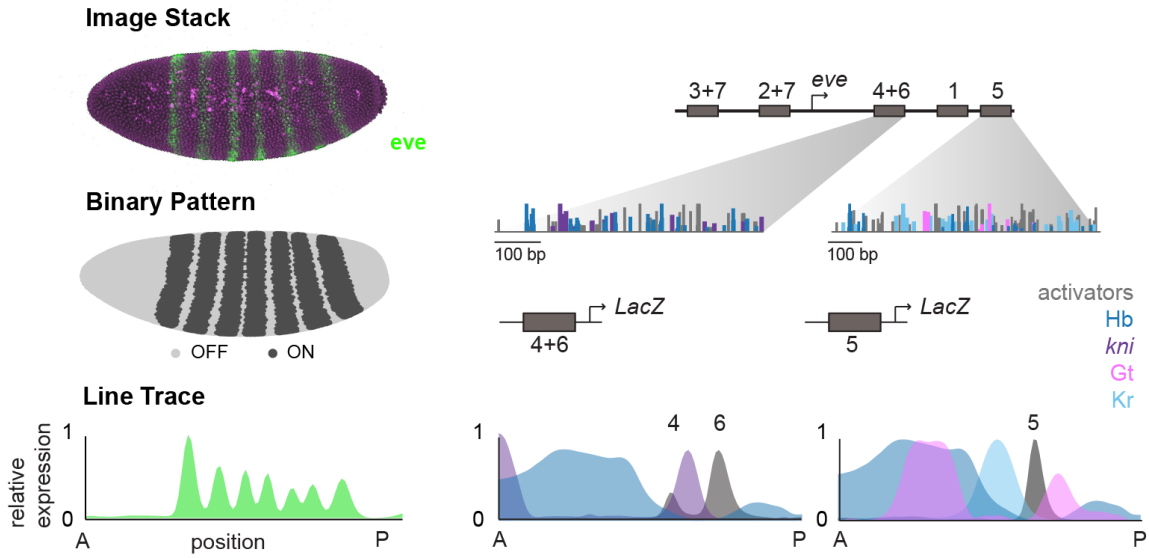


Figure 4.1: *even-skipped* is an ideal system to study locus evolution. Left: *even-skipped* (*eve*) is expressed in seven transverse stripes along the anterior-posterior (AP) axis of the blastoderm embryo in *Drosophila melanogaster*. We collect quantitative gene expression data at cellular resolution for *eve* and its regulators using fluorescent *in situ* hybridization or immunohistochemistry and we image individual embryos using 2-photon confocal microscopy. Shown here is a maximum-intensity projection of a representative image stack taken from a single embryo at time point 3 in the Virtual Embryo (Top). *eve* mRNA is labelled in green while nuclei stained with SYTOX green are labelled in magenta. We use a computational pipeline to assign relative gene expression measurements to individual cells, and we can register measurements from many embryos into a common morphological template to generate a gene expression atlas. We can plot atlas data as a binary pattern after applying a threshold of 0.2 (Middle) or as a line trace along the lateral side of the embryo (Bottom). Right: The seven *eve* stripes are generated by 5 annotated enhancers, each of which generates one or two individual stripes. These enhancers contain predicted bindings sites (vertical bars, height is proportional to the PATSER-predicted score for each site) for transcription factors that are themselves expressed in patterns. *eve4+6* is regulated by two repressors, *hunchback* (*hb*) and *knirps* (*kni*), while *eve5* is regulated by three repressors, *hb*, *giant* (*gt*) and *Krüppel* (*Kr*). Both enhancers are also sensitive to ubiquitously-expressed activators. Atlas data for *LacZ* reporter patterns are plotted as line traces (dark grey), as are the expression patterns for Hb, Gt and Kr protein and *knirps* mRNA.

To identify alternative solutions for the seven-stripe *eve* pattern, we use computational models of enhancer function that predict expression in single cells (Ilsley et al. 2013). These empirical models can identify candidate enhancer regulators (Ilsley et al. 2013), dissect the causes of expression pattern divergence (Wunderlich et al. 2012), pinpoint additional sources of regulatory DNA in a complex locus (Staller et al. 2015), and predict how enhancers respond to

engineered transcription factors (Crocker et al. 2016). Here, we used them to predict how different *eve* stripe combinations can be generated from endogenous enhancers in *Drosophila melanogaster*. We then validate these predictions by measuring the output of engineered enhancers in embryos. These methods suggest that alternative regulatory solutions are possible for the *eve* locus, which makes it an attractive candidate for a thorough investigation of enhancer divergence.

Results

Our goal in this study was to define the solution space for the *eve* expression pattern: how many different ways can we generate seven *eve* stripes from a suite of enhancers? To accomplish this goal, we used logistic regression to screen all possible stripe combinations for plausible patterns given the constraints imposed by the upstream regulatory network. This approach suggested alternatives to the solution encoded by the endogenous set of *eve* enhancers in *Drosophila melanogaster*. We then generated new stripe combinations by enhancer engineering. Finally, we computationally screened orthologous *eve* enhancers for diverged function and tested candidates in reporter constructs. Our results suggest that even in developmental genes with complex expression patterns, there are many ways to encode the same regulatory information.

Logistic regression predicts new eve *stripe combinations within the constraints of the upstream regulatory network.*

We hypothesized that computational models of *eve* enhancer function could accurately fit both endogenous and nonendogenous *eve* stripe combinations. To test this hypothesis, we used logistic regression to relate measurements for a common set of input regulators to all possible combinations of stripes (Ilsley et al. 2013). This approach determines whether the input regulators contain sufficient positional information to specify a given expression pattern. All data for fitting these models were taken from the Virtual Embryo dataset, which includes

quantitative cellular-resolution expression measurements for 95 genes at 6 time points during stage 5 of *Drosophila melanogaster* embryogenesis (Fowlkes et al. 2008). We used data from the third time point for our analysis. As input TFs, we included protein data for Bicoid (Bcd), Hunchback (Hb), Giant (Gt) and Krüppel (Kr) and mRNA data for *caudal* (*cad*), *knirps* (*kni*) and *tailless* (*tll*) (Figure 4.2A). Our models also included a constant term and a quadratic term for Bcd (Bcd^2) that was previously found to be useful in modeling the regulation of *eve* stripe 2 (Ilsley et al 2013). To generate target patterns, we thresholded the *eve* expression pattern into ON and OFF cells (threshold = 0.2), identified and selected ON cells within each stripe, and joined them in all 127 possible combinations of seven stripes. Models were trained on each of these binary patterns and model predictions were themselves thresholded (threshold = 0.2) (Figure 4.2A) and evaluated using the Matthews correlation coefficient (MCC), a convenient metric which summarizes the confusion matrix for binary predictions (Matthews 1975). MCC values range from -1 to 1, where 1 represents perfect model predictions, 0 represents random predictions, and -1 represents complete disagreement between predictions and observations.

We found that logistic regression could fit both endogenous and nonendogenous stripe combinations with high accuracy. Models for all endogenous patterns predicted expression in the expected domains of one or two stripes (Figure 4.2C). Although three endogenous enhancers (*eve2+7*, *eve3+7*, and *eve4+6*) express a total of five stripes in pairs, models fit on each individual stripe performed very well (MCC > 0.7, Figure 4.2 B and C). These results suggested that these stripes may be produced separately. Finally, our approach predicted that some stripes generated by separate endogenous enhancers may be produced together. Specifically, models fit on combinations of stripes 2 and 5 as well as 5 and 7 accurately predicted expression in two distinct stripes (Figure 4.2C). We recently showed that an extended version (*eve2+7*) of the minimal stripe 2 enhancer contributes to patterning of endogenous stripe 7 (Staller et al. 2015), but neither stripe 2 nor stripe 7 activity have not been previously associated with the stripe 5 enhancer (*eve5*). Finally, a single three stripe pattern combining stripes 2, 5 and 7 was fit accurately (Figure 4.2 B and C, MCC = 0.799). MCC scores for some other patterns containing

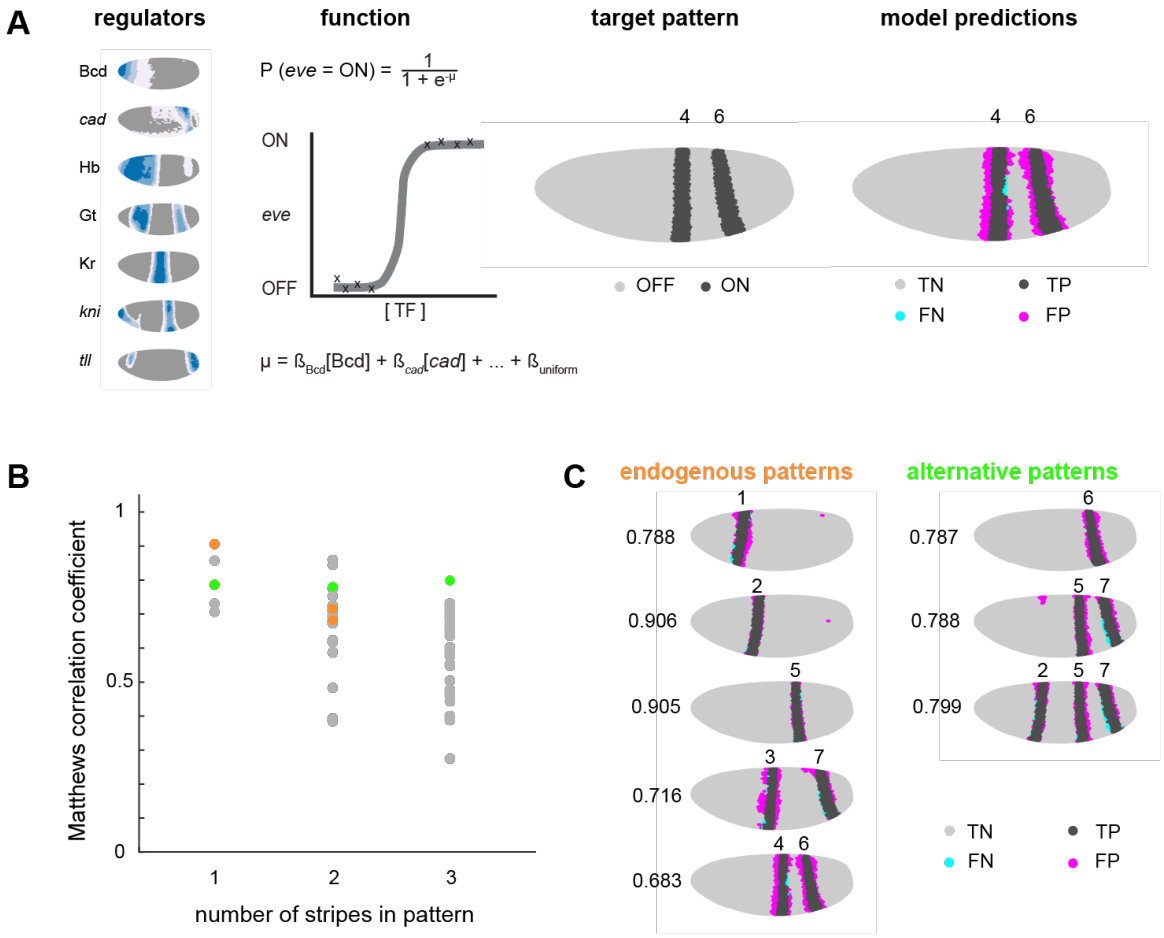


Figure 4.2: A computational screen to identify alternative solutions for the *eve* expression pattern. (A) Overview of our approach. We perform logistic regression to fit functions that use quantitative expression measurements for *eve* regulators (plotted here where darker blue indicates higher relative expression levels in the embryo) to predict binary *eve* patterns in all 6,078 cells in the Virtual Embryo. As an example, we fit one parameter per regulator using a binary pattern consisting of *eve* stripes 4 and 6 from our gene expression atlas ("target pattern", thresholded at a fixed value of 0.2). We then used these parameters to predict expression in each cell, and again thresholded expression values at a fixed value of 0.2 to compare our predictions to the target pattern ("model predictions", TP = true positives, TN = true negatives, FP = false positives, FN = false negatives). Our model of *eve* stripes 4 and 6 fits the data well – it predicts expression in two distinct stripes with some misspecified cells on the stripe borders. (B) We used this approach to generate models for all 1, 2 and 3 stripe combinations of *eve* stripes. We evaluated the performance of each model by calculating the Matthews correlation coefficient (MCC). (C) We find that models fit to all endogenous patterns performed well (orange). Functions fit to other 1 or 2 stripe patterns also performed well, which indicated that these patterns may be encoded by modifying existing *eve* enhancers. Finally, the function fit to one 3 stripe pattern, stripes 2, 5 and 7, performed well. We chose three patterns for experimental follow-up (green).

three or more stripes were relatively high (Figure 4.2B), but none of these other models predicted the expression of three distinct stripes (data not shown). Model predictions for all combinations of one and two stripes are shown in Supplemental Figure 4.1. We also performed this same analysis using the area under the receiver operating curve (AUC), which evaluates predictions over many thresholds, and our results were consistent (Supplemental Figure 4.2).

Our results demonstrate that logistic regression is a simple and effective tool for predicting alternative ways to encode a seven-stripe expression pattern. However, this analysis alone was not sufficient to predict how to generate alternative expression patterns from existing enhancers or screen enhancer sequences in other species for diverged activity. We addressed these challenges for two candidate enhancers: *eve4+6* and *eve5*.

Adding Krüppel sites to eve4+6 *generates stripe 6 alone.*

Although they did not explicitly include DNA sequence, our models can nonetheless predict how to generate new expression patterns by changing the sequence features of annotated enhancers. Our general approach was to fit models using expression measurements driven by *LacZ* reporter constructs containing endogenous enhancers, hereafter referred to as "reporter patterns." We then simulated the effects of adding or removing binding sites for individual regulators by altering their fit parameters and predicting the effects on expression. Because these models are simple to fit and evaluate, we were able to generate testable predictions quickly and efficiently.

We applied this approach to our first target: *eve* stripe 6. Stripe 6 is normally generated with stripe 4 from the *eve4+6* enhancer (Fujioka et al. 1999). To determine how to alter *eve4+6* to generate stripe 6 alone, we used logistic regression to fit a model for the *eve4+6* reporter pattern, which we measured at cellular resolution and thresholded at a fixed value of 0.2 (Supplemental Figure 4.3A). This model predicted expression in two stripes in the embryo (MCC = 0.694, Supplemental Figure 4.3C). To predict how to remove stripe 4 expression without affecting stripe 6, we generated a target pattern that consisted of the cells in stripe 6

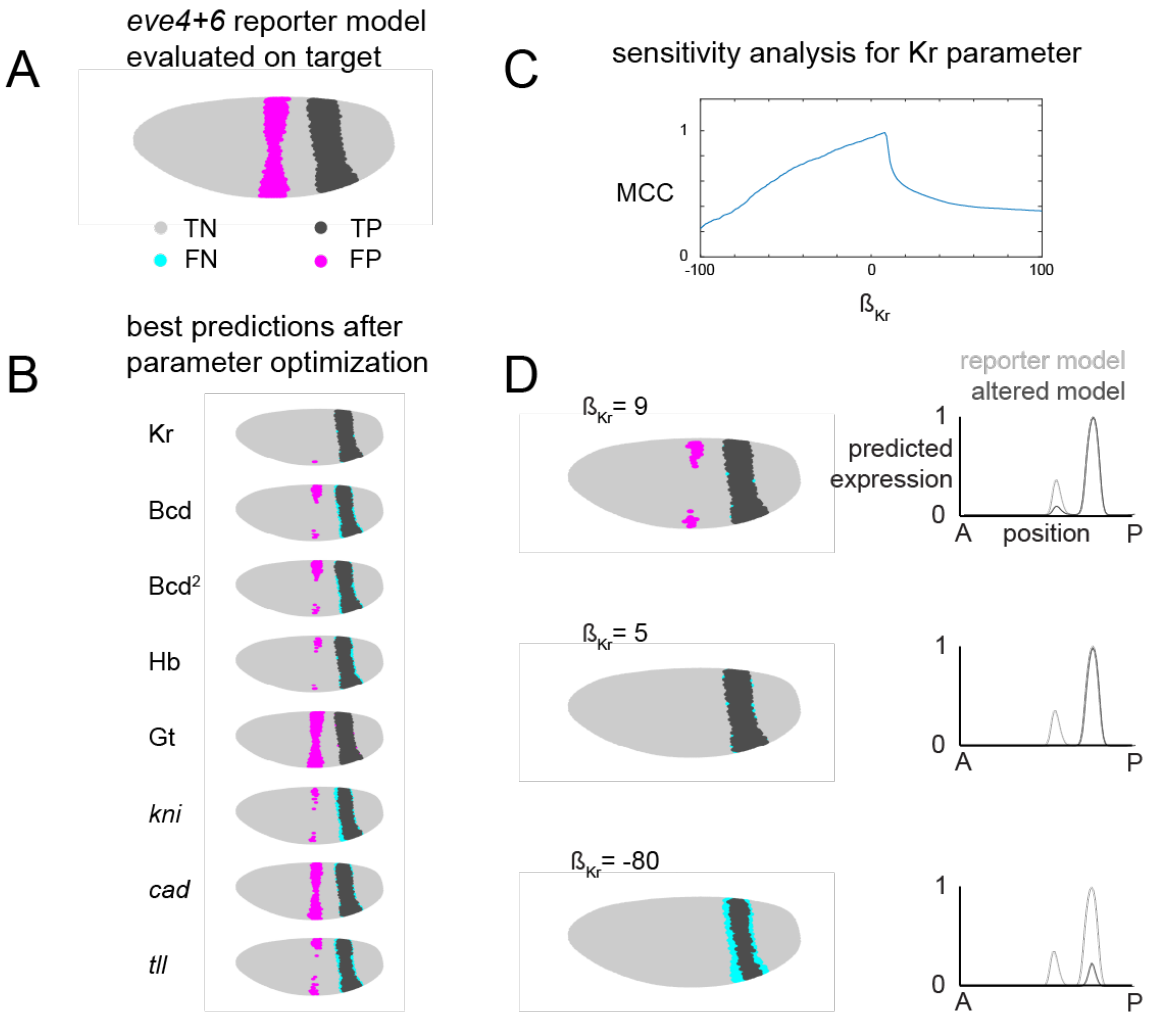


Figure 4.3: A sensitivity analysis predicts how to generate stripe 6 by engineering *eve4+6*. (A) We fit a model using thresholded data collected from an *eve4+6 LacZ* reporter construct. We then used this model to predict expression in all cells and generate a target pattern consisting of predicted ON cells in stripe 6 alone. When we evaluate our eve4+6 model on this target pattern, we see false positives in stripe 4 and perfect predictions in stripe 6. (B) We varied the parameter associated with each regulator between values of -2500 and 2500 and evaluated model performance by calculating the MCC. Optimizing the Kr parameter yielded the greatest improvement in model performance. (C) Model performance is plotted as a function of the Kr parameter value. Small decreases in the Kr parameter improve model performance, but larger decreases diminish model performance due to predicted effects on stripe 6 expression. (D) Predicted expression patterns for three Kr parameter values. Decreasing the Kr parameter to a value of 9 decreases predicted stripe 4 expression, while decreasing the Kr parameter to a value of 5 abolishes predicted stripe 4 expression completely. Decreasing the Kr parameter further lowers predicted expression of stripe 6.

predicted ON by our model (Supplemental Figure 4.3D). Evaluating our *eve4+6* model on this target pattern resulted in false positives in stripe 4 and perfect predictions for stripe 6 (MCC = 0.7515, Figure 4.3A).

Our goal was to predict the simplest sequence manipulations required to generate stripe 6 alone. Formally, this required performing a sensitivity analyses for each individual parameter in the *eve4+6* reporter pattern. We varied each parameter between -2500 and 2500, predicted the effect on expression, and calculated the MCC using our stripe 6 target pattern as a gold standard. We found that while predicted expression of stripe 4 decreases after optimizing many individual parameters, decreasing the parameter for Krüppel (Kr) could abolish predicted stripe 4 expression with no predicted effect on stripe 6 (Figure 4.3B). However, our model predicts that expression is highly sensitive to Kr influence: small decreases in the Kr parameter decrease stripe 4 expression without abolishing it, whereas large decreases abolish stripe 4 but also affect stripe 6 expression (Figure 4.3C and D).

To test these modeling predictions, we introduced Kr binding sites into *eve4+6* and measured the effect on expression. This experiment is challenging for two reasons. First, introducing binding sites within an enhancer can create or destroy binding sites for other important transcription factors. To control for this, we predicted binding sites for known *eve4+6* regulators before and after we introduced new binding sites (Figure 4.4A). Second, a transcription factor's efficacy within an enhancer is a function of the number, location and affinity of its binding sites. How many Kr sites should we introduce, where should we put them, and what affinity should they be? We chose three different configurations to test in reporter constructs (Figure 4.4). The first two configurations included annotated Kr sites from the *eve* stripe 2 enhancer (Gallo et al. 2011); these sites have low predicted affinity based on bacterial-1 hybrid experiments ("weak sites", Noyes et al. 2008). In one configuration, we introduced 4 additional sites in the middle of the enhancer (Figure 4.4B, *eve4+6 construct 1*), while in the other, we introduced 10 additional sites across the enhancer (Figure 4.4C, *eve4+6 construct 2*). The last configuration (*eve4+6 construct 3*) included 4 sites with high predicted affinity based

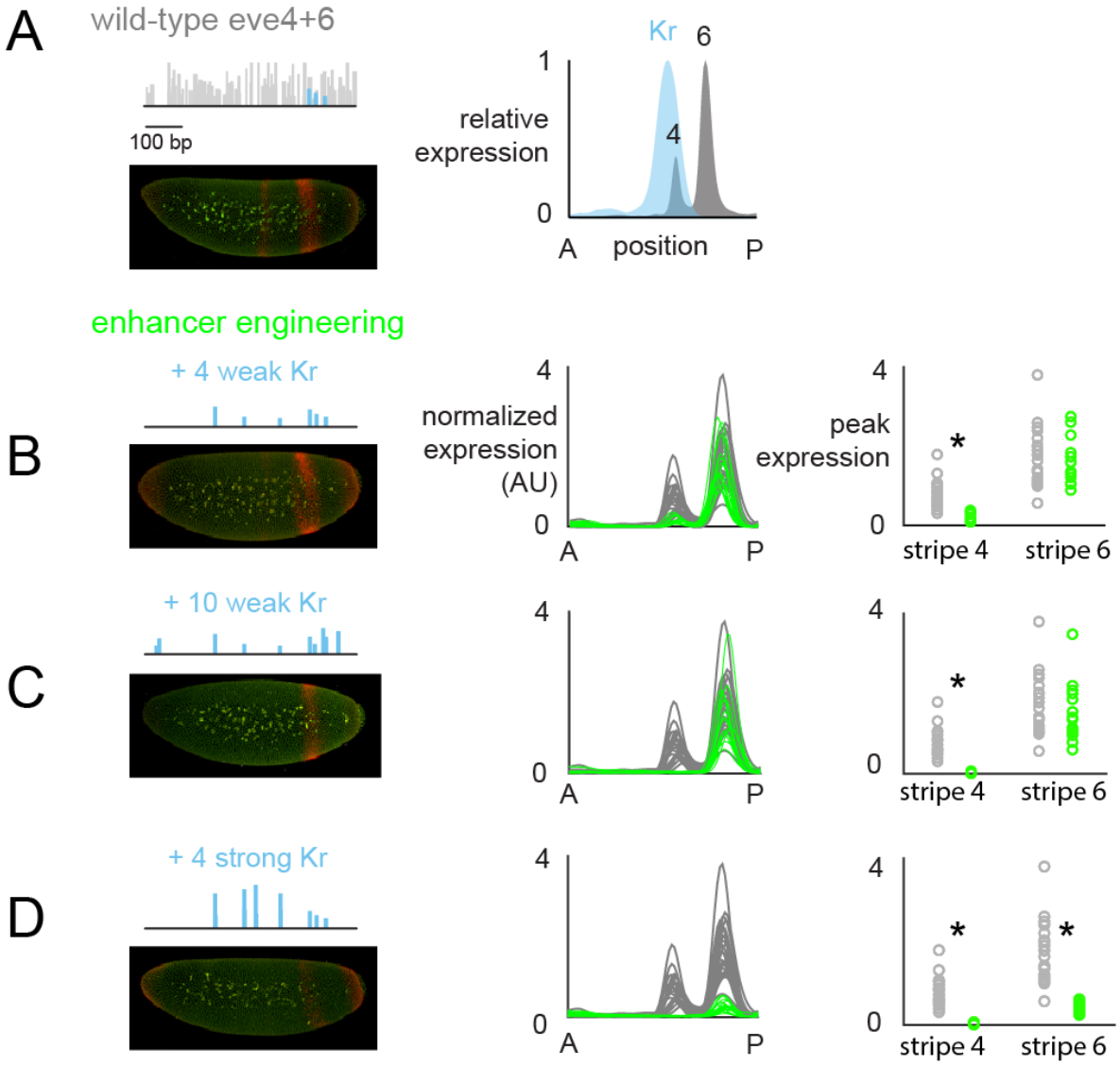


Figure 4.4: Adding Kr sites to *eve4+6* generates stripe 6 alone. (A) Left: Predicted binding sites for known *eve4+6* regulators displayed as in Figure 4.1. Predicted binding sites for Kr are shown in blue. Embryos containing an *eve4+6* reporter construct express *LacZ* in two stripes. Shown here is a maximum intensity projection taken from a representative image stack of an embryo at time point 3 stained for *LacZ* mRNA (red) with nuclei labelled using SYTOX green (green). Right: Line traces from atlas data for Kr protein and *eve4+6 LacZ* mRNA.(B) Left: Predicted Kr binding sites in *eve4+6 construct 1*. Middle: Line traces from individual embryos containing reporter constructs for *eve4+6* (grey) or *eve4+6 construct 1* (green). Line traces were normalized using a *hkb* co-stain. Right: Peak stripe levels from individual line traces. Asterisks indicate significant differences in peak level between *eve4+6* and *eve4+6 construct 1* (p-value < 0.001, Mann-Whitney U test). We observed no significant difference in stripe 6 peaks (p-value > 0.2). (C) *eve4+6 construct* 2. Data represented as in (B). (D) *eve4+6 construct 3*.

on bacterial-1 hybrid experiments ("strong sites") in the same position as those added to *eve4+6 construct 1* (Figure 4.4D). We synthesized these modified enhancers, cloned them into reporter constructs, and measured their expression patterns quantitatively at cellular resolution using a *huckebein* co-stain to detect differences in relative stripe levels between lines (Wunderlich, et al. 2014).

We found that addition of 4 weak Kr sites decreased stripe 4 expression without affecting stripe 6 expression (Figure 4.4B). Addition of 10 weak sites abolished stripe 4 completely without affecting stripe 6 expression (Figure 4.4C). Addition of 4 strong sites abolished expression of stripe 4, but also decreased expression of stripe 6 (Figure 4.4D). This effect is due to the direct influence of Kr protein, as the *eve4+6 construct 3* reporter pattern was indistinguishable from the *eve4+6* reporter pattern when the constructs were introduced into a Kr mutant background (Supplemental Figure 4.4). These results validated our modeling predictions and demonstrate that an alternative pattern can be generated from an endogenous enhancer by increasing sensitivity to a single regulator. We also demonstrated that strong binding sites can repress gene expression even in areas of the embryo with very low concentrations of that factor. This result could explain why endogenous *eve* enhancers such as *eve2* contain many weak binding sites as opposed to a few strong binding sites (Crocker et al. 2016; Farley et al. 2016).

Orthologous enhancers drive different relative stripe levels.

Our synthetic experiments suggested that stripe 4 expression can be modulated relative to stripe 6. We hypothesized that we might observe the same phenomenon in orthologous *eve4+6* enhancers. We used the LiftOver tool in the UCSC Genome Browser to identify orthologous *eve4+6* sequences in *Drosophila yakuba* (*D. yak*), *Drosophila pseudoobscura* (*D. pse*) and *Drosophila virilis* (*D. vir*), and we measured expression driven by *LacZ* reporter constructs containing these enhancers.

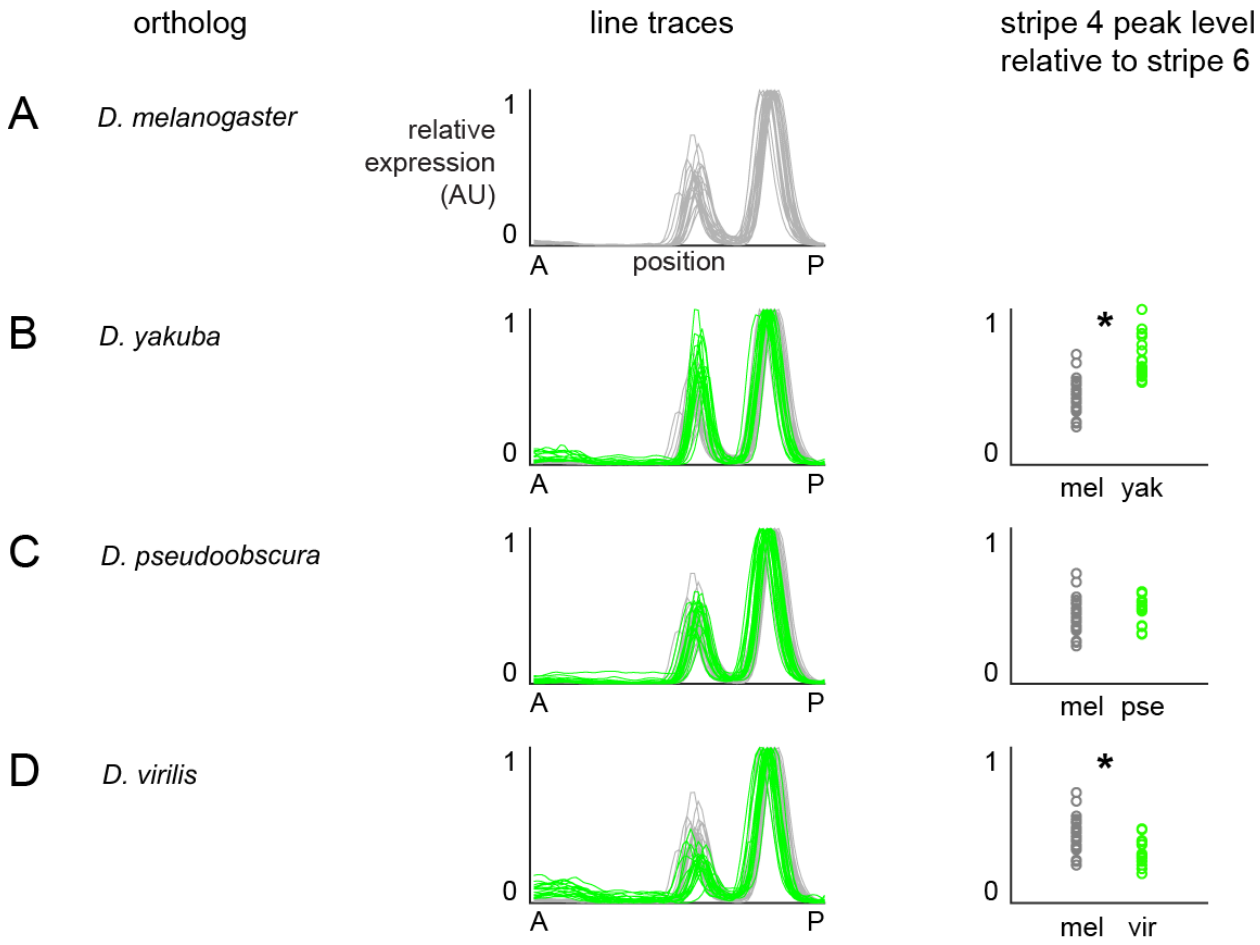


Figure 4.5: Orthologous *eve4+6* enhancers drive different levels of stripe 4 relative to stripe 6. (A) Middle: Line traces from individual embryos containing an *eve4+6* reporter construct from *D. mel*. Traces are normalized by maximum value of stripe 6. (B) Middle: Line traces from individual embryos containing *eve4+6* reporter constructs from *D. mel* (grey) or *D. yak* (green). Right: Stripe 4 peaks were determined by taking the local maximum value of the normalized stripe 4 line trace. Asterisks indicate significant differences in stripe peaks between *D. mel* and *D. yak* reporter patterns (p-value < 0.001, Mann-Whitney U test). (C and D) Same as (B), but for enhancers from *D. pse* and *D. vir*, respectively.

We found that relative to stripe 6, stripe 4 levels in the *D. yak* reporter pattern were significantly higher compared to the *D. mel* reporter pattern (Figure 4.5B). The *D. pse eve4+6* reporter pattern was indistinguishable from the *D. mel* reporter pattern (Figure 4.5C), while stripe 4 levels in the *D. vir* reporter pattern were significantly lower compared to the *D. mel* reporter pattern (Figure 4.5D). However, it remains difficult to map these changes onto specific sequence features because binding sites for many TFs are different within these enhancers

(Supplemental Figure 4.5). These results suggest that stripe 4 levels can be modulated relative to stripe 6 levels in orthologous *eve4+6* enhancers.

A computational search for divergence in eve4+6.

Our synthetic experiments suggested that Kr sites within *eve4+6* could modulate stripe 4 expression relative to stripe 6, and in the extreme case, abolish stripe 4 completely. It is therefore possible that an '*eve6*' enhancer exists within the Drosophila phylogeny. However, maintenance of the seven-stripe *eve* pattern would require a complementary '*eve4*' enhancer within such a species. How could an *eve* stripe 4 pattern be generated?

We repeated our computational analysis with a second target: *eve* stripe 4. In this case, our target pattern consisted of cells in stripe 4 predicted ON by our *eve4+6* reporter model (Figure 4.6A, see logic outlined in Supplemental Figure 4.3) We varied each parameter in the *eve4+6* model to maximize performance when evaluated on this target. We found that decreasing the parameter for Giant was sufficient to decrease expression of stripe 6 with limited impact on stripe 4 (Figure 4.6B). Giant is a repressor that is expressed in stripe 6 cells (Stanojevic et al. 1991; Strunk et al. 2001). This analysis suggests that Gt sites within *eve4+6* may modulate stripe 6 expression relative to stripe 4.

While our computational and experimental work suggest that separation of stripes 4 and 6 into distinct enhancers is a formal possibility, additional methods are required to locate such sequences in endogenous DNA. This goal is challenging because enhancer boundaries are ill-defined and the regulatory information that specifies a given pattern may be distributed over a large region (Fujioka and Jaynes 2012; Montavon et al. 2011). We therefore chose to scan windows of endogenous sequence and calculate binding site enrichment over a background distribution consisting of predicted sites within all DNase-accessible regions in stage 5 (Wunderlich et al. 2015; Thomas et al. 2011). For *D. mel eve4+6*, we chose a window size of 500bp, which is similar to the size of experimentally-minimized *eve* enhancers (Small et al. 1992; Small et al. 1996). We also calculated TF enrichment in sequences surrounding *eve4+6*,

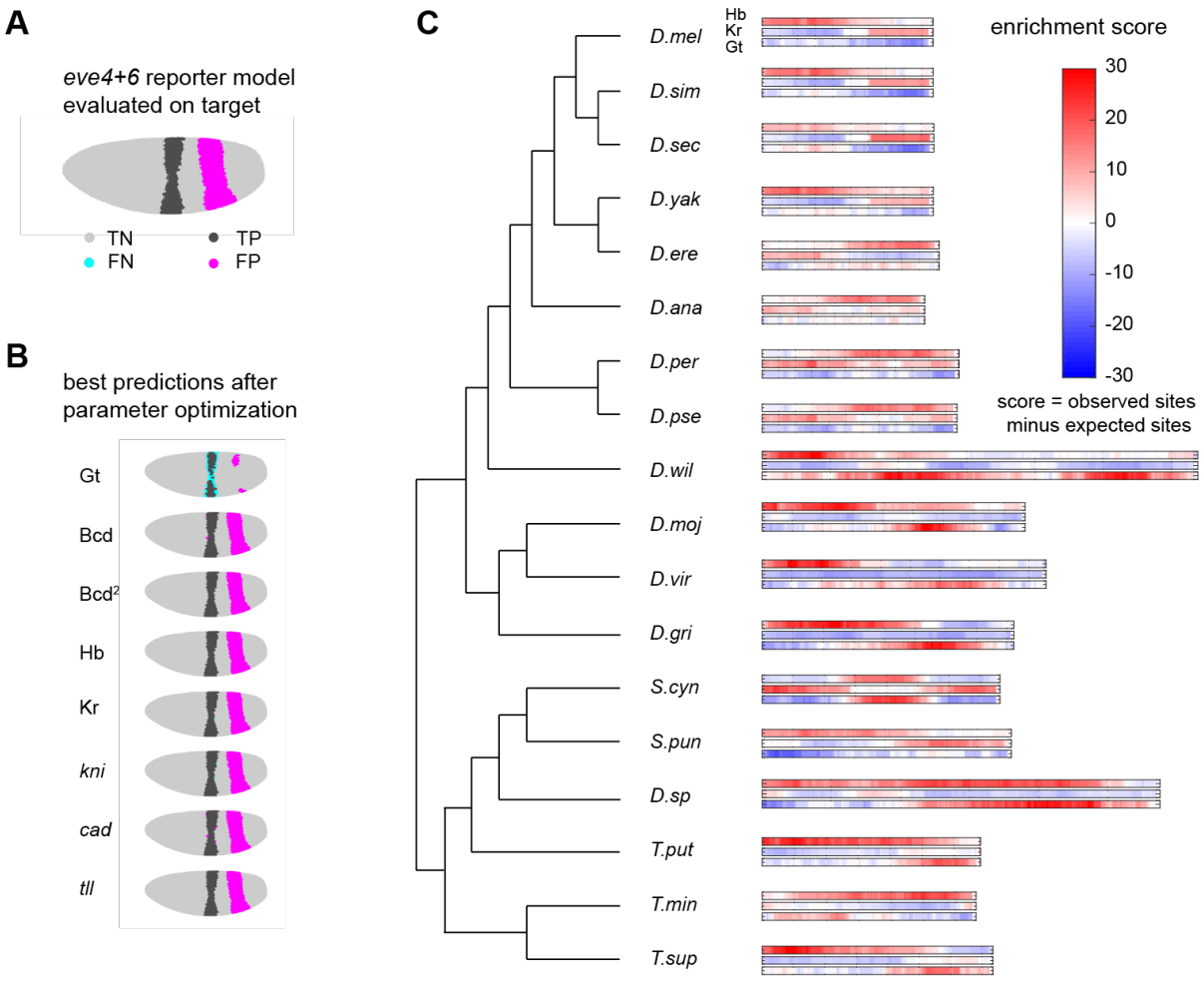


Figure 4.6: A bioinformatic search for functional divergence in orthologous *eve4+6* enhancers. (A) *eve4+6* model predictions evaluated on the target pattern of stripe 4 alone. (B) Sensitivity analyses predicted how to generate a target pattern of stripe 4 alone. In this case, optimizing the Giant parameter (Gt) decreased expression of stripe 6, but also decreased expression of stripe 4. (C) We hypothesized that Kr or Gt binding sites may modulate *eve4+6* expression in orthologous enhancers. To find regions of TF enrichment, we predicted binding sites for Hb, Kr and Gt in 500bp windows within an extended version of *eve4+6* from *Drosophila melanogaster* (*D. mel*). Using predicted binding sites within DNase accessible regions during state 5 as a background distribution (Thomas et al. 2011), we calculated an enrichment score for each factor within each window (Wunderlich, et al. 2015). As expected, the annotated *eve4+6* enhancer in *D. mel* is highly enriched in Hb sites, but a flanking region is also enriched in Kr sites. This pattern is consistent in Drosophila species through *Drosophila pseudoobscura* (*D. pse*). However, in other species, such as *Drosophila mojavensis* (*D. moj*) and *Drosophila virilis* (*D. vir*), flanking *eve4+6* sequences are enriched in Gt sites but not Kr sites. Within orthologous *eve4+6* enhancers from Sepsid species (Hare, et al. 2008), this pattern is consistent: a 'core' regions enriched in Hb sites, and a flanking region is enriched in either Gt or Kr sites but not both. There are two exceptions to this rule: in both *Sepsis cynipsea* (*S. cyn*) and *Dicranosepsis sp.* (*D. sp*), the 'core' region is enriched in both Hb and Gt sites.

since we found that different fragments of the *eve4+6* enhancer region drove expression patterns with different relative stripe levels (Supplemental Figure 4.6).

We found that the annotated *eve4+6* sequence is highly enriched in sites for Hb, the repressor that determines the anterior boundary of stripe 4 and the posterior boundary of stripe 6 (Figure 4.5C; see Figure 4.1). As Hb is critical for the *eve4+6* expression pattern, it is a convenient marker for the 'core' enhancer element. This same sequence is not enriched for either Kr or Gt sites, which is consistent with robust expression of both stripes (Figure 4.5C). However, the sequence downstream of *eve4+6* is highly enriched in Kr sites, which suggests that Kr may influence stripe 4 expression (Figure 4.5C). We therefore created a series of reporter constructs that test the effect of this flanking element on expression. However, we detected no effect on the level of stripe 4 expression after this enriched flanking region was replaced with computationally designed "neutral" DNA (Supplementary Figure 4.7; Estrada et al. 2016).

To identify candidate sequences with diverged function, we performed the same binding site enrichment analysis on a series of previously identified orthologous *eve4+6* enhancers (Hare et al. 2008). All orthologous sequences contained regions that were enriched in Hb sites (Figure 4.5C). In addition, we identified regions in enhancers from *Sepsis cynipsea* and *Dicranosepsis sp*. that were enriched in both Hb sites and Gt sites (Figure 4.5C). We predicted that these regions would drive expression of stripe 6 alone. However, when we introduced a candidate enhancer fragment from *Sepsis cynipsea* into a reporter construct, it drove no detectable expression (data not shown, See Appendix D for sequence). These results indicate that further analysis is needed to detect functional divergence in orthologous *eve4+6* enhancers.

Engineering eve5 *generates new patterns including stripes 2 and 7.*

We applied our modeling approach to another target: *eve* stripes 2, 5 and 7. We fit a model on thresholded expression measurements of a reporter construct containing the annotated *eve5* enhancer sequence (Fujioka et al. 1999; Gallo et al. 2011). We used this model to predict expression in all cells and generate a target pattern consisting of the predicted stripe 5

pattern and the endogenous pattern for stripes 2 and 7. When the *eve5* model is evaluated on this target pattern, it makes perfect predictions for stripe 5 and predicts false negatives in stripes 2 and 7 (Figure 4.7A). We then optimized the parameter associated with each regulator in the *eve5* model by varying it between -2500 and 2500, predicting the effect on expression, and evaluating the prediction on the target pattern using the MCC. We found that optimizing the Hb parameter or the *cad* parameter improved model performance: increasing the Hb parameter from -25.7 to 20 predicted expression in all three stripes, predicted expression in all three stripes, while increasing the *cad* parameter from -27 to -8 predicted expression in stripe 7 (Figure 4.7B). This sensitivity analysis suggests that either increasing *cad* activation or converting Hb from a repressor to an activator in *eve5* would be sufficient to generate expression in other stripes.

Guided by the modeling, we decided to predict the effect of removing Hb repression in *eve5*. Hb is an intrinsic repressor that can be counter-repressed by Cad, but the sequence features governing this interaction are unknown (see Chapter 3). Moreover, there is little experimental evidence that Hb can function as a strong activator in *eve* stripe 2 regulatory sequence (Supplemental Figure 3.2, Small et al. 1992; Arnosti et al. 1996). We simulated removal of Hb repression by setting the Hb parameter in the *eve5* model to zero and predicting expression in all cells (*eve5 mut Hb* model). Our model predicted that removing Hb repression should generate low level expression in stripe 7 (Figure 4.7C). We then performed a second sensitivity analysis using the *eve5 mut Hb* model as a starting point. As before, optimizing the *cad* parameter increased predicted expression in stripe 7 (Figure 4.7D). However, in this case, increasing the Bcd parameter from 147.4 to 224 or increasing the parameter for Bcd^2 from -878.9 to -485 predicted expression in all three stripes (Figure 4.7D).

Overall, our modeling approach made three testable predictions. First, increasing *cad* activation in *eve5* should generate stripe 7. Second, removing Hb repression in *eve5* should generate low level expression in stripe 7. Finally, removing Hb repression and adding Bcd activation in *eve5* should generate expression in stripes 2, 5 and 7. We tested these predictions

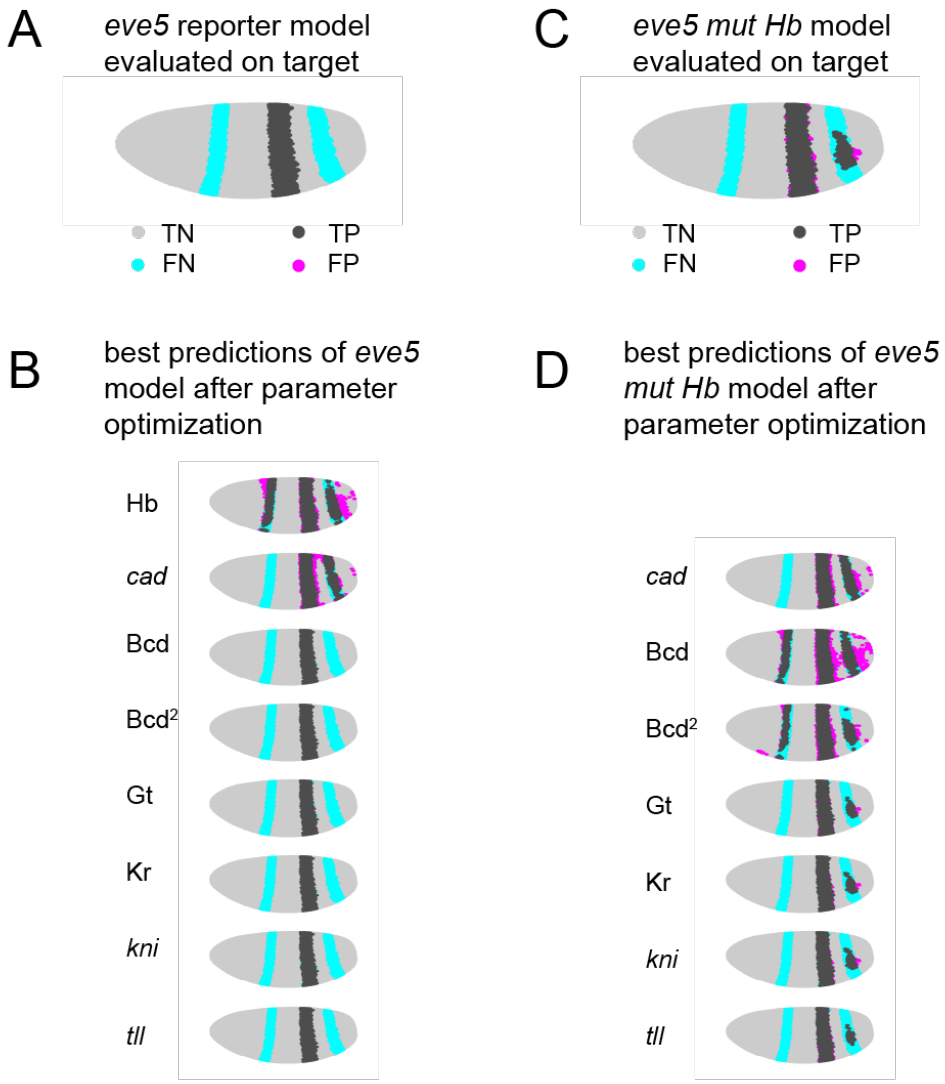


Figure 4.7: Sensitivity analyses predict how to generate multiple stripes from *eve5*. (A) We fit a model using thresholded data collected from an *eve5 LacZ* reporter construct. We then used this model to predict expression in all cells and generate a target pattern consisting of predicted ON cells in stripes 2, 5 and 7. The *eve5* model predicts false negatives in stripes 2 and 7 when evaluated on this target pattern. (B) We varied the parameter associated with each regulator between values of -2500 and 2500 and evaluated model performance by calculating the MCC using the target pattern as a gold standard. Optimizing the Hb parameter improved performance most, followed by optimizing the *caudal* (*cad*) parameter. (C) We simulated the effect of mutating Hb binding sites by setting the Hb parameter to 0, which predicted low level expression in stripe 7. (D) We varied the parameter associated with all other regulators after setting the Hb parameter to 0. Optimizing the linear and quadratic Bicoid terms (Bcd and Bcd^2) as well as the *cad* parameter improved model performance when evaluated on the target pattern.

by engineering the *eve5* enhancer (Figure 4.8A) and measuring gene expression in embryos staged between 9-35% membrane invagination (Luengo Hendriks et al. 2006). For each embryo, we measured the levels of stripe 2 and stripe 7 relative to stripe 5. Contrary to our modeling predictions, adding high affinity Cad sites did not generate expression in stripe 2 or stripe 7 (*eve5 add Cad*, Figure 4.8B). However, mutating predicted Hb sites in eve5 generated low level expression in both stripe 2 and stripe 7 (*eve5 mut Hb*, Figure 4.8C). We confirmed that we mutated all important Hb sites in *eve5* by expressing *eve5* and *eve5 mut Hb* in embryos misexpressing ventral *hb* from the *snail* promoter (*sna::hb* embryos; Clyde et al. 2003). The *eve5* reporter pattern retreats from ventral Hb in *sna::hb* embryos, while the *eve5 mut Hb* construct does not (Supplemental Figure 4.8). When we combined mutation in Hb and Cad sites (*eve5 mut Hb add Cad*), expression in stripe 2 increased further compared to the Hb mutations alone (Figure 4.8D, p-value < 0.005, Mann-Whitney U test). While stripe 7 levels driven by this double mutant construct were not significantly different from wild-type in embryos at this stage (Figure 4.8D), they were significantly different from stripe 7 levels driven by *eve5* and *eve5 mut Hb* at later time points (Supplemental Figure 4.9). We therefore concluded that the Cad sites we added increased expression in stripes 2 and 7, but only when Hb sites were mutated in *eve5*.

Finally, addition of Bcd sites alone (*eve5 add Bcd)* generated low level expression in stripe 2 (Figure 4.8E) and higher levels of stripe 2 expression when paired with mutations in Hb sites (*eve5 mut Hb add Bcd*, Figure 4.8F). Combining all 3 perturbations in one construct (*eve5 mut Hb add Cad add Bcd*) generated high levels of stripe 2 expression, as well as ectopic expression in the anterior of the embryo (Figure 4.8G). These results demonstrated that stripes 2, 5 and 7 can be generated together by engineering *eve5*, although there are discrepancies between our modeling predictions and our experimental results.

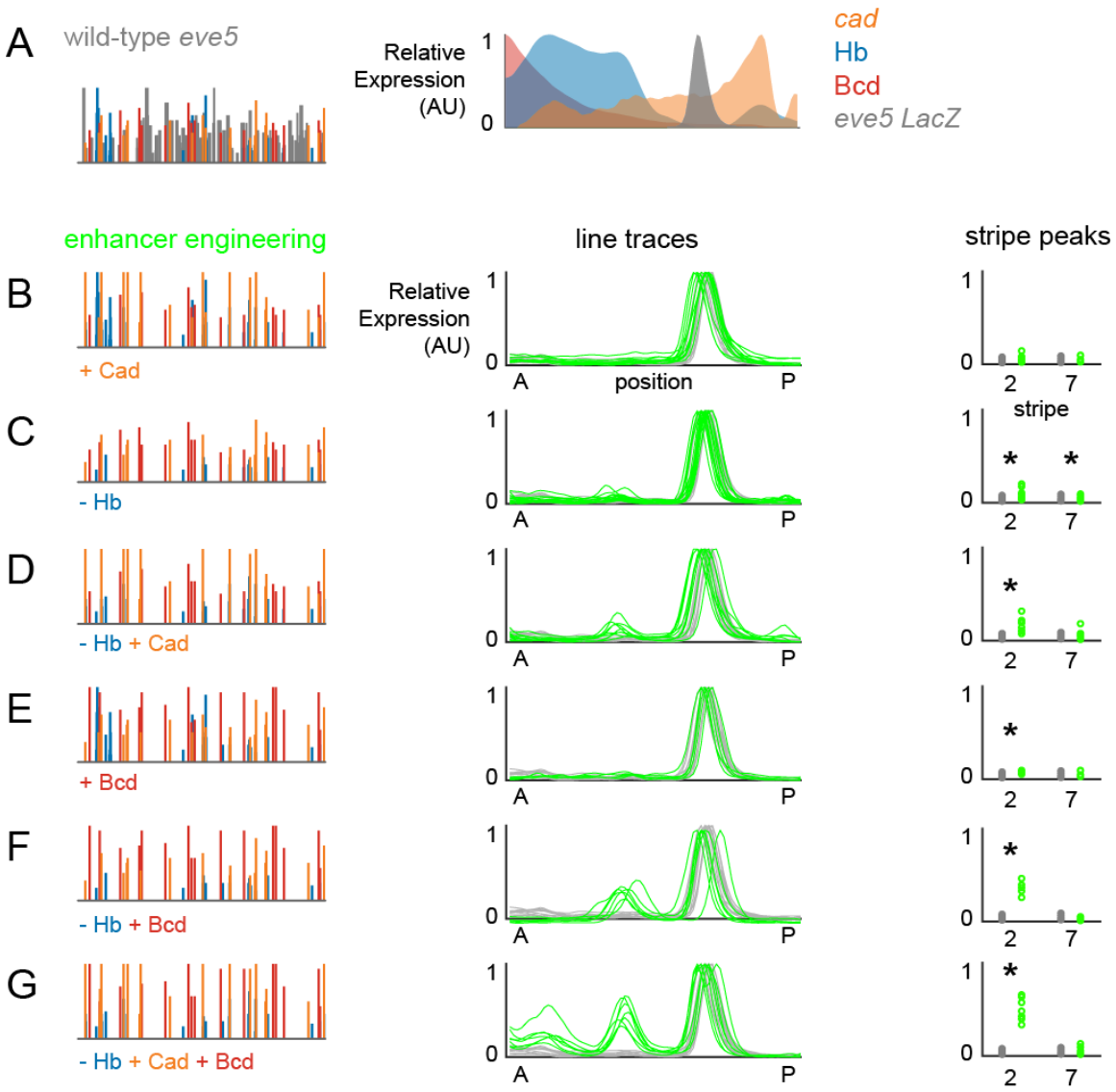


Figure 4.8: Engineering *eve5* generates stripe 2 and 7. (A) Predicted binding sites in eve5 for known regulators (grey), Bcd (red), Cad (orange) and Hb (blue). (B) Left: Predicted binding sites in *eve5 add Cad*. Middle: Line traces from individual embryos containing reporter constructs for wild-type *eve5* (grey) or *eve5 add Cad* (green). Embryos were staged between 8-35% membrane invagination in stage 5. Line traces were background subtracted and normalized to the peak expression level of stripe 5. Right: Maximum expression level in the area of stripe 2 and stripe 7 were extracted from individual line traces and plotted. Asterisks indicate significant differences from wild-type (p-value < 0.005, Mann-Whitney rank sum test). (C through G) Same data representation as (B) but for *eve5 mut Hb*, *eve5 mut Hb add Cad*, *eve5 add Bcd*, *eve5 mut Hb add Bcd*, and *eve5 mut Hb add Bcd add Cad* respectively.

Discussion

In this work, we found that the *even-skipped* gene expression pattern can be generated with multiple sets of enhancers that encode the stripes differently. We used computational modeling to demonstrate that the upstream *eve* regulatory network can, in principle, direct expression of multiple *eve* stripe combinations, some of which are not observed in the set of endogenous enhancers in *Drosophila melanogaster*. We predicted how to generate these alternative patterns using computational modeling, and we validated alternative solutions *in vivo* using sequence engineering. We used these results to inform a bioinformatic search for enhancers with diverged function. While follow-up experiments from that search are still in progress, we detected changes in relative stripe levels between orthologous enhancers. Here, we discuss the implications of our results for *eve* enhancer function and evolution.

Many solutions for a complex pattern

Our results suggest that there may be multiple ways to partition expression of the seven *eve* stripes among a set of enhancers. Our computational models suggest that the *eve* regulators carry sufficient positional information to specify nonendogenous stripe patterns, and our experiments demonstrate that some of these patterns can be generated by modifying endogenous regulatory DNA.

It has been shown conclusively that compensatory evolution can occur within modular enhancers (Ludwig et al. 2000; Ludwig et al. 2005) and between enhancers that drive overlapping expression patterns (Wunderlich et al. 2015). This work suggests that compensatory evolution may also occur between modular enhancers in the same locus. However, it remains unclear whether these alternative evolutionary solutions are "accessible," i.e. whether intermediate regulatory states confer no fitness cost. We could make progress on this question in two ways. First, we could test alternative patterns in the endogenous *eve* locus by replacing endogenous enhancers with engineered sequences. While advances in genome editing make this type of experiment possible (Bassett et al. 2013; Gratz et al. 2014; Ren et al. 2014), fitness costs

may only be revealed under genetic or environmental perturbation (Frankel et al. 2010; Ludwig et al. 2011). Second, we could locate enhancers with diverged function in orthologous sequences. This goal may require sequencing additional insect species, and we note that regulatory sequence alignment becomes more difficult with increased evolutionary distance (Hare et al. 2008). While our simple binding site overrepresentation analysis revealed a promising candidate in the *eve4+6* enhancer from *S. cynipsea*, the enhancer fragment we assayed failed to drive expression *in vivo*. Progress on this front may require improved models that predict expression level from sequence, many of which have been previously applied to Drosophila enhancers (Janssens et al. 2006; Kim et al. 2013; Martinez et al. 2014; Martinez et al. 2013; He et al. 2010; Kazemian et al. 2010; Samee & Sinha 2013; Duque & Sinha 2015; Samee & Sinha 2014; Duque et al. 2014). We anticipate that the quantitative data collected in this study will be useful as training sets to improve these types of models.

If developmental gene expression patterns can be generated in multiple ways, this conclusion would impact how we interpret functional genomic data in different species. Many groups have observed differences in TF binding or accessibility in orthologous enhancers (Paris et al. 2013; Arnold et al. 2014; Villar et al. 2015; Villar et al. 2014). While these signatures may imply functional divergence, they may also be compensated for by other changes in the locus. Functional genomic data is especially difficult to interpret in cell culture because the assay is performed in a single *trans* landscape. Candidate enhancers should be assayed in intact animals, and Drosophila embryos are an excellent system for further studies on this question. New experimental reagents also allow for reciprocal measurements of transgenic reporter constructs in non-*melanogaster* species, which will prove invaluable by controlling for changes in the patterning network between species (Stern et al. 2017).

Finally, we note that the logical intermediates between different evolutionary solutions are shadow enhancers – enhancers that drive overlapping patterns of expression (Hong et al. 2008; Barolo 2012). Shadow enhancers may confer adaptive properties, such as precision or robustness to genetic or environmental perturbation (Frankel et al. 2010; Perry et al. 2010;

Perry et al. 2011). However, shadow enhancers may also be byproducts of neutral evolution (Lynch 2007). Shadow enhancers for *eve* stripe 7 have already been identified in the locus (Staller et al. 2015), and the regulatory information for stripe 7 may drift between different enhancers over evolutionary time (Hare et al. 2008; Peterson et al. 2009). Further measurements may reveal additional *eve* shadow enhancers in other species, which would advance our understanding of their function in development.

Simple methods to engineer complex expression patterns

Developmental enhancers are complex - they contain many binding sites for regulators that vary in space and time, and these binding sites often exhibit non-additive interactions (Reiter et al. 2017). While many studies have focused on testing the necessity of individual binding sites or sequence fragments for enhancer function (Stanojevic et al. 1991; Small et al. 1992), attempts to build functional enhancers from their component binding sites are comparatively rare and often produce negative results (Vincent et al. 2016).

Here, we demonstrate the success of an intermediate approach: engineering new expression patterns by modifying existing enhancers. To this end, we developed and implemented a computational approach to predict if known regulators carry enough positional information to specify a given pattern. Though simple, this method requires quantitative measurements of all regulating TFs and putative target expression patterns at cellular resolution. While the Virtual Embryo dataset is uniquely suited for this purpose, we anticipate that similar datasets will be generated for other animal systems as imaging and sequencing technologies mature (Wotton et al. 2015; Combs and Eisen 2013; Klein et al. 2015; Rotem et al. 2015; Lee et al. 2015).

We predicted how to generate target patterns from existing sequences without the use of a sequence-based model. We were successful in this regard for two reasons. First, our target patterns could be generated by modulating the influence of one or two regulators on existing enhancers. Second, adding binding sites in locations that did not interfere with predicted sites

for other regulators was sufficient to increase the influence of that regulator. This latter finding suggests that 1) existing methods for predicting binding sites using PWMs are sufficient for the purpose of enhancer engineering (Noyes et al. 2008; Zhu et al. 2011; Gallo et al. 2011), and 2) in agreement with the 'billboard' model of enhancer activity, regulator efficacy may not depend on the exact position of binding sites within *eve* enhancers (Arnosti and Kulkarni 2005). Whether or not these features are generalizable to other regulatory networks remains unclear. We anticipate that enhancers exhibiting high cooperativity or competition between binding sites will require sequence-based models to engineer effectively (Carey 1998; Estrada et al. 2016). Overall, we were surprised by the relative success of our approach given the simplicity of our methods and the complexity of developmental enhancers.

A synthetic approach reveals what we don't understand

At its best, synthetic biology reveals unknown features of a system by testing whether existing principles are sufficient to build something new. In this regard, the predictions generated by our computational models were useful even when they were wrong. For example, *eve5* is thought to be activated by the spatially uniform TFs Stat92E and Zelda and repressed by Hb, Gt and Kr (Fujioka et al. 1999). Our modeling framework formalized this hypothesis and made testable predictions – removal of all Hb binding sites should result in some expression near stripe 2 and robust expression near stripe 7. The discrepancy between model and experiment points to a gap in our understanding of *eve5* regulation and generates new hypotheses. *eve5* may be sensitive to an unknown repressor expressed in the area of stripe 7, or Stat92E and/or Zelda may require a coactivator that is not expressed in this region. The use of models to suggest and contextualize experiments is especially important when studying complex systems with many interacting components, as the behavior of such systems is difficult to intuit (Wunderlich and DePace 2011).

In addition to revealing our incomplete understanding of *eve5*, these experiments indirectly test putative mechanisms of Hb bifunctionality. Hb activates expression when bound

to some enhancers and represses when bound to others; indeed, our recent work suggests that Hb both activates and represses *eve* stripe 7 through distinct enhancers (Staller et al. 2015). However, the sequence features that control Hb activity remain unclear. Synergy between Bcd and Hb has been observed (Simpson-Brose et al. 1994), and computational work suggests that the presence of nearby Cad sites can convert Hb from a repressor into an activator (Kim et al. 2013; see Chapter 3). If these hypotheses are true, we expected that adding high affinity Bcd or Cad sites to *eve5* would be sufficient to convert Hb into an activator, thereby producing expression in the regions of stripe 2 and stripe 7. However, our data show that this is not the case. Though not a stringent test of this mechanism, our results suggest that Bcd/Hb and Cad/Hb synergy either does not occur or requires precise spacing between binding sites, as is the case for other factors (Latchman 2010). We anticipate that the quantitative data generated by these experiments will provide an useful training set for sequence-based models to investigate the sequence features controlling Hb bifunctionality.

Conclusion

We find that the *eve* enhancers may exhibit plasticity due to the existence of alternative ways to partition their seven-stripe expression pattern. We also find that complex developmental enhancers can be rationally designed using simple computational and experimental approaches to produce new gene expression patterns. While these methods requires substantial knowledge of existing enhancers and the regulatory network that controls them, emerging technologies for measuring and perturbing gene expression in single cells may allow for widespread future application to other complex systems. Our results provide a useful case study for interpreting sequence divergence in orthologous enhancers – divergence in one enhancer may be compensated for elsewhere in the locus.

Materials and Methods

Computational modeling

Model training and fitting were performed as described (Ilsley et al. 2013). Briefly, we used atlas data from time point 3 of the Virtual Embryo. For our model inputs, we used protein data for Bcd, Hb, Gt, and Kr and mRNA data for *kni*, *cad* and *tll*; there is not currently protein data for Kni, Cad and Tll in the Virtual embryo. We also used squared values of Bcd protein data as another input (Bcd^2). To generate target patterns, we thresholded *eve* mRNA data from time point 3 using a fixed value of 0.2 and we extracted coordinates of individual stripes using the label function in the PointCloud Toolbox (http://bdtnp.lbl.gov/Fly-Net/bioimaging.jsp?w=analysis). Logistic regression was performed on all combinations of *eve* stripes with the common set of input regulators using the glmfit command in MATLAB. We then used these fit models to predict expression in all cells using the glmval command, and we thresholded these predictions at the same value of 0.2. We evaluated our predictions using the Matthews correlation coefficient (MCC, Matthews 1975) or the area under the receiver operating characteristic curve (Swets 1988). We also manually inspected our results to determine if our models predicted expression in separate stripes.

To predict how to generate alternative patterns from endogenous enhancers, we fit a model on atlas data of *LacZ* mRNA in embryos containing enhancer reporter constructs (reporter data). As above, we used a fixed threshold of 0.2 and data from time point 3. We fit models using the same input regulators to reporter data, and then used those fit model to predict expression in all cells. We thresholded the model predictions at the same value of 0.2 to generate the target patterns for our sensitivity analyses. For the stripe 6 target, we selected the stripe 6 cells predicted ON by the *eve4+6* reporter model using the label function. For the target including stripes 2, 5 and 7, we combined the cells predicted ON by the *eve5* model with thresholded data for endogenous stripes 2 and 7. To perform sensitivity analyses on the reporter models, we varied each parameter individually between values of -2500 and 2500, predicted expression for each value, evaluated predictions using the MCC, and selected the parameter

value that produced the highest MCC. We plotted predictions for each parameter after optimization. For *eve5*, we performed a second sensitivity analysis in the same way after setting the Hb parameter to 0.

Construct design

Similar to Chapter 3, we used PATSER (stormo.wustl.edu) to predict binding sites for known regulators of *eve4+6* and *eve5* before and after other binding site manipulations to minimize unintended effects. For *eve4+6*, we used position weight matrices (PWMs) generated by bacterial 1-hybrid experiments for D, Stat92E, Gt, Hb, Kni, Kr, Tll and Zelda (Noyes et al. 2008, Schroeder et al. 2011). We used a pseudocount of 1 to generate frequency matrices from count matrices, and a p-value cutoff of 0.01 to threshold binding sites predictions. To create weak Kr sites, we introduced mutations to generate sequences that matched annotated binding sites in *eve2* (Gallo et al. 2011). To create strong Kr sites, we introduced mutations to generate sequences that matched the consensus motif of the Kr PWM. For *eve5*, we used PWMs for Bcd, Cad, D, Stat92E, Gt, Hb, Kni, Kr, Nub, Tll and Zelda. We used a pseudocount of 1 and a p-value cutoff of 0.01. For *eve5 mut Hb*, we introduced mutations that disrupted predicted Hb binding sites with the highest PATSER scores. For other *eve5* constructs, we introduced mutations that generated sequences that matched the consensus motif of the Bcd and Cad PWMs.

Fly work

All reporter constructs were cloned into the pBΦY reporter plasmid (Hare et al. 2008) and integrated into the attP2 landing site (Groth et al. 2004), with the exception of *eve4+6 S. cyn*, which was cloned into a modified version of pBΦY containing an MS2 cassette (see Chapter 5) and a designed spacer sequence devoid of binding sites for *eve* regulators (Estrada et al. 2016). All transformants were homozygosed, and embryos were fixed, stained and imaged as described in Chapter 2. *sna::hb* embryos were generated as described in Chapter 2. We also generated fly lines containing the Kr[1] allele balanced with CyO as well as homozygous

transgenes containing *eve4+6* or *eve4+6 construct 3*. The Kr[1] stock was obtained from the Bloomington Stock Center (line #3494).

Binding site enrichment analysis

Enrichment of Hb, Gt and Kr sites was calculated as described in Chapter 3. We used PWMs generated by bacterial 1-hybrid methods coupled with SOLEXA sequencing for Hb and Kr (Zhu et al 2011). For Gt, we used a PWM in the literature (Schroeder et al 2011). We used pseudocounts of 0.01 for all PWMs to generate frequency matrices from count matrices. We calculated enrichment for all 500bp windows of orthologous *eve4+6* enhancers by subtracting expected sites from observed sites in the sequence. We did not threshold the PATSER scores when predicting binding sites for this analysis.

References

Akam, M., 1989. Drosophila development: making stripes inelegantly. Nature, 341(6240), pp. 282–283.

Alberts, B. et al., 2014. Molecular Biology of the Cell, Sixth Edition, Garland Science.

Arnold, C.D. et al., 2014. Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. Nature genetics, 46(7), pp.685–692.

Arnosti, D.N. et al., 1996. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development, 122(1), pp.205–214.

Arnosti, D.N. & Kulkarni, M.M., 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? Journal of cellular biochemistry, 94(5), pp.890–898.

Barolo, S., 2012. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. BioEssays: news and reviews in molecular, cellular and developmental biology, 34(2), pp.135–141.

Barolo, S. & Posakony, J.W., 2002. Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. Genes & development, 16(10), pp.1167–1181.

Bassett, A.R. et al., 2013. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell reports, 4(1), pp.220–228.

Cannavò, E. et al., 2016. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. Current biology, 26(1), pp.38–51.

Carey, M., 1998. The enhanceosome and transcriptional synergy. Cell, 92(1), pp.5–8.

Combs, P.A. & Eisen, M.B., 2013. Sequencing mRNA from cryo-sliced Drosophila embryos to determine genome-wide spatial patterns of gene expression. PloS one, 8(8), p.e71820.

Crocker, J., Ilsley, G.R. & Stern, D.L., 2016. Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors. Nature genetics, 48(3), pp.292–298.

Duque, T. et al., 2014. Simulations of enhancer evolution provide mechanistic insights into gene regulation. Molecular biology and evolution, 31(1), pp.184–200.

Duque, T. & Sinha, S., 2015. What does it take to evolve an enhancer? A simulation-based study of factors influencing the emergence of combinatorial regulation. Genome biology and evolution, 7(6), pp.1415–1431.

Estrada, J. et al., 2016. Information Integration and Energy Expenditure in Gene Regulation. Cell, 166(1), pp.234–244.

Estrada, J. et al., 2016. SiteOut: An Online Tool to Design Binding Site-Free DNA Sequences. PloS one, 11(3), p.e0151740.

Farley, E.K. et al., 2016. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proceedings of the National Academy of Sciences of the United States of America, 113(23), pp.6508–6513.

Fowlkes, C.C. et al., 2008. A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell, 133(2), pp.364–374.

Frankel, N. et al., 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature, 466(7305), pp.490–493.

Frasch, M. et al., 1987. Characterization and localization of the even-skipped protein of Drosophila. The EMBO journal, 6(3), pp.749–759.

Fujioka, M. et al., 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(11), pp.2527–2538.

Fujioka, M. & Jaynes, J.B., 2012. Regulation of a duplicated locus: Drosophila sloppy paired is replete with functionally overlapping enhancers. Developmental biology, 362(2), pp. 309–319.

Gallo, S.M. et al., 2011. REDfly v3. 0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic acids research, 39(suppl 1), pp.D118–D123.

Glassford, W.J. et al., 2015. Co-option of an Ancestral Hox-Regulated Network Underlies a Recently Evolved Morphological Novelty. Developmental cell, 34(5), pp.520–531.

Goltsev, Y. et al., 2004. Different combinations of gap repressors for common stripes in Anopheles and Drosophila embryos. Developmental biology, 275(2), pp.435–446.

Gompel, N. et al., 2005. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature, 433(7025), pp.481–487.

Gratz, S.J. et al., 2014. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics, 196(4), pp.961–971.

Harding, K. et al., 1986. Cross-regulatory interactions among pair-rule genes in Drosophila. Science, 233(4767), pp.953–959.

Hare, E.E. et al., 2008. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS genetics, 4(6), p.e1000106.

He, X. et al., 2010. Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. PLoS computational biology, 6(9), p.e1000935.

Hong, J.-W., Hendrix, D.A. & Levine, M.S., 2008. Shadow enhancers as a source of evolutionary novelty. Science, 321(5894), pp.1314–1314.

Ilsley, G.R. et al., 2013. Cellular resolution models for even skipped regulation in the entire Drosophila embryo. eLife, 2, p.e00522.

Indjeian, V.B. et al., 2016. Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. Cell, 164(1-2), pp.45–56.

Janssens, H. et al., 2006. Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. Nature genetics, 38(10), pp.1159–1165.

Jiang, P. et al., 2015. Natural variation of the expression pattern of the segmentation gene even-skipped in melanogaster. Developmental biology, 405(1), pp.173–181.

Kazemian, M. et al., 2010. Quantitative analysis of the Drosophila segmentation regulatory network using pattern generating potentials. PLoS biology, 8(8). Available at: http://dx.doi.org/10.1371/journal.pbio.1000456.

Kim, A.-R. et al., 2013. Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic. PLoS genetics, 9(2), p.e1003243.

Kimura, M., 1985. The role of compensatory neutral mutations in molecular evolution. Journal of genetics, 64(1), p.7.

Klein, A.M. et al., 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell, 161(5), pp.1187–1201.

Latchman, D.S., 2010. Eukaryotic Transcription Factors 5 edition., Academic Press.

Lee, J.H. et al., 2015. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nature protocols, 10(3), pp.442–458.

Long, H.K., Prescott, S.L. & Wysocka, J., 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell, 167(5), pp.1170–1187.

Ludwig, M.Z. et al., 2011. Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. PLoS genetics, 7(11), p.e1002364.

Ludwig, M.Z. et al., 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature, 403(6769), pp.564–567.

Ludwig, M.Z. et al., 2005. Functional evolution of a cis-regulatory module. PLoS biology, 3(4), p.e93.

Ludwig, M.Z., Patel, N.H. & Kreitman, M., 1998. Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development, 125(5), pp.949–958.

Luengo Hendriks, C.L. et al., 2006. Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. Genome biology, 7(12), p.R123.

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

Martinez, C. et al., 2014. Ancestral resurrection of the Drosophila S2E enhancer reveals accessible evolutionary paths through compensatory change. Molecular biology and evolution, 31(4), pp.903–916.

Martinez, C.A. et al., 2013. A synthetic biology approach to the development of transcriptional regulatory models and custom enhancer design. Methods, 62(1), pp.91–98.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et biophysica acta, 405(2), pp.442–451.

Maurano, M.T. et al., 2012. Systematic localization of common disease-associated variation in regulatory DNA. Science, 337(6099), pp.1190–1195.

Montavon, T. et al., 2011. A regulatory archipelago controls Hox genes transcription in digits. Cell, 147(5), pp.1132–1145.

Noyes, M.B. et al., 2008. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nucleic acids research, 36(8), pp.2547–2560.

Nüsslein-Volhard, C. & Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature, 287(5785), pp.795–801.

Paris, M. et al., 2013. Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression. PLoS genetics, 9(9), p.e1003748.

Perry, M.W. et al., 2010. Shadow enhancers foster robustness of Drosophila gastrulation. Current biology, 20(17), pp.1562–1567.

Perry, M.W., Boettiger, A.N. & Levine, M., 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 108(33), pp.13570–13575.

Peterson, B.K. et al., 2009. Big genomes facilitate the comparative identification of regulatory elements M. W. Hahn, ed. PloS one, 4(3), p.e4688.

Prud'homme, B. et al., 2011. Body plan innovation in treehoppers through the evolution of an extra wing-like appendage. Nature, 473(7345), pp.83–86.

Prud'homme, B. et al., 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature, 440(7087), pp.1050–1053.

Rebeiz, M., Patel, N.H. & Hinman, V.F., 2015. Unraveling the Tangled Skein: The Evolution of Transcriptional Regulatory Networks in Development. Annual review of genomics and human genetics, 16, pp.103–131.

Reed, R.D. et al., 2011. optix drives the repeated convergent evolution of butterfly wing pattern mimicry. Science, 333(6046), pp.1137–1141.

Reiter, F., Wienerroither, S. & Stark, A., 2017. Combinatorial function of transcription factors and cofactors. Current opinion in genetics & development, 43, pp.73–81.

Ren, X. et al., 2014. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. Cell reports, 9(3), pp.1151–1162.

Rotem, A. et al., 2015. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nature biotechnology, 33(11), pp.1165–1172.

Samee, A.H. & Sinha, S., 2013. Evaluating thermodynamic models of enhancer activity on cellular resolution gene expression data. Methods, 62(1), pp.79–90.

Samee, M.A.H. & Sinha, S., 2014. Quantitative modeling of a gene's expression from its intergenic sequence. PLoS computational biology, 10(3), p.e1003467.

Simpson-Brose, M., Treisman, J. & Desplan, C., 1994. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. Cell, 78(5), pp. 855–865.

Small, S. et al., 1991. Transcriptional regulation of a pair-rule stripe in Drosophila. Genes & development, 5(5), pp.827–839.

Small, S., Blair, A. & Levine, M., 1992. Regulation of even-skipped stripe 2 in the Drosophila embryo. The EMBO journal, 11(11), pp.4047–4057.

Small, S., Blair, A. & Levine, M., 1996. Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. Developmental biology, 175(2), pp.314–324.

Staller, M.V. et al., 2015. Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 112(3), pp.785–790.

Stanojevic, D., Small, S. & Levine, M., 1991. Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science, 254(5036), pp.1385–1387.

Stern, D. et al., 2017. Genetic and transgenic reagents for Drosophila simulans, D. mauritiana, D. yakuba, D. santomea and D. virilis. bioRxiv, p.096644. Available at: http://biorxiv.org/content/early/2017/02/20/096644 [Accessed March 7, 2017].

Strunk, B. et al., 2001. Role of CtBP in transcriptional repression by the Drosophila giant protein. Developmental biology, 239(2), pp.229–240.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science, 240(4857), pp. 1285-1293.

Thomas, S. et al., 2011. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. Genome biology, 12(5), p.R43.

Villar, D. et al., 2015. Enhancer Evolution across 20 Mammalian Species. Cell, 160(3), pp.554–566.

Villar, D., Flicek, P. & Odom, D.T., 2014. Evolution of transcription factor binding in metazoans – mechanisms and functional implications. Nature reviews. Genetics, 15(4), pp.221–233.

Vincent, B.J., Estrada, J. & DePace, A.H., 2016. The appeasement of Doug: a synthetic approach to enhancer biology. Integrative biology, 8(4), pp.475–484.

Wittkopp, P.J., Haerum, B.K. & Clark, A.G., 2008. Regulatory changes underlying expression differences within and between Drosophila species. Nature genetics, 40(3), pp.346–350.

Wotton, K.R. et al., 2015. Quantitative system drift compensates for altered maternal inputs to the gap gene network of the scuttle fly Megaselia abdita. eLife, 4. Available at: http://dx.doi.org/10.7554/eLife.04785.

Wunderlich, Z. et al., 2012. Dissecting sources of quantitative gene expression pattern divergence between Drosophila species. Molecular systems biology, 8, p.604.

Wunderlich, Z. et al., 2015. Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. Cell reports, 12(11), pp.1740–1747.

Wunderlich, Z. & DePace, A.H., 2011. Modeling transcriptional networks in Drosophila development at multiple scales. Current opinion in genetics & development, 21(6), pp. 711–718.

Zhu, L.J. et al., 2011. FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system. Nucleic acids research, 39(Database issue), pp.D111–7.

CHAPTER 5: DISCUSSION

Overview

In this dissertation, I described how we investigated the regulatory logic and evolution of *even-skipped* (*eve*) enhancers by challenging computational models with quantitative gene expression measurements in Drosophila embryos. In Chapter 2, we found that the pattern driven by *eve3+7* behaves differently under perturbation compared to endogenous stripes. This discrepancy was explained by *eve2+7*, a shadow enhancer that generates the stripe 7 pattern using different regulatory logic. In Chapter 3, we found that Hunchback (Hb) bifunctionality in these shadow enhancers was due to direct repression in *eve3+7*, and Caudal (Cad) counter-repression in *eve2+7*. Counter-repression of Hb by Cad is a conserved feature of *eve* stripe 2 regulation, and counter-repression in general may be more common in developmental systems than currently appreciated. In Chapter 4, we used computational modeling to examine the regulatory logic of the *eve* locus as a whole. We predicted alternative solutions for the *eve* expression pattern, and we validated our predictions by engineering endogenous enhancers and screening orthologous sequences for diverged enhancer function. Here, I describe how we can make progress on questions raised by this work using the vast experimental and computational resources available for Drosophila embryos.

Can we reconstitute a developmental enhancer?

This thesis begins with a set of negative results: we could not reconstitute *eve2* from its annotated binding sites in RedFly (Gallo et al. 2011). There are two explanations for this failure: 1) we do not know all of the regulators of *eve2*; or 2) Redfly annotations do not include flanking sequences that are critical for binding site function. Based on the results in Chapter 3, Cad is an excellent candidate for a missing regulator – Cad binding sites are necessary to counter-repress Hb sites in *eve2+7*. It is tempting to suggest that repression by the single annotated Hb binding

site is responsible for our negative results, although the ectopic anterior expression patterns driven by reconstituted constructs suggest otherwise (Figure 1.2). We can test this hypothesis in new constructs by adding predicted Cad binding sites or mutating the annotated Hb site.

Flanking sequences around binding sites may also be important for their function in *eve2*. Unpublished results suggest that flanking sequences around Bcd sites are necessary to reconstitute the function of the *hunchback P2* enhancer (Jeehae Park, personal communication; Driever et al. 1989). Motivated by these findings, I included 7 bp flanks around annotated Bcd binding sites as well as the Slp1 site identified in Andrioli et al. 2002 in a new round of reconstituted *eve2* constructs. These construct should also include the Cad sites mutated in Chapter 3 experiments (Supplemental Figure 5.1). Unfortunately, these new and improved *eve2* constructs failed to drive expression in embryos (data not shown, see Enhancer Sequences in Appendix E). While these results are personally devastating, they are also scientifically useful. The most parsimonious explanation is that reconstituted *eve2* constructs are missing binding sites for one or more critical activators. Potential candidates include *Dichaete* (Nambu and Nambu 1996; Ma et al. 1998) and *Trithorax-like* (*GAGA factor*, Moshe and Kaplan 2017). We could investigate these candidates directly in additional constructs, or take a more unbiased approach by adding back blocks of endogenous sequence (Johnson et al. 2008).

While we could simply abandon our dreams of enhancer reconstitution, success would enable tantalizing follow-up experiments. For example, a detailed map of regulatory information in *eve2* would allow for careful sequence engineering. As discussed in Chapters 3 and 4, adding and removing binding sites within enhancers can have unintended consequences due to effects on neighboring or overlapping sites. By mapping 'neutral' areas for binding site addition, *eve2* reconstitution could provide a powerful *in vivo* platform for testing the function of individual regulators (discussed in more detail below).

Conversely, mapping the exact location of all *eve2* binding sites would enable direct tests of *cis*-regulatory grammar in the context of a native enhancer. While many labs are interested in the rules underlying TF interactions (Kim et al. 2009), tests of these rules are limited to small

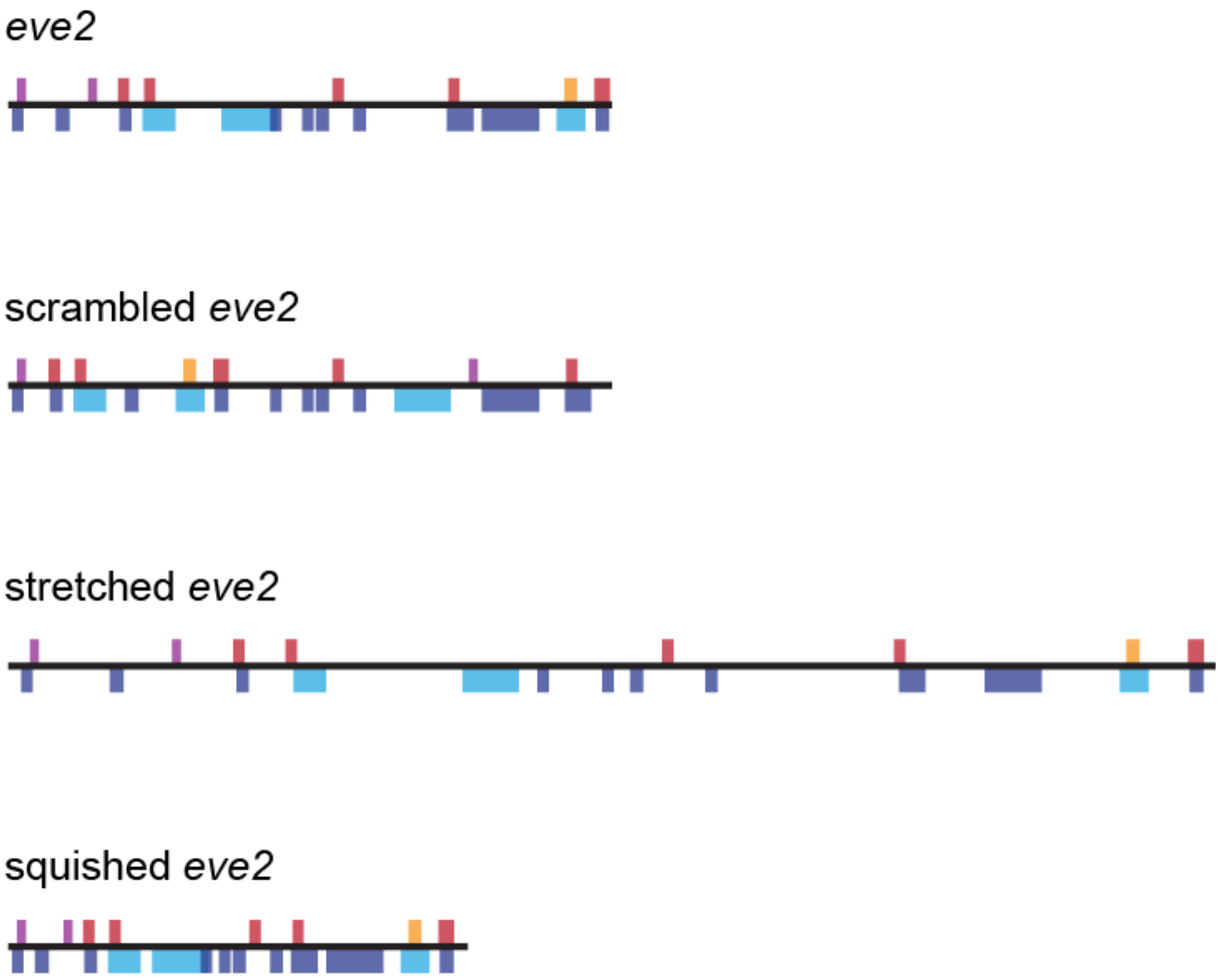


Figure 5.1: Reconstituting *eve2* would enable direct tests of *cis*-regulatory grammar. Cartoons of potentially interesting follow-up constructs if *eve2* reconstitution succeeds in the future. Binding sites displayed as in Figure 1.2. Red: Bicoid; tangerine: Hunchback; violet: Zelda; light blue: Giant; dark blue: Krüppel.

arrays of binding sites (Simpson-Brose et al. 1994; Fakhouri et al. 2010; Liberman and Stathopoulos 2009; Erceg et al. 2014) or fusions and deletions of entire enhancers (Gray et al. 1994; Lydiard-Martin et al. 2014). Once we know all the binding sites necessary for *eve2* function, we can rigorously test the role of binding site order, number, spacing and affinity in generating the stripe 2 pattern (see Figure 5.1 for examples). If we can couple controlled sequence perturbations with computational models of enhancer function, we may make some serious headway on defining the sequence-to-function relationship for developmental

enhancers. Indeed, Jeehae Park and Javier Estrada are currently pioneering this type of approach for the *hunchback P2* enhancer (*hbP2*). Though much simpler than *eve2*, *hbP2* exhibits 'higher-order' cooperativities between binding sites that cannot be explained by pairwise protein-protein interactions (Estrada et al. 2016; Gregor et al. 2007). Now that they have a reconstituted *hbP2* sequence in hand, Jeehae and Javi are better equipped to dissect the molecular basis of higher-order cooperativity and identify additional sequence features that contribute to *hbP2* function. Doug, I believe, would be proud of their efforts.

How do different activators and repressors work?

This is the $64,000 question, and I wish I had infinite time and youth to address it. While the Drosophila enhancer field seems quite content to classify TFs as activating, repressing or bifunctional, TFs in other systems have been characterized in much greater biochemical detail (Fuda et al. 2009). TFs may act on one or more distinct steps in the eukaryotic transcription cycle (Green 2005; Adelman and Lis 2012; Beagrie and Pombo 2016; Harlen and Churchman 2017), and their precise biochemical roles may influence expression output (Scholes et al. 2016). Interactions between TFs and the transcriptional machinery may require cofactors that contain many components, and these complexes may not be identical across cell types (Mannervik 2014). To understand how enhancers function and evolve, we need to characterize the biochemical functions of their regulators and identify their cofactors. The reagents described in this dissertation can help.

My results demonstrate that we can rationally design enhancers that contain added or mutated binding sites for individual TFs. While we have characterized these enhancers in fixed embryos, we can now take advantage of recently developed methods for imaging nascent transcription in live embryos (Garcia et al. 2013; Bothma et al. 2014; Bothma et al. 2015). Specifically, we can measure the influence of individual TFs on particular dynamic parameters, which may help us infer their biochemical function (Figure 5.2). I have already designed versions of *eve2* that contain additional binding sites for different activators (see Enhancer

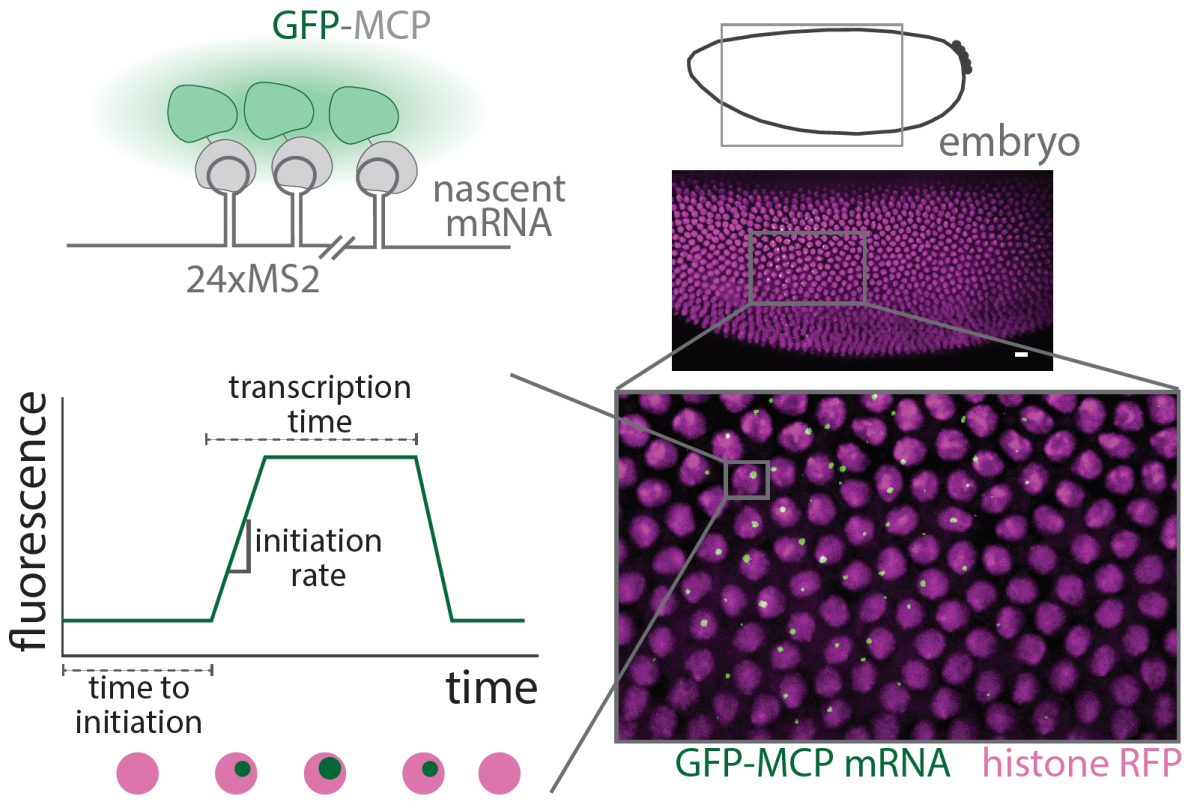


Figure 5.2: The MS2 system enables live imaging of nascent transcription in embryos. In live embryos, we observe dynamic nascent transcription using the MS2 reporter system and a simple model of transcription (Garcia et al 2013). The two-color maximum-projection image is of an embryo during the 14th nuclear cycle of development. The projection is from 210 z-stack images, separated by 0.5 µm, acquired by light sheet-based fluorescence microscopy; scale bar 10 µm. The large field of view allows simultaneous imaging of many more nuclei than previous methods employing confocal microscopy. Here, we observe transcriptional output driven by the *eve2* enhancer an an MS2-*LacZ* reporter construct. In addition to nascent transcript levels within individual nuclei, we also can measure the number and fraction of active nuclei as well as their spatiotemporal location over the time course of embryo development. Figure by Tim Harden.

Sequences in Appendix E). In pilot experiments with my collaborator Tim Harden and the Nikon Imaging Center at Harvard Medical School, we found that adding Zelda binding sites in *eve2* was sufficient to activate gene expression in mitotic cycles 11, 12 and 13 (Figure 5.3A). In addition, adding Zelda binding sites increased expression levels in individual cells (Figure 5.3B). This latter result contradicts recent work suggesting that Zelda binding influences the probability of activation, but not expression level (Crocker et al. 2017). As those experiments

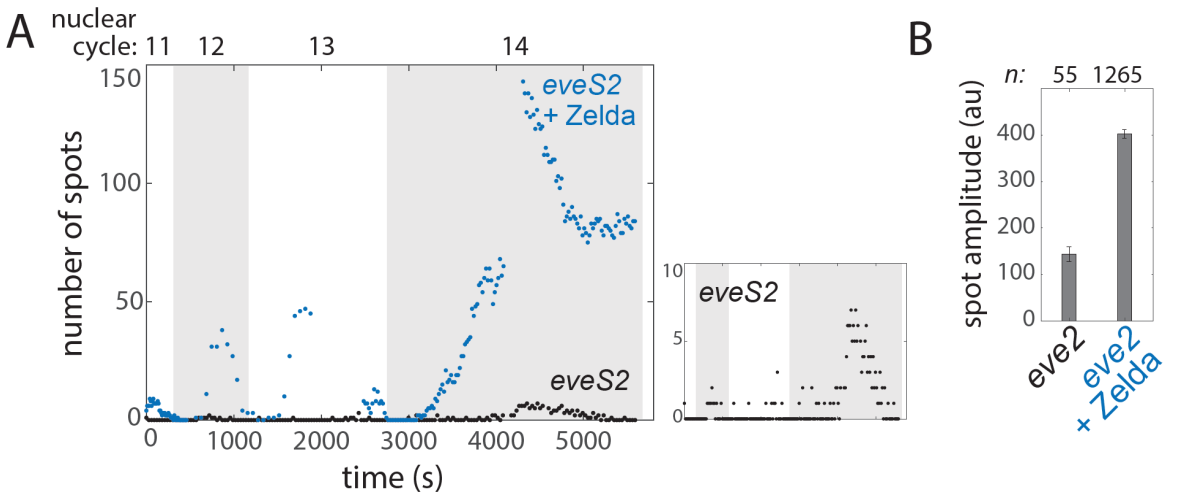


Figure 5.3: Zelda increases the number of active cells as well as the expression level in those cells. (A) Measurements of the number of active transcription sites during development across individual embryos containing reporter constructs for *eve2 MS2* (black) or *eve2 MS2 add Zelda* (blue, see Enhancer Sequences in Appendix E). Discontinuities in the data are interruptions in image acquisition due to optical re-alignment. Inset shows the endogenous enhancer activity on a scale emphasizing that endogenous expression during nuclear cycle 14 is robust. Times between nuclear divisions are indicated by alternating gray and white intervals with corresponding nuclear cycles listed above. The time between nuclear divisions is as expected for healthy flies; these imaging conditions do not detectably hinder development through zygotic genome activation during nuclear cycle 14. These data are consistent with recently published data using a synthetic transcriptional platform to measure the effect of Zelda on transcription (Crocker et al 2017). (B) Spot amplitude from a subset of the data in (A) during nuclear cycle 14. Spot amplitude is a proxy for the amount of expression at each transcription site. Contrary to the data in A, these data are inconsistent with a model in which Zelda activity is confined to chromatin remodeling. Although the data shown here are enlightening, they are preliminary. Once optimized, these experiments will provide us with additional information about the spatiotemporal characteristics of transcription (see Figure 5.2). Figure by Tim Harden.

were performed in a more synthetic context, the impact of Zelda binding on expression dynamics may depend on the identity and function of surrounding TFs.

This general approach is not limited to activators in *eve2*. We can also measure the influence of gap gene repressors on expression dynamics in *eve* enhancers. As with Krüppel and *eve4+6*, the key is to choose regulators that overlap particular *eve* stripes but do not define stripe boundaries. For example, *eve4+6 construct 1* is ready-made to study Krüppel (see Chapter

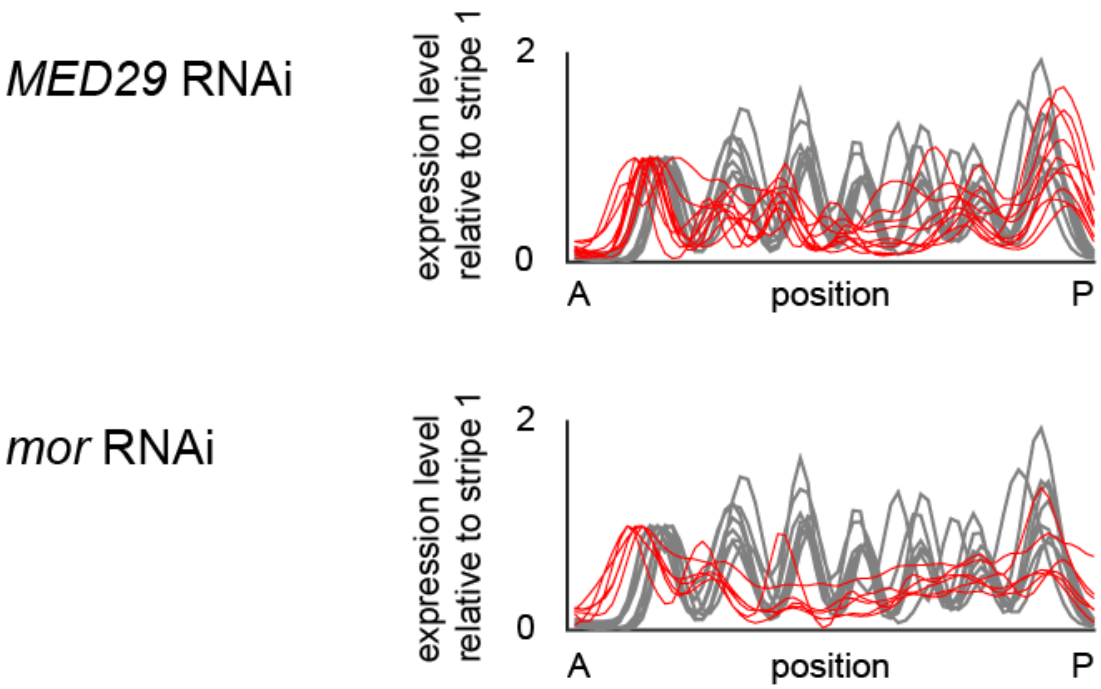


Figure 5.4: Depleting maternal cofactors affects *eve* expression. Line traces of *eve* mRNA levels in wild-type embryos (grey) and embryos depleted of maternal cofactors (red). Line traces were normalized to the peak of stripe 1. Depleting both cofactors appears to decrease levels of stripes 2-6. Depleting *Mediator complex subunit 29* (*MED29*) may cause a posterior shift of stripe 7, while depleting *moira* (*mor*) may cause an anterior shift of stripe 1. Note that 'wild-type' embryos were generated as part of the *sna::hb* cross that generates approximately equal numbers of wild-type and *sna::hb* embryos. They are not perfect controls, but, shockingly, they are the best 'wild-type' sample currently stained for *eve* mRNA in the DePace Lab Bio Imaging Meta Database.

4), while the effects of Hunchback, Giant and Knirps could be examined in engineered versions of *eve2*, *eve4+6* and *eve5* respectively.

I have also designed genetic crosses to couple live imaging with RNAi, which would allow for similar studies of maternally-deposited cofactors (Supplemental Figure 5.2). In collaboration with Max Staller, Adam Carte, Evi Van Italie, Kelly Biette and Olivia Foster, we have already detected effects of maternal cofactor depletion on *eve* expression (Figure 5.4). In these experiments, stripe-specific effects are especially notable because they can suggest interactions between enhancer-specific regulators and cofactors. For example, *MED29* and *mor* depletion

appears to affect stripes 2-6 more than stripe 1 or stripe 7 (Figure 5.4). These interactions can be investigated further with genetics and biochemistry.

Of course, to understand the biochemical mechanisms of TF function, we could also make biochemical measurements. While functional genomic experiments have been performed on Drosophila embryos (Paris et al. 2013; Lott et al. 2011; Chen et al. 2013; Ghavi-Helm et al. 2014; Blythe and Wieschaus 2016), they often require homogenization, which can dilute signal for enhancers expressed in particular regions of the embryo. However, we've been encouraged by a recent study, which used ChIP-qPCR to characterize the distinct biochemical mechanisms of *runt* repression in *slp* shadow enhancers (Hang and Gergen 2017). We could use an analogous approach on reporter constructs containing wild-type or engineered *eve* enhancers to measure the impact of particular transcription factors on preinitiation complex (PIC) assembly, polymerase recruitment and pause release. Coupling biochemical measurements with live imaging of engineered *eve* enhancers could make the Drosophila embryo a powerhouse for mechanistic studies of TF function that even Doug could envy.

What are the molecular mechanisms underlying counter-repression?

The main conclusion of Chapter 3 is the Cad counter-represses Hb in *eve2+7*. How does this interaction occur? We can take some cues from the literature on short-range repression in the Drosophila embryo, which has been characterized in terms of required cofactors (Nibu et al. 1998; Nibu and Levine 2001; Nibu et al. 2003; Payankaulam and Arnosti 2009) and induced chromatin state (Li and Arnosti 2011). By coupling the genetic reagents described in this dissertation with biochemical measurements, we can dissect the mechanisms underlying counter-repression.

The simplest mechanistic hypothesis is that Cad binding prevents Hb binding in *eve2+7*. We can test this hypothesis by performing ChIP-qPCR on embryos containing *eve2+7* and *eve2+7 mut Cad*. If our hypothesis is correct, we should detect increased Hb binding in *eve2+7 mut Cad* compared to *eve2+7*. As discussed above, this experiment may be complicated by the

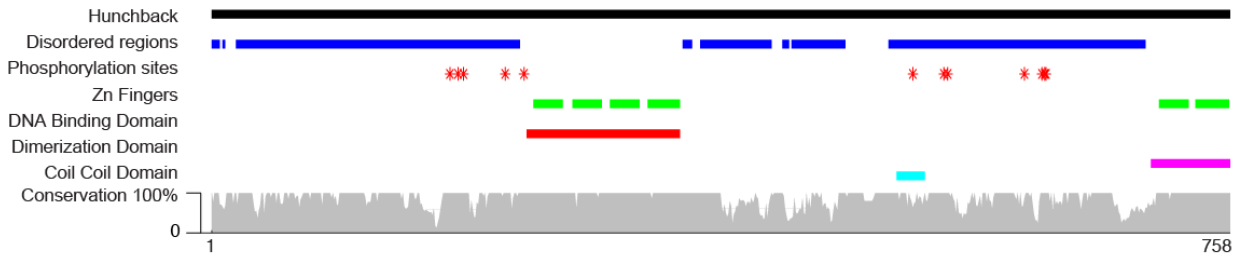


Figure 5.5: Annotated Hunchback protein features. Figure by Max Staller.

fact that both enhancers are active in a limited part of the embryo, and ChIP-qPCR requires homogenization. Development of region-specific ChIP methods could address this shortcoming (Bonn et al. 2012; Bowman et al. 2014).

Because Hb appears to bind to the *eve2+7* region *in vivo* (Paris et al. 2013), we need to consider the alternative hypothesis that Cad binding interferes with the repressive function of Hb but not its binding. To address this hypothesis rigorously, we need to define the mechanism of Hb repression in *eve2+7*. We have in hand an *eve2+7* construct with added Hb sites that drives no stripe 2 expression and retreats from ventral Hb (*eve2+7 add Hb*; see Supplemental Figure 5.3). As described above, we could perform ChIP-qPCR on this construct to determine if additional Hb sites block PIC formation, polymerase recruitment or pause release and correlate these measurements with effects on expression dynamics using the MS2 system.

We can also use the *sna::hb* system to define the Hb protein domains that are responsible for repression. Hb contains many annotated protein features including the putative dimerization domain (McCarty et al. 2003), a coiled-coil domain, and phosphorylation sites mapped by mass spectrometry (Figure 5.5; Norbert Perrimon, personal communication). It also contains other conserved domains that are necessary for function in *Drosophila melanogaster* neurons (Tran et al. 2010). To assess the importance of these features, we can misexpress mutant forms of *hb* and measure expression driven by *eve2+7*, *eve2+7 mut Cad*, and *eve2+7 add Hb*. We have already generated fly lines containing transgenes in the same genomic location that allow us to misexpress wild-type *hb* (DePace stock #0616) and *hb* with a deleted

dimerization domain (DePace stock #0617). Preliminary results suggested that mutating the dimerization domain abolishes repression of *ftz* and *eve3+7* LacZ (0/18 embryos exhibited repression compared to 7/12 embryos that misexpressed wild-type *hb;* p-value = 4×10^{-6}). These results are consistent with findings in Drosophila neurons (Tran et al. 2010). However, as shown by the *slp* example, the same TF can employ different modes of repression in neighboring enhancers (Hang and Gergen 2017). We should therefore measure the response of *eve2+7*, *eve2+7 mut Cad*, and *eve2+7 add Hb* to misexpressed mutant forms of *hb*. While our goal is to understand the molecular mechanisms underlying counter-repression, we must begin by dissecting the mechanism of Hb repression before we can tackle its interactions with Cad.

Are there other bifunctional transcription factors in the gap gene network?

Out of all the findings presented in this dissertation, the one that shocks me most is Hb repression in *eve2+7*. While consistent with qualitative studies in the early *eve2* literature (Small et al. 1992; Arnosti et al. 1996), this example taught me that textbook cartoons can constrain your thinking. If Hb doesn't activate *eve* stripe 2, does it function as an activator anywhere in the embryo? Qualitative measurements from binding site arrays provide convincing evidence that it has the capacity to activate transcription (Simpson-Brose et al. 1994), but whether it does so within endogenous enhancers is a different question. We are currently testing whether Hb directly activates the proximal *Krüppel* shadow enhancer (see Figure 2.5). This sequence contains many predicted binding sites for Hb and Cad, so we used a computational tool to design enhancer sequences with mutated binding sites for one or both factors (Estrada et al. 2016). While we have transgenic animals containing reporter constructs for these designed sequences, we have not yet measured their expression patterns (see Enhancer Sequences in Appendix E). This experiment will determine whether Cad counter-repression also occurs in the Krüppel locus.

Recent computational studies have proposed that other gap genes act bifunctionally in the Drosophila embryo. Regression models suggest that *kni* bifunctionally regulates *eve4+6*

(Crocker et al. 2016), while thermodynamic models suggest that *giant* acts as an autoactivator in the *gt-3* enhancer (Hoermann et al. 2016). The latter study included *in vivo* mutagenesis of predicted Gt binding sites to support its conclusions. As we learned from the Hb example, computational models can appear correct for the wrong reasons, and careful perturbations in *cis* and *trans* are necessary to distinguish true bifunctionality from other regulatory interactions. We can test the *eve4+6* predictions by mutating predicted Kni binding sites in *eve4+6* and measuring WT and mutated constructs in a *sna::kni* background (Clyde et al. 2003; DePace stock #0037). Measuring expression of *eve4+6* regulators in *sna::kni* embryos will be critical for interpreting the results correctly, as simulations of genetic perturbations may not recapitulate behavior *in vivo* (compare predictions in Ilsley et al. 2013 with measurements in Staller et al. 2015). We can also test the predictions for *giant* bifunctionality by expressing reporter constructs containing WT and mutated forms of *gt-3* in *sog::gt* embryos (See Chapter 2). We now have the genetic tools to rigorously define the regulatory roles of the gap genes, which for years have been nebulous at best (Jaeger et al. 2012).

Can we investigate *cis*-regulatory rules in high throughput?

Perturbing enhancer sequence in reporter constructs is a powerful method to investigate regulatory logic in intact animals. However, this approach is limited by the time required to generate, maintain and assess individual transgenic lines – we can only measure output of a few sequences in a reasonable time frame. We can measure expression driven by thousands of enhancer variants in cell culture (Shlyueva et al. 2014), but as discussed in Chapter 1, these methods are limited by sequence length and the single *trans* environment of the cell line used. We would like to perturb enhancer sequence and measure the consequences in embryos without the annoying step of generating transgenic lines. How can we can we make this dream a reality?

I have worked in collaboration with many lab members (most recently Anna Cha) to develop CRISPR-interference (CRISPRi; (Larson et al. 2013; Qi et al. 2013; Gilbert et al. 2014) as a high throughput method for perturbing endogenous regulatory DNA. Our strategy is to

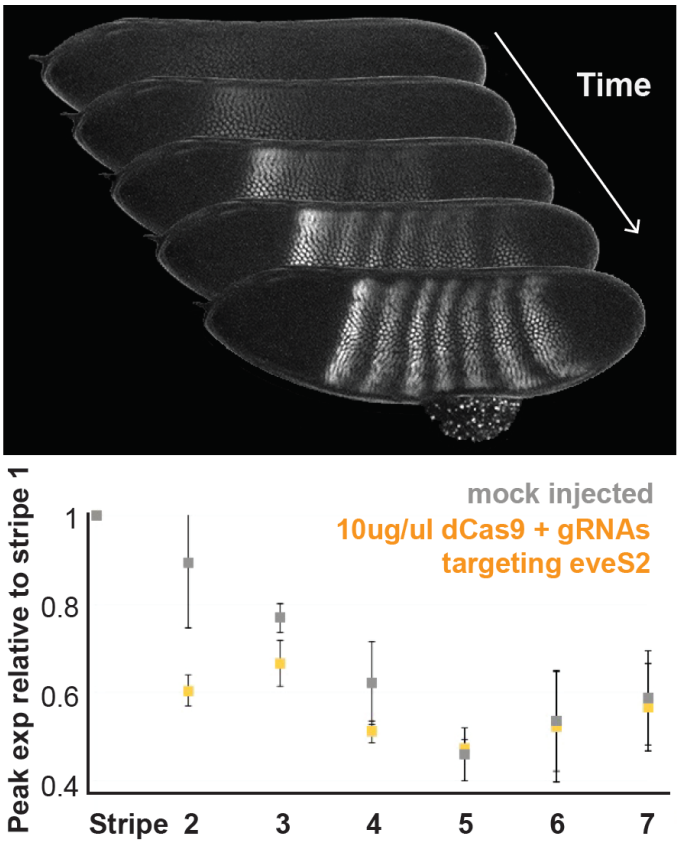


Figure 5.6: Preliminary data suggests that CRISPRi is feasible in Drosophila embryos. Top: We image YFP in live embryos using confocal microscopy. Maximum intensity projections of a single stage 5 embryo are shown here. Bottom: We used ImageJ to extract peak expression levels of all stripes relative to stripe 1. Embryos were staged by eye; n = 3 for mock injected embryos (grey) and n= 4 for embryos injected with dCas9/gRNAs (orange). Error bars indicate standard deviation. Figure by Anna Cha and Angela DePace.

inject guide RNAs (gRNAs) that target enzymatically dead Cas9 (dCas9) to regulatory DNA and measure the consequences by live imaging of a yellow fluorescent protein (YFP) reporter under the control of the *eve* regulatory sequences (Ludwig et al. 2011). I designed a construct that expressed dCas9 fused to the Kr repression domain (Crocker and Stern 2013) under the control of the *nanos* (*nos*) promoter (see Appendix E). I also designed gRNAs to target the regulatory sequences for *eve* stripe 2 (see Appendix E). Preliminary data suggests that co-injection of purified dCas9 with gRNAs targeting *eve* stripe 2 decreased stripe 2 expression relative to stripe 1 (Figure 5.6). However, in this experiment, embryos were not rigorously staged and controls including either component alone were not included.

We are currently improving the CRISPRi protocol. First, we need to optimize dCas9 levels in the embryo. Anna Cha has crossed a UAS-inducible dCas9 (a gift from Ben Ewen-Campen and Norbert Perrimon) with maternal *GAL4* lines and measured dCas9 protein levels

by western blot. Second, as an intermediate step, we are targeting gRNAs to the *eve* promoter and assessing morphology of the larval cuticle as compared to an *eve* null mutant (Nüsslein-Volhard and Wieschaus 1980). This method is faster and easier compared to live measurements of YFP levels in embryos. Finally, we are genetically encoding gRNAs for the *eve* promoter to establish a benchmark for our injection method (Ren et al. 2014). These experiments are part of a long-term investment in an injection-based approach that will allow us to rapidly explore combinatorial effects of inhibiting discontiguous regulatory sequences. Developing this type of approach will be critical for dissecting the complex region that regulates *eve* stripe 7 expression. Moreover, Cas9 can be fused to a wide variety of different effector proteins or fluorophores to manipulate and measure gene expression.

Does regulator separation in shadow enhancers confer robustness?

Shadow enhancers confer robustness of patterning to environmental perturbation (Frankel et al. 2010; Perry et al. 2010). The sequence features underlying this phenomenon are currently unknown. Our results demonstrate that shadow enhancers can generate the same pattern using different regulators: *eve* stripe 7 shadow enhancers use different repressors (Chapters 2 and 3), while *Kr* shadow enhancers use different activators (Wunderlich et al. 2015). These findings suggest a tempting hypothesis: by building expression patterns using different regulators, shadow enhancers confer robustness to perturbations in those regulators. We can test this hypothesis by measuring expression driven by shadow enhancer reporter constructs in genetically perturbed embryos. Specifically, we expect single shadow enhancers to be sensitive to perturbations in their regulators and insensitive to perturbations in other regulators. Furthermore, we expect that combined shadow enhancer pairs should be insensitive to perturbations in 'unshared' regulators.

As a first step, we should confirm direct TF influence on all *eve* and *Kr* shadow enhancers. We know that Kni directly binds *eve3+7* to define the anterior stripe 7 border (Struffi et al. 2011). Indirect evidence in Chapter 3 suggests that Gt binds *eve2+7* to define the anterior

stripe 7 border, but we should test this interaction directly. Olivia Foster has designed an *eve2+7* construct containing mutated Kni sites; this will also be a critical control to establish regulator separation in the *eve* stripe 7 shadow enhancers. We have not yet measured the effects of perturbing TF binding sites in *Kr* enhancers. In addition to these single enhancer controls, we need reporter constructs containing combined shadow enhancer pairs. We have a reporter construct containing *Kr* shadow enhancers separated by a neutral spacer (Wunderlich et al. 2015), as well as a reporter construct containing the entire *eve* regulatory region spanning *eve3+7* and *eve2* (*eve3+7 and eve2,* see Enhancer Sequences in Appendix E). To make rigorous comparisons, we need to generate a new reporter construct containing *eve3+7* and *eve2+7* separated by a neutral spacer.

With regard to specific genetic perturbations to incorporate into our experiments, we can deplete *Kr* activators using RNAi (Staller et al. 2013), and we can perturb *eve* repressors using heterozygous mutant lines (Frasch et al. 1987). We note that due to the extensive regulatory interactions in the embryo, perturbing individual regulators without affecting others may be impossible (Jaeger 2011). Quantitative comparisons between shadow enhancer reporter constructs will likely be necessary. Despite this complication, I'm excited about the prospect of investigating a specific hypothesis on how shadow enhancers confer robustness in developmental systems.

Can we detect divergence in enhancer function?

Coming full circle, this dissertation also ends with a set of negative results: despite some attractive candidates, we were unable to identify orthologous *eve* enhancers that drive different stripe patterns. We can explain these results in 2 ways: 1) there are no orthologous *eve* enhancers with diverged function; or 2) I didn't try hard enough in my screening approach. I favor the second hypothesis. We have shown that different *eve4+6* enhancer fragments drive different relative stripe levels (Supplemental Figure 4.6), as do orthologous versions of the same *eve4+6* enhancer fragment (Figure 4.5). Furthermore, an orthologous version of *eve3+7* from

Ceratitis capitata only drives expression of stripe 3 when assayed in *Drosophila melanogaster* embryos (Peterson et al. 2009). I believe that diverged enhancers are out there, but we lack the predictive methods to find them. I can think of three ways forward.

First, we can screen different fragments of the *eve4+6* candidate sequences we have in hand. The *eve4+6* fragment from *S. cynipsea* was assayed with spacer sequences on either side in an *MS2-LacZ* reporter construct. These spacer sequences were included to control for the effect of Kr binding sites on the *eve* promoter. However, we should assay the same fragment in our standard pBΦY reporter plasmid. We should also assay this fragment with additional flanking sequence on each end in case this sequence includes critical binding sites. Finally, we should assay fragments of *eve4+6* from *Dicranosepsis sp.*, since portions of that sequence are also enriched for both Hb and Gt binding sites. In terms of the *eve* enhancer screen in Chapter 4, while there was an attempt, that attempt could have been more thorough.

Second, we can apply the binding site enrichment analysis to other sets of orthologous enhancers. Based on our experimental results, depletion of Hb binding sites and enrichment of Bcd binding sites in *eve5* may generate stripes 2 and 5 together. While we can identify orthologous *eve5* enhancers in the sequenced Drosophila species using the LiftOver tool (http://genome.ucsc.edu/), identifying *eve5* sequences in Sepsid species may require additional sequencing or alignment methods (Schwartz et al. 2003). We can also speculate on other binding site signatures of diverged function based on our knowledge of *eve* enhancer logic. Published results in *kni* mutant embryos suggest that depletion of Kni binding sites in *eve2* enhancers may be sufficient to generate stripe 5 (Janssens et al. 2006). We also expect that enrichment of Kr binding sites in *eve3+7* enhancers should generate stripe 7 alone, since the *Kr* expression pattern overlaps stripe 3 but not stripe 7 (Supplemental Figure 2.7). We note that enrichment analysis should be performed with caution, as results appear to be sensitive to the threshold value for PATSER score (Supplemental Figure 3.5). Our general approach will also be improved by sequencing additional insect species, better methods for predicting and mapping

binding sites (Zhu et al. 2011; Rhee and Pugh 2012; He et al. 2015) and a complete understanding of *eve* regulatory logic.

Finally, we could use more sophisticated computational methods for predicting expression from sequence. Binding site enrichment is simple to understand and implement, but may not identify enhancer fragments that autonomously generate expression. There are computational methods available to pinpoint fragments that have the potential to drive expression (Kazemian et al. 2010) and others that explicitly predict expression patterns from enhancer sequence (He et al. 2010; Kim et al. 2013). More recent models combine these approaches to identify sequence fragments that drive particular gene expression patterns (Samee and Sinha 2014); these models appear ideal for our purposes. Successful identification of diverged *eve* enhancers could demonstrate an unprecedented level of predictive power for these quantitative models; I personally hope they're up to the challenge.

Final thoughts

I'll conclude this work with some general statements on scientific philosophy and a few moments of honesty. As a computational neophyte, I've discovered the awesome power of the 'model first' doctrine espoused by leaders in the field (Gunawardena 2014; Phillips 2015). Though sometimes an impossible standard, this approach improves your experiments because, by design, both positive and negative results are useful. Positive results can demonstrate the predictive utility of a given model, while negative results can reveal features of the system we do not understand. I've also learned the secret alchemy of systems biology: numbers can turn old biology into new biology. The *eve* enhancers have been studied for decades, yet quantitative measurements and computational models have revealed new features that challenge basic assumptions in the field. Finally, as a geneticist at heart, I stand in awe of the *eve* locus and the Drosophila embryo. This system is an experimental cornucopia, overflowing with opportunity and insight for anyone who wants to look. I believe it represents our best hope to connect DNA sequence features, network topology, organismal morphology and evolution. I'm grateful for the

opportunity to contribute to the development of the system, and to developmental and systems biology in general. For any budding geneticists who have made it this far, ***eve*** really is the new ***Lac***.

References

Adelman, K. & Lis, J.T., 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature reviews. Genetics, 13(10), pp.720–731.

Andrioli, L.P.M. et al., 2002. Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. Development, 129(21), pp.4931–4940.

Arnosti, D.N. et al., 1996. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development, 122(1), pp.205–214.

Beagrie, R.A. & Pombo, A., 2016. Gene activation by metazoan enhancers: Diverse mechanisms stimulate distinct steps of transcription. BioEssays: news and reviews in molecular, cellular and developmental biology, 38(9), pp.881–893.

Blythe, S.A. & Wieschaus, E.F., 2016. Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. eLife, 5. Available at: http://dx.doi.org/10.7554/eLife.20148.

Bonn, S. et al., 2012. Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. Nature protocols, 7(5), pp.978–994.

Bothma, J.P. et al., 2014. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. Proceedings of the National Academy of Sciences of the United States of America, 111(29), pp.10598–10603.

Bothma, J.P. et al., 2015. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. eLife, 4. Available at: http://dx.doi.org/10.7554/eLife.07956.

Bowman, S.K. et al., 2014. H3K27 modifications define segmental regulatory domains in the Drosophila bithorax complex. eLife, 3, p.e02833.

Chen, K. et al., 2013. A global change in RNA polymerase II pausing during the Drosophila midblastula transition. eLife, 2, p.e00861.

Clyde, D.E. et al., 2003. A self-organizing system of repressor gradients establishes segmental complexity in Drosophila. Nature, 426(6968), pp.849–853.

Crocker, J., Ilsley, G.R. & Stern, D.L., 2016. Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors. Nature genetics, 48(3), pp.292–298.

Crocker, J. & Stern, D.L., 2013. TALE-mediated modulation of transcriptional enhancers in vivo. Nature methods, 10(8), pp.762–767.

Crocker, J., Tsai, A. & Stern, D.L., 2017. A Fully Synthetic Transcriptional Platform for a Multicellular Eukaryote. Cell reports, 18(1), pp.287–296.

Driever, W., Thoma, G. & 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(6232), pp.363–367.

Drosophila 12 Genomes Consortium et al., 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature, 450(7167), pp.203–218.

Erceg, J. et al., 2014. Subtle changes in motif positioning cause tissue-specific effects on robustness of an enhancer's activity. PLoS genetics, 10(1), p.e1004060.

Estrada, J. et al., 2016. Information Integration and Energy Expenditure in Gene Regulation. Cell, 166(1), pp.234–244.

Estrada, J. et al., 2016. SiteOut: An Online Tool to Design Binding Site-Free DNA Sequences. PloS one, 11(3), p.e0151740.

Fakhouri, W.D. et al., 2010. Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo. Molecular systems biology, 6, p.341.

Frankel, N. et al., 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature, 466(7305), pp.490–493.

Frasch, M. et al., 1987. Characterization and localization of the even-skipped protein of Drosophila. The EMBO journal, 6(3), pp.749–759.

Fuda, N.J., Ardehali, M.B. & Lis, J.T., 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature, 461(7261), pp.186–192.

Gallo, S.M. et al., 2011. REDfly v3. 0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic acids research, 39(suppl 1), pp.D118–D123.

Garcia, H.G. et al., 2013. Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Current biology, 23(21), pp.2140–2145.

Ghavi-Helm, Y. et al., 2014. Enhancer loops appear stable during development and are associated with paused polymerase. Nature, 512(7512), pp.96–100.

Gilbert, L.A. et al., 2014. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell, 159(3), pp.647–661.

Gray, S., Szymanski, P. & Levine, M., 1994. Short-range repression permits multiple enhancers to function autonomously within a complex promoter. Genes & development, 8(15), pp. 1829–1838.

Green, M.R., 2005. Eukaryotic transcription activation: right on target. Molecular cell, 18(4), pp. 399–402.

Gregor, T. et al., 2007. Probing the limits to positional information. Cell, 130(1), pp.153–164.

Hang, S. & Gergen, J.P., 2017. Different modes of enhancer-specific regulation by Runt and Even-skipped during Drosophila segmentation. Molecular biology of the cell. Available at: http://dx.doi.org/10.1091/mbc.E16-09-0630.

Harlen, K.M. & Churchman, L.S., 2017. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nature reviews molecular cell biology. Available at: http://dx.doi.org/10.1038/nrm.2017.10.

He, Q., Johnston, J. & Zeitlinger, J., 2015. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. Nature biotechnology, 33(4), pp.395–401.

He, X. et al., 2010. Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. PLoS computational biology, 6(9), p.e1000935.

Hoermann, A., Cicin-Sain, D. & Jaeger, J., 2016. A quantitative validated model reveals two phases of transcriptional regulation for the gap gene giant in Drosophila. Developmental biology, 411(2), pp.325–338.

Ilsley, G.R. et al., 2013. Cellular resolution models for even skipped regulation in the entire Drosophila embryo. eLife, 2, p.e00522.

Jaeger, J., 2011. The gap gene network. Cellular and molecular life sciences: CMLS, 68(2), pp. 243–274.

Jaeger, J., Manu & Reinitz, J., 2012. Drosophila blastoderm patterning. Current opinion in genetics & development, 22(6), pp.533–541.

Janssens, H. et al., 2006. Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. Nature genetics, 38(10), pp.1159–1165.

Johnson, L.A. et al., 2008. Reverse-Engineering a Transcriptional Enhancer: A Case Study in Drosophila. Tissue engineering. Part A, 14(9), pp.1549–1559.

Kazemian, M. et al., 2010. Quantitative analysis of the Drosophila segmentation regulatory network using pattern generating potentials. PLoS biology, 8(8). Available at: http://dx.doi.org/10.1371/journal.pbio.1000456.

Kim, A.-R. et al., 2013. Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic. PLoS genetics, 9(2), p.e1003243.

Kim, H.D. et al., 2009. Transcriptional regulatory circuits: predicting numbers from alphabets. Science, 325(5939), pp.429–432.

Larson, M.H. et al., 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nature protocols, 8(11), pp.2180–2196.

Liberman, L.M. & Stathopoulos, A., 2009. Design flexibility in cis-regulatory control of gene expression: synthetic and comparative evidence. Developmental biology, 327(2), pp. 578–589.

Li, L.M. & Arnosti, D.N., 2011. Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes. Current biology, 21(5), pp.406–412.

Lott, S.E. et al., 2011. Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq. PLoS biology, 9(2), p.e1000590.

Ludwig, M.Z. et al., 2011. Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. PLoS genetics, 7(11), p.e1002364.

Lydiard-Martin, T. et al., 2014. Locus architecture affects mRNA expression levels in Drosophila embryos. bioRxiv, p.005173. Available at: http://biorxiv.org/content/early/2014/05/14/005173?rss=1 [Accessed October 21, 2015].

Mannervik, M., 2014. Control of Drosophila embryo patterning by transcriptional co-regulators. Experimental cell research, 321(1), pp.47–57.

Ma, Y. et al., 1998. Gene regulatory functions of Drosophila fish-hook, a high mobility group domain Sox protein. Mechanisms of development, 73(2), pp.169–182.

McCarty, A.S. et al., 2003. Selective dimerization of a C2H2 zinc finger subfamily. Molecular cell, 11(2), pp.459–470.

Moshe, A. & Kaplan, T., 2017. Genome-wide Search for Zelda-like Chromatin Signatures Identifies GAF as a Pioneer Factor in Early Fly Development. bioRxiv, p.104067. Available at: http://biorxiv.org/content/early/2017/01/29/104067 [Accessed March 18, 2017].

Nambu, P.A. & Nambu, J.R., 1996. The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. Development, 122(11), pp. 3467–3475.

Nibu, Y. & Levine, M.S., 2001. CtBP-dependent activities of the short-range Giant repressor in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 98(11), pp.6204–6208.

Nibu, Y., Senger, K. & Levine, M., 2003. CtBP-independent repression in the Drosophila embryo. Molecular and cellular biology, 23(11), pp.3990–3999.

Nibu, Y., Zhang, H. & Levine, M., 1998. Interaction of short-range repressors with Drosophila CtBP in the embryo. Science, 280(5360), pp.101–104.

Nüsslein-Volhard, C. & Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature, 287(5785), pp.795–801.

Paris, M. et al., 2013. Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression. PLoS genetics, 9(9), p.e1003748.

Payankaulam, S. & Arnosti, D.N., 2009. Groucho corepressor functions as a cofactor for the Knirps short-range transcriptional repressor. Proceedings of the National Academy of Sciences of the United States of America, 106(41), pp.17314–17319.

Perry, M.W. et al., 2010. Shadow enhancers foster robustness of Drosophila gastrulation. Current biology, 20(17), pp.1562–1567.

Peterson, B.K. et al., 2009. Big genomes facilitate the comparative identification of regulatory elements. PloS one, 4(3), p.e4688.

Qi, L.S. et al., 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell, 152(5), pp.1173–1183.

Ren, X. et al., 2014. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. Cell reports, 9(3), pp.1151–1162.

Rhee, H.S. & Pugh, B.F., 2012. ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. Current protocols in molecular biology, Chapter 21, p.Unit 21.24.

Samee, M.A.H. & Sinha, S., 2014. Quantitative modeling of a gene's expression from its intergenic sequence. PLoS computational biology, 10(3), p.e1003467.

Scholes, C., DePace, A.H. & Sánchez, Á., 2017. Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle. Cell systems, 4(1), pp.97–108.e9.

Schwartz, S. et al., 2003. Human-mouse alignments with BLASTZ. Genome research, 13(1), pp. 103–107.

Shlyueva, D., Stampfel, G. & Stark, A., 2014. Transcriptional enhancers: from properties to genome-wide predictions. Nature reviews. Genetics, 15(4), pp.272–286.

Simpson-Brose, M., Treisman, J. & Desplan, C., 1994. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. Cell, 78(5), pp. 855–865.

Small, S., Blair, A. & Levine, M., 1992. Regulation of even-skipped stripe 2 in the Drosophila embryo. The EMBO journal, 11(11), pp.4047–4057.

Staller, M.V. et al., 2013. Depleting gene activities in early Drosophila embryos with the "maternal-Gal4-shRNA" system. Genetics, 193(1), pp.51–61.

Staller, M.V. et al., 2015. Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 112(3), pp.785–790.

Struffi, P. et al., 2011. Combinatorial activation and concentration-dependent repression of the Drosophila even skipped stripe 3+7 enhancer. Development, 138(19), pp.4291–4299.

Tran, K.D., Miller, M.R. & Doe, C.Q., 2010. Recombineering Hunchback identifies two conserved domains required to maintain neuroblast competence and specify early-born neuronal identity. Development, 137(9), pp.1421–1430.

Wunderlich, Z. et al., 2015. Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. Cell reports, 12(11), pp.1740–1747.

Zhu, L.J. et al., 2011. FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system. Nucleic acids research, 39(Database issue), pp.D111–7.

APPENDIX A: SUPPLEMENTAL MATERIAL FOR CHAPTER 1

Supplemental Methods

Construct design and transgenic fly line creation

Reconstituted enhancers were designed using annotated binding sites from RedFly (Gallo et al. 2011) separated by spacer sequences generated by SiteOut (Estrada et al. 2016). We screened against sites for Bicoid, Hunchback, Zelda, Krüppel, Giant, Caudal, Tailless, Dichaete, Stat92E, Knirps, and Nubbin. Enhancer sequences were synthesized by Integrated DNA Technologies and cloned into the pBΦY vector between the NotI and BglII restriction sites using Gibson assembly (Gibson 2009). The pBΦY vector contains the *even-skipped* basal promoter driving *lacZ*, AmpR and mini-white cassettes, and an attB site for site-specific genome integration using the PhiC31 system (Hare et al. 2008; Groth et al. 2004). Constructs were injected into w118 flies containing the attP2 integration site using commercial services provided by BestGene. Successful transformants were screened for mini-white and homozygosed.

In situ hybridization and imaging

In situ hybridization was performed as previously described (Luengo Hendriks et al. 2006). Briefly, embryos were collected for 0-4 hours at 25˚C and fixed in heptane and 5% paraformaldehyde for 25 minutes. We then incubated the embryos for 2 days with digoxigenin (DIG)-labeled probes for *fushi tarazu* (*ftz*) and dinitrophenol (DNP)-labeled probes for *lacZ* and *huckebein* (*hkb*). *hkb* is expressed at the poles of the embryos and is used as a co-stain to normalize expression levels between different transgenic lines (Wunderlich et al. 2014). Probes were then detected following sequential incubation with commercial anti-DIG or anti-DNP antibodies conjugated to horseradish peroxidase from PerkinElmer, as well as a color reaction with coumarin-tyramide or Cy3-tyramide, respectively. Nuclei were labelled using Sytox Green from Life Technologies, and embryos were mounted in DePex from Electron Microscopy

Sciences. Stage 5 embryos were sorted into one of six time classes using the extent of membrane invagination as a morphological marker. Data presented in the main text was taken from the third cohort (9-25% membrane invagination). Three-dimensional images of individual embryos were obtained using two-photon laser scanning microscopy. We also include data from embryos that did not contain a *hkb* co-stain (see Supplemental Data, below).

Image processing and analysis

Image stacks from individual embryos were converted into PointCloud files containing the xyz coordinates and expression levels of *ftz* or *lacZ*/*hkb* for each nucleus (Luengo Hendriks et al. 2006). Individual pointclouds were averaged together into a gene expression atlas using the software described in (Fowlkes et al. 2008). Normalized *lacZ* levels were calculated for individual embryos using the *hkb* co-staining method (Wunderlich et al. 2014). Line traces were plotted using the extractpattern function contained in the PointCloud toolbox (http://bdtnp.lbl.gov/Fly-Net/bioimaging.jsp?w=analysis) which calculates the mean expression level in 100 bins along the anterior-posterior axis for 16 strips around the circumference of the embryo. We used the average of the 4th and 5th strip for our analysis.

Binding site predictions

Binding sites for Caudal were predicted using a position weight matrix generated by bacterial 1-hybrid methods (Noyes et al. 2008) and the PATSER software using a p-value cutoff of 0.01 (Hertz & Stormo 1999). Binding sites were visualized using the inSite software (http://www.cs.utah.edu/~miriah/insite/).

Supplemental data

Available at figshare.com:

https://figshare.com/articles/Cellular_resolution_data_from_wild_type_and_reconstituted_eve_stripe_2_enhancer_reporter_constructs/2948377

Enhancer sequences

The following sequences correspond to the wild-type and reconstituted versions of *eve2* that were assayed in reporter constructs. Lowercase letters indicate wild-type sequence, while uppercase letters indicate designed spacer sequences. One base pair in a footprinted Bicoid site was altered in both reconstituted constructs; this base pair is indicated in magenta.

>*eve2,* DePace stock #0216
ggttacccggtactgcataacaatggaacccgaaccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaat
ccgtttgccatcagcgagattattagtcaattgcagttgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttga
acaatcgtcgcagtttggtaacacgctgtgccatactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggtt
gcgaagtcagggcattccgccgatctagccatcgccatcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaa
tccaatcccgatccctagcccgatcccaatcccaatcccaatcccttgtccttttcattagaaagtcataaaaacacataataatgatgtcgaa
gggattagggg

>*reconstituted eve2 – spacer set 1,* DePace stock #0553
ACTGGACAGTttacccggtacCCGACTGCGTTACTATGGCGACTAtaactgggacaCACATCAAACACCAT
ctggtttCACGGTCGATACTAAGTgttaatccgttAGATGATGgcgagattattagtcaattgcagttgcTTGAGATCT
TTTAATTCTGTTGAATGGGATCCTAAACgactttattgcagcatcttgaacaatcgtcgcagtttggtaacacgctCATT
ACTTAGACGCACCagacggaatcgagggaccctggactataatcgcTCCGAATTaccgggttgcGGTACGTTTCTCA
TTCTGTGACCGGAGGCCTATGTCTGGTATCATGTAGGAgtttgtttgtttgctgggattagccaagggcttgaAG
ATGGGtccaatcccgatccctagcccgatcccaatcccaatcccaatccctTATTTACTTAGACTgaaagtcataaaaacacat
aataTCTTTGAcgaagggattaggCCTAACTACA

>*reconstituted eve2 – spacer set 2,* DePace stock #0554
ACTCGTATACttacccggtacAAACTGTGTTGGGGATTTCTCGTAtaactgggacaGGAAACGACGTACA
GctggtttGACAATGTAAGATAAATgttaatccgttACAACTCCgcgagattattagtcaattgcagttgcGTGCATGT
TCACTCGCTGTTTATCTATACGAATAGAAgactttattgcagcatcttgaacaatcgtcgcagtttggtaacacgctAGG
GGTTGTGCTGTACTagacggaatcgagggaccctggactataatcgcTAGTCAATaccgggttgcTGAGCGAAATA
CACGTTGTCGTCAGTACTGATTTGTACTAGGCATATGCTCgtttgtttgtttgctgggattagccaagggcttgaG
TGTTGAtccaatcccgatccctagcccgatcccaatcccaatcccaatccctGTTTAGCAGCGAAGgaaagtcataaaaacac
ataataGCTAGTAcgaagggattaggCATACTACAC

References

Estrada, J. et al., 2016. SiteOut: An Online Tool to Design Binding Site-Free DNA Sequences. PloS one, 11(3), p.e0151740.

Fowlkes, C.C. et al., 2008. A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell, 133(2), pp.364–374.

Gallo, S.M. et al., 2011. REDfly v3. 0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic acids research, 39(suppl 1), pp.D118–D123.

Gibson, D.G., 2009. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic acids research, 37(20), pp.6984–6990.

Groth, A.C. et al., 2004. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics, 166(4), pp.1775–1782.

Hare, E.E. et al., 2008. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS genetics, 4(6), p.e1000106.

Hertz, G.Z. & Stormo, G.D., 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics, 15(7-8), pp.563–577.

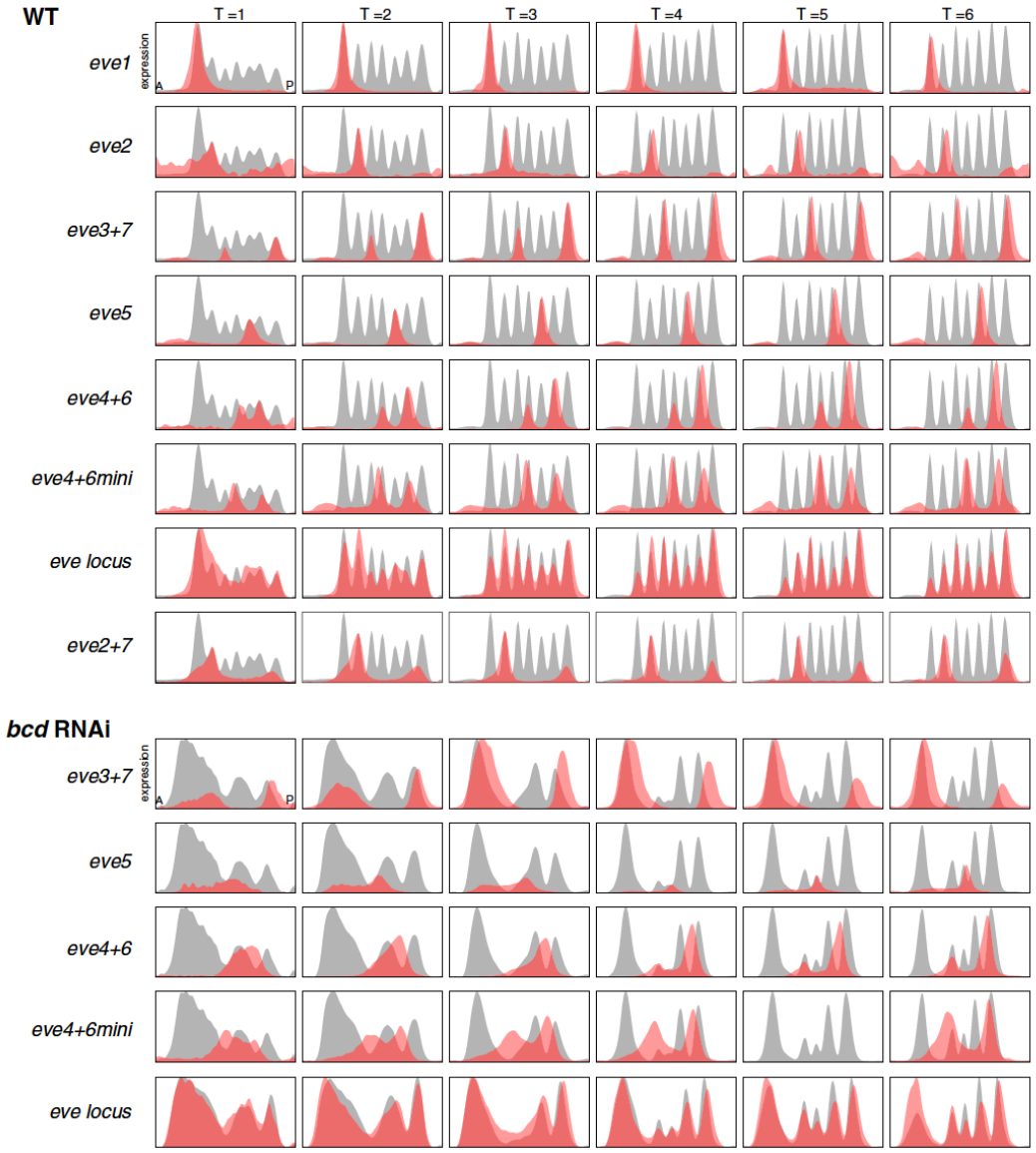
Luengo Hendriks, C.L. et al., 2006. Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. Genome biology, 7(12), p.R123.

Noyes, M.B. et al., 2008. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nucleic acids research, 36(8), pp.2547–2560.

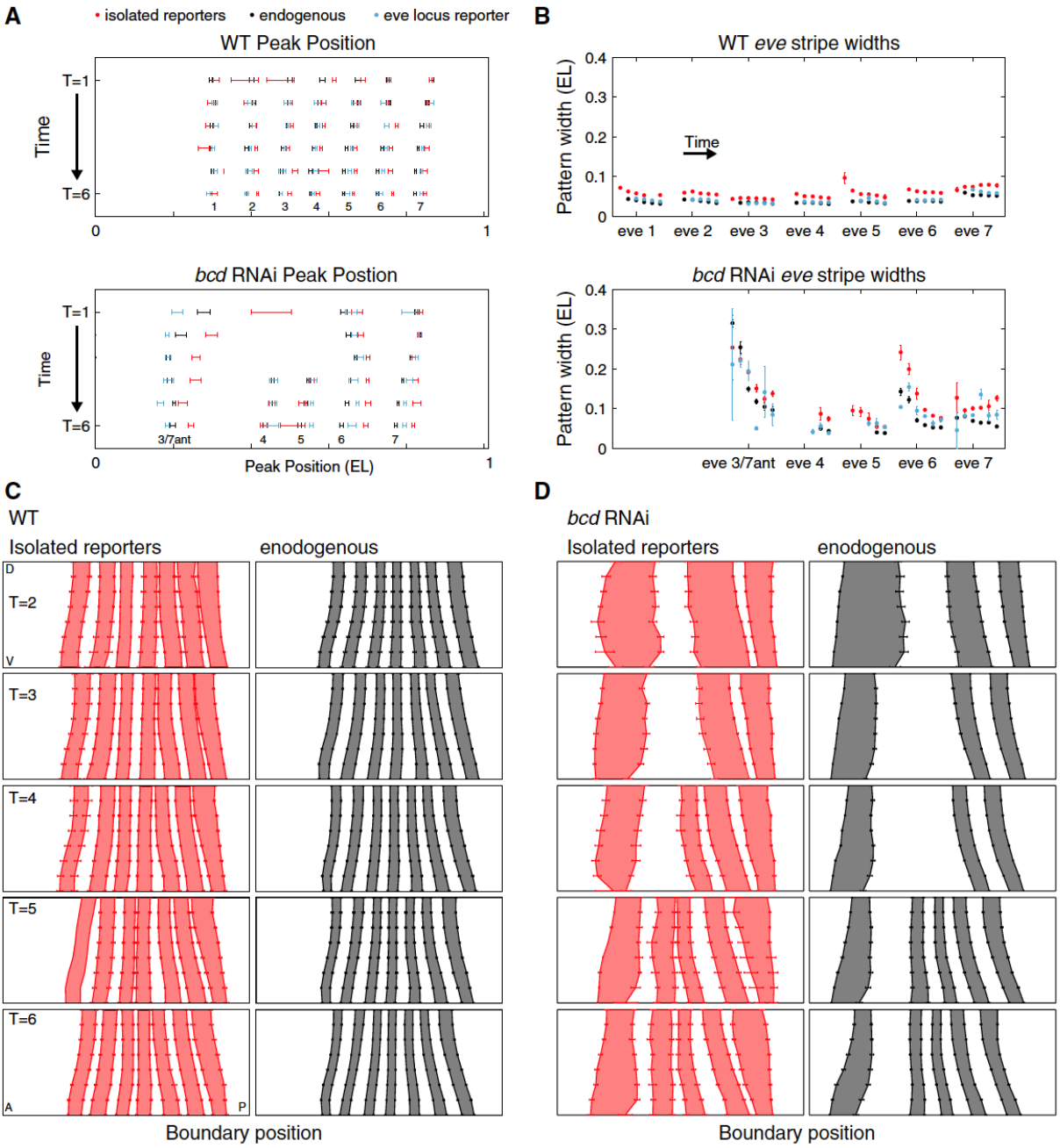
Wunderlich, Z., Bragdon, M.D. & DePace, A.H., 2014. Comparing mRNA levels using in situ hybridization of a target gene and co-stain. Methods, 68(1), pp.233–241.

APPENDIX B: SUPPLEMENTAL MATERIAL FOR CHAPTER 2

Supplemental Figures

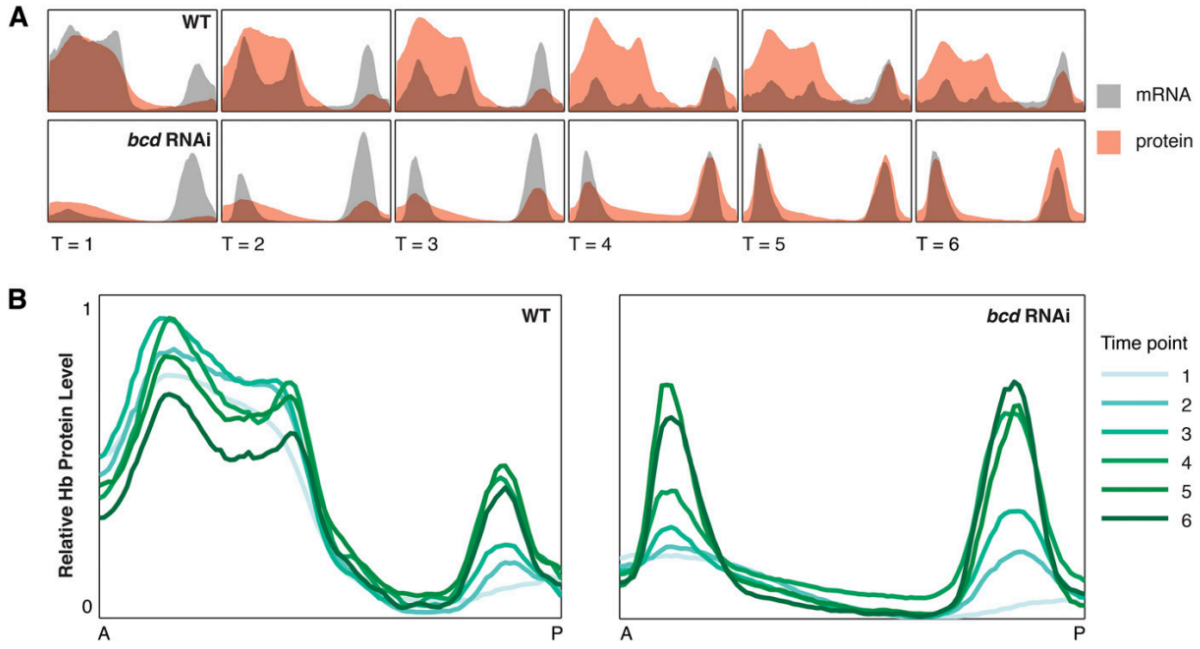


Supplemental Figure 2.1: *eve* enhancer *lacZ* reporters overlap the corresponding endogenous patterns with varying fidelity. Line traces of *lacZ* enhancer reporters (red) and the endogenous *eve* (gray) mRNA pattern in wild-type and *bcd* RNAi gene expression atlases. AP position is plotted on the x axis and expression level on the y axis for a lateral strip of the embryo. Reporter patterns were manually scaled to match endogenous peaks.

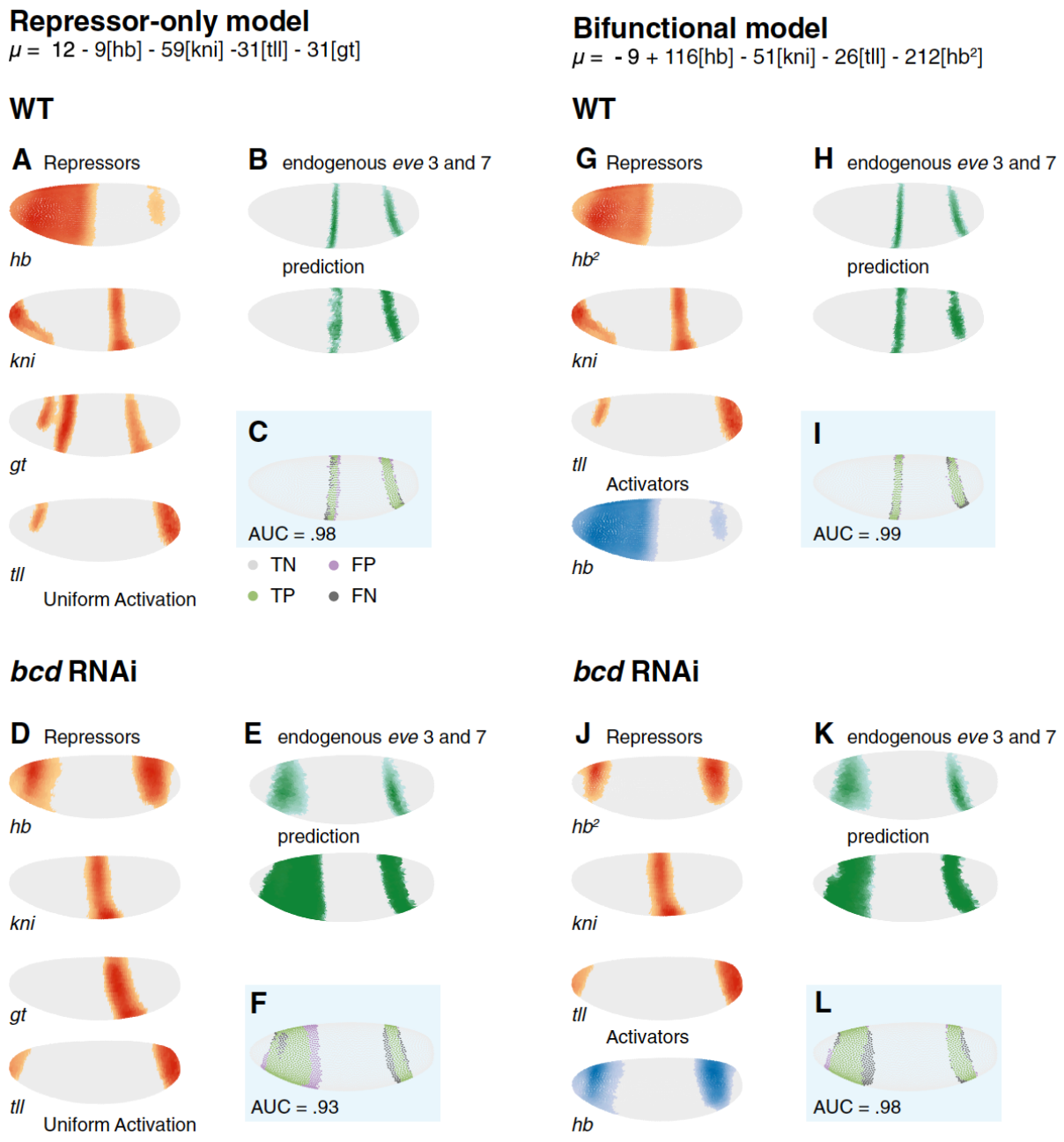


Supplemental Figure 2.2: The expression pattern driven by the whole locus reporter is more similar to the endogenous pattern than the traditional reporters. (A) The reporter peak positions (red) are slightly posterior to the endogenous eve peaks (black) and whole locus reporter peaks (blue). Peak positions calculated from lateral line traces in Supplemental Figure 2.1. The anterior *eve3+7* pattern is faint and broad at time point 1 and the peak is close to the middle of the embryo as seen in the lateral line trace in Supplemental Figure 2.1 (B) Stripes driven by the traditional reporters (red) are wider than endogenous stripes (black) and whole locus reporter (blue) in wild-type and *bcd* RNAi embryos. Widths calculated from lateral line traces in Supplemental Figure 2.1. In wild-type, some of the error bars are

Supplemental Figure 2.2 (Continued). smaller than the diameter of the point. (C and D) Boundary positions of the traditional reporters (red) and endogenous stripes (gray) in wild-type (C) and *bcd* RNAi embryos (D). Note that the ventral-most part of the *eve3+7* reporter anterior pattern is very faint in *bcd* RNAi embryos and this boundary is not reliably detected by our software.

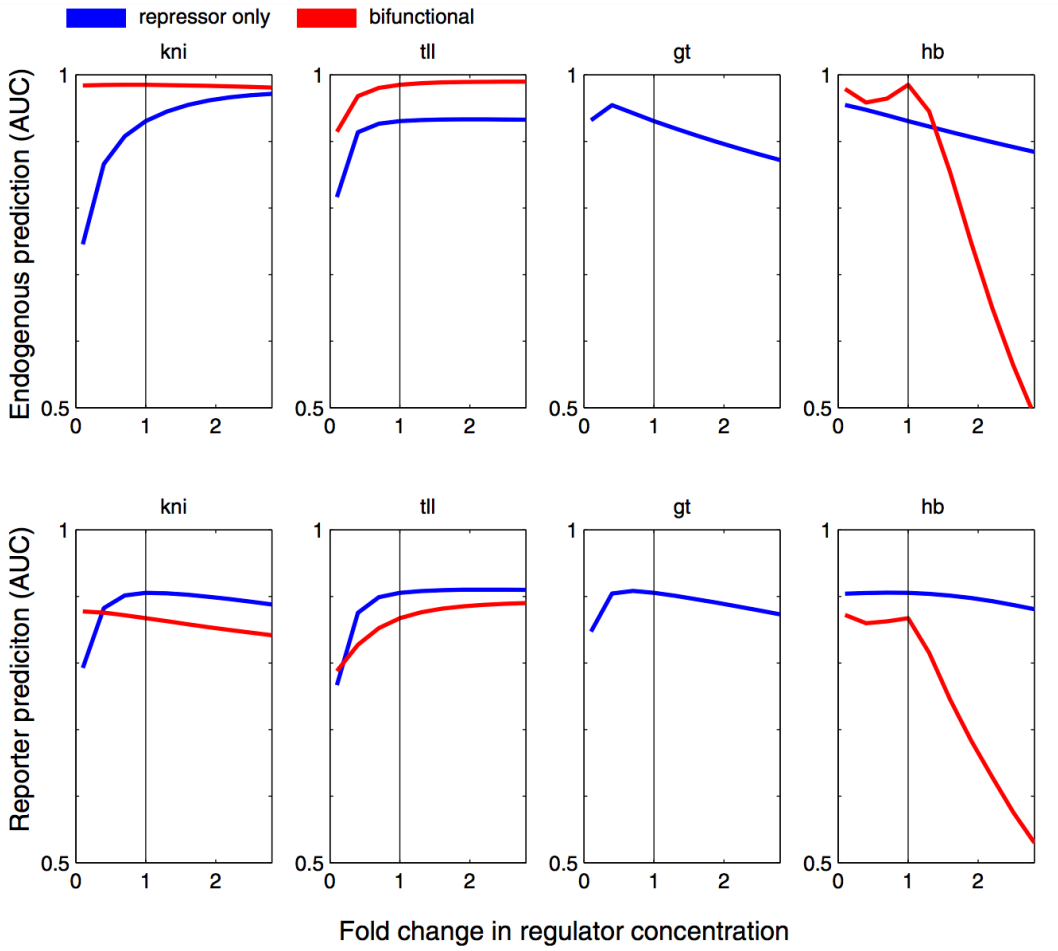


Supplemental Figure 2.3: *bcd* RNAi perturbs *hb* mRNA and protein levels. (A) We used Hb protein data for the computational modeling because in both WT and *bcd* RNAi embryos *hb* mRNA (gray) and protein (red) patterns are different. (B) The Hb protein expression pattern changes over stage 5 in both wild-type and *bcd* RNAi embryos. In wild-type embryos, both maternal and *bcd*-activated zygotic mRNA contribute to the anterior pattern, whereas in *bcd* RNAi only maternal mRNA contributes (Tautz et al. 1988). Note that each atlas is normalized separately, so absolute levels are not comparable between atlases. Relative levels change extensively. Data reproduced from Staller et al. 2015.

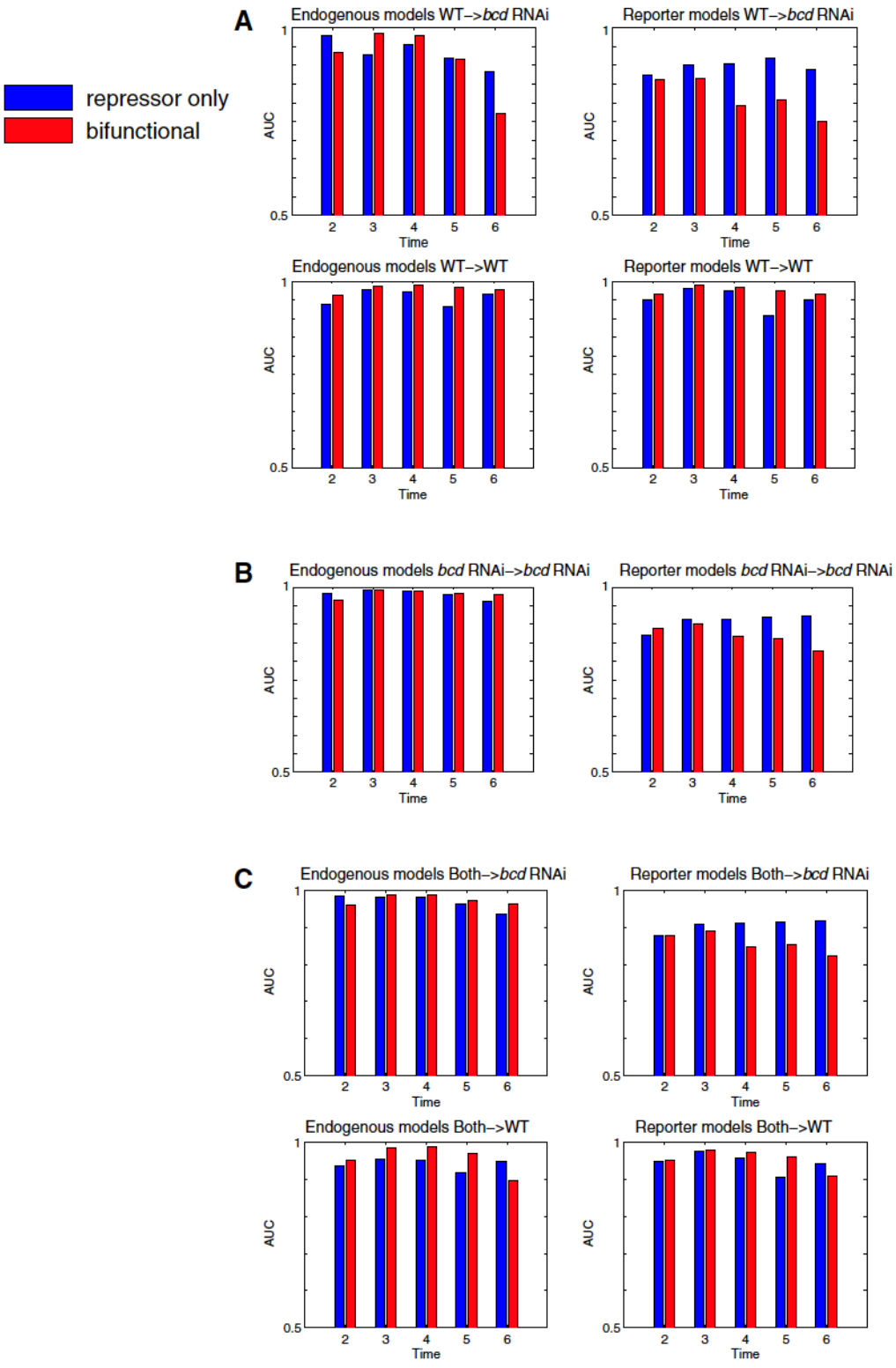


Supplemental Figure 2.4: Under bcd perturbation, the expression patterns of endogenous *eve* stripes 3 and 7 are more accurately predicted by the bifunctional model. (A) WT expression patterns of the regulators in the repressor-only model. The expression level of each TF is shown for every cell. Cells with expression below an ON/OFF threshold (Materials and Methods) are plotted in gray. For cells above this threshold, color intensity represents expression level. Repressors are red and activators are blue. (B) The expression pattern of the endogenous eve stripes 3 and 7 and the predictions of the repressor-only model in WT. (C) Comparison of predictions to measurement in WT embryos. Green cells are true positives (TP), purple cells are false positives (FP), dark gray cells are false negatives (FN), and light gray cells are true negatives (TN). For visualization, the threshold is set to 80%

Supplemental Figure 2.4 (Continued). sensitivity, but the AUC metric quantifies performance over all thresholds. (D) The expression patterns of the regulators in the repressor-only model in *bcd* RNAi embryos. (E) The expression pattern of the endogenous *eve* stripes 3 and 7 and the predictions of the repressor-only model in *bcd* RNAi. (F) Comparison of repressor-only model predictions to data in *bcd* RNAi. (G–L) Same as A–F, respectively, for the bifunctional model.

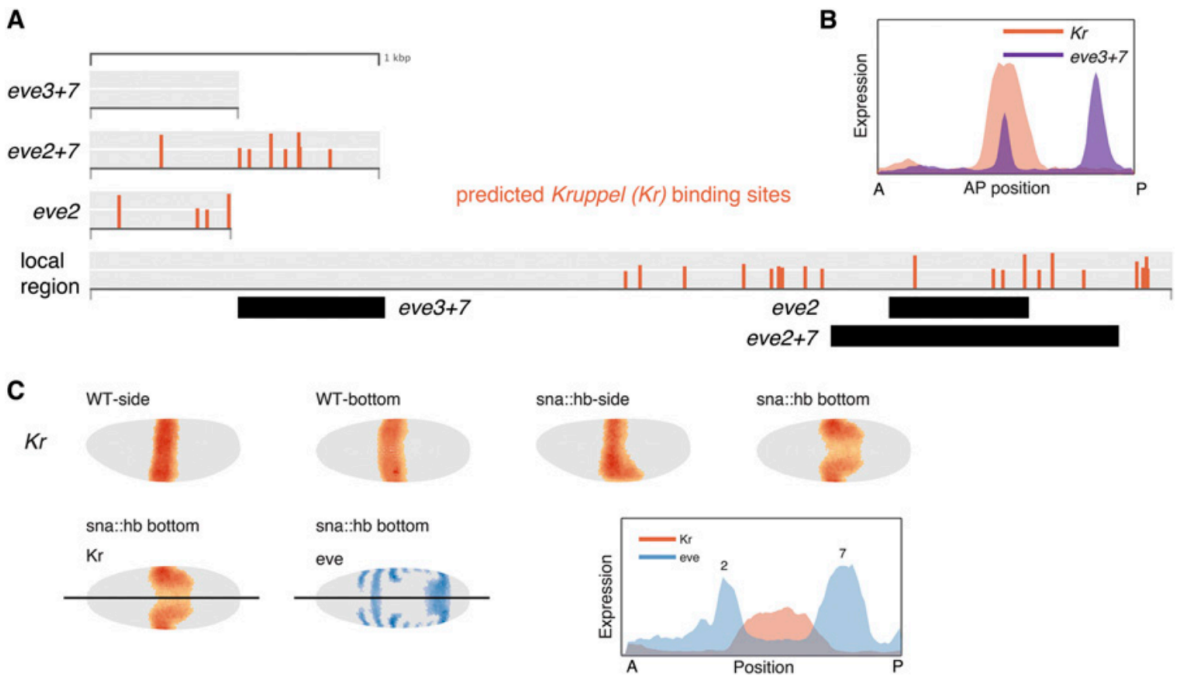


Supplemental Figure 2.5: Scaling the relative level of a TF between atlases generally does not change model relative performance. We varied the concentration of each TF separately in the *bcd* RNAi atlas and recalculated the AUC of the repressor-only and bifunctional models. This scaling simulates possible global changes in levels between genotypes. For the endogenous pattern, for all scalings of *kni* and *tll*, the bifunctional model is more accurate than the repressor-only model. For Hb, the bifunctional model is more accurate than the repressor-only model so long as maximal Hb levels in *bcd* RNAi are less than 1.38x maximal WT levels. Because *bcd* is a potent activator of Hb, Hb levels are very likely reduced in *bcd* RNAi embryos. For the reporter pattern all scalings of Hb preserve relative model performance. The repressor-only model is more accurate for a broad scaling of *kni* and *tll* levels.

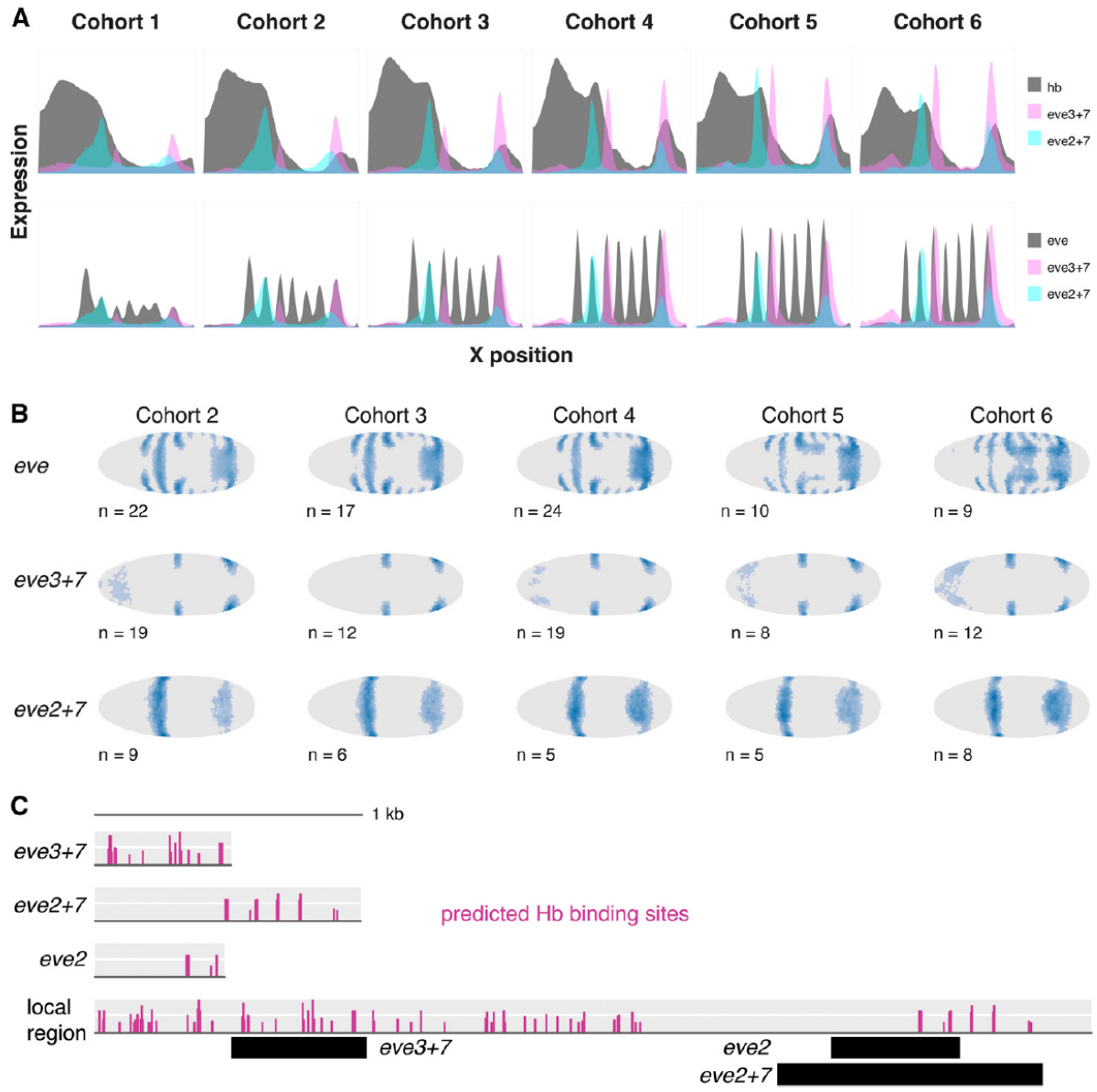


Supplemental Figure 2.6: Fitting the repressor-only and bifunctional models on different datasets yielded similar results. (A) Results after fitting the models in WT at different time points and predicting the corresponding time points. The repressor-only model

Supplemental Figure 2.6 (Continued). always more accurately predicted the reporter in the *bcd* RNAi background. Although both models are very accurate in WT, the bifunctional model is more accurate. (B) Results after fitting the models in *bcd* RNAi and predicting *bcd* RNAi. The repressor-only model more accurately predicted the reporter pattern. For the endogenous pattern, both models performed well. (C) Fitting the models on both the WT and *bcd* RNAi datasets led to similar results: The bifunctional model more accurately predicted the endogenous pattern and the repressor-only model more accurately predicted the reporter pattern in the *bcd* RNAi background.



Supplemental Figure 2.7: The expansion of the *Krüppel* expression pattern explains the shape of the *eve2+7* reporter pattern in *sna::hb* embryos. (A) The *eve2* enhancer is enriched for predicted Kr binding sites (red), whereas the *eve3+7* enhancer is depleted for Kr binding sites. We predicted binding sites using PATSER (Hertz and Stormo 1999) with a position weight matrix derived from bacterial one-hybrid data (Noyes et al. 2008). (B) *Kr* expression overlaps stripe 3 of the *eve3+7* reporter mRNA in wild-type embryos. *Kr* does not repress this pattern, which is consistent with the absence of binding sites. (C) The distribution of *Kr* mRNA in wild-type and *sna::hb* embryos. The expanded ventral region of the *Kr* expression pattern appears to set the boundary of the expanded *eve* stripe 7 pattern.



Supplemental Figure 2.8: Quantitative differences between the *eve* stripe 7 shadow enhancers. (A) Top: Line traces of Hb protein, *eve3+7 lacZ* (pink), and *eve2+7 lacZ* (cyan). Hb overlaps the stripe 7 pattern driven by both reporters. Bottom: Line traces of *eve*, *eve3+7 lacZ*, and *eve2+7 lacZ*. Neither reporter perfectly matches the endogenous pattern. Traces were manually scaled to match the corresponding endogenous stripes. (B) Computational renderings of gene expression atlas data from *sna::hb* embryos at different time points. The number of embryos included in each time point of the gene expression atlas is shown. (C) Predicted Hb binding sites in enhancer sequences and intervening sequences upstream of the *eve* coding sequence. Binding sites were predicted and represented as in Supplemental Figure 2.7.

Supplemental Tables

Supplemental Table 2.1: Model parameters

		Linear					Quadratic				
	Time	Constant	*hb*	*kni*	*tll*	*gt*	Constant	*hb*	*kni*	*tll*	hb^2
Endogenous	2	5	−5	−21	−9	−15	−5	68	−22	−7	−134
hb protein	3	12	−9	−59	−31	−31	−9	116	−51	−26	−212
Other mRNA	4	11	−5	−40	−51	−27	−9	101	−42	−67	−137
	5	5	−2	−17	−25	−14	−7	73	−26	−49	−95
	6	10	−4	−48	−37	−33	2	49	−50	−58	−77
Endogenous	3	12	−9	−59	−31	−31	−9	116	−51	−26	−212
Ilsley et al. (1) (includes gt protein)	3	11	−10	−66	−22	−13	−8	110	−45	−25	−201
Reporter	2	6	−6	−30	−6	−16	−3	64	−37	−4	−134
hb protein	3	12	−11	−72	−18	−29	−10	145	−73	−14	−286
Other mRNA	4	11	−7	−37	−22	−43	−13	105	−25	−16	−144
	5	4	−1	−13	−14	−15	−12	83	−15	−22	−103
	6	8	−4	−26	−22	−40	−5	59	−19	−23	−91

Supplemental Table 2.2: AUC scores

	Linear		Quadratic	
T = 3	Ilsley et al. (1)	Refit	Ilsley et al. (1)	Refit
WT	0.9622	0.9786	0.9874	0.9874
bcd RNAi	0.9321	0.9275	0.985	0.9849

Enhancer Sequences

>*eve3+7,* DePace stock #0204
GGATCCTCGAAATCGAGAGCGACCTCGCTGCATTAGAAAACTAGATCAGTTTTTTGTTTTGGCC
GACCGATTTTTGTGCCCGGTGCTCTCTTTACGGTTTATGGCCGCGTTCCCATTTCCCAGCTTCTT
TGTTCCGGGCTCAGAAATCTGTATGGAATTATGGTATATGCAGATTTTTATGGGTCCCGGCGAT
CCGGTTCGCGGAACGGGAGTGTCCTGCCGCGAGAGGTCCTCGCCGGCGATCCTTGTCGCCCGT
ATTAGGAAAGTAGATCACGTTTTTTGTTCCCATTGTGCGCTTTTTTCGCTGCGCTAGTTTTTTTC
CCCGAACCCAGCGAACTGCTCTAATTTTTTAATTCTTCACGGCTTTTCATTGGGCTCCTGGAAA
AACGCGGACAAGGTTATAACGCTCTACTTACCTGCAATTGTGGCCATAACTCGCACTGCTCTCG
TTTTTAAGATCCGTTTGTTTGTGTTTGTTTGTCCGCGATGGCATTCACGTTTTTACGAGCTC

>*eve2+7,* DePace stock #409 [note that this construct has a polymorphism (in green) and 6bp deletion (ATCCCA between orange bases) that were corrected in DePace stock #0547]
AGAAGGCTTGCATGTGGGCCTTTTCCAGGTCGGCCAGTAGGTAGAGTTGTTGCGATGCGGCTA
TGCCGGGCGAGTTAATGCCAATGCAAATTGCGGGCGCAATATAACCCAATAATTTGAAGTAAC
TGGCAGGAGCGAGGTATCCTTCCTGGTTACCCGGTACTGCATAACAATGGAACCCGAACCGTA
ACTGGGACAGATCGAAAAGCTGGCCTGGTTTCTCGCTGTGTGTGCCGTGTTAATCCGTTTGCCA
TCAGCGAGATTATTAGTCAATTGCAGTTGCAGCGTTTCGCTTTCGTCCTCGTTTCACTTTCGAG
TTAGACTTTATTGCAGCATCTTGAACAATCGTCGCAGTTTGGTAACACGCTGTGCCATACTTTC
ATTTAGACGGAATCGAGGGACCCTGGACTATAATCGCACAACGAGACCGGGTTGCGAAGTCAG
GGCATTCCGCCGATCTAGCCATCGCCATCTTCTGCGGGCGTTTGTTTGTTTGTTTGCTGGGATT
AGCCAAGGGCTTGACTTGGAATCCAATCCTGATCCCTAGCCCGATCCCAATCCCAATCCCTTGT
CCTTTTCATTAGAAAGTCATAAAAACACATAATAATGATGTCGAAGGGATTAGGGGCGCGCAG
GTCCAGGCAACGCAATTAACGGACTAGCGAACTGGGTTATTTTTTTGCGCCGACTTAGCCCTGA
TCCGCGAGCTTAACCCGTTTTGAGCCGGGCAGCAGGTAGTTGTGGGTGGACCCCACGATTTTT
TTGGCCAAACCTCCAAGCTAACTTGCGCAAGTGGCAAGTGGCCGGTTTGCTGGCCCAAAAGAG
GAGGCACTATCCCGGTCCTGGTACAGTTGGTACGCTGGGAATGATTATATCATCATAATAAATG
TTTTGCCCAACGAAACCGAAAACTTTTCAAATTAAGTCCCGGCAACTGGGTTCCCATTTTCCATT
TTCCATGTTCTGCGGGCAGGGGCGGCCATTATCTCGCT

>*eve1,* DePace stock #0213
GGCCTAATCACTTCCCTGAAATGCATAATTGTGCCGCGGCTTTTGATACGCTCCTGGCGGAGAG
GGAGATGAGGAAAGGATGCACGGGAACCGCAGCCAAGTGGCAGTCGAGATTGGCAAATCCGC
CAGCGGACAATGCCCAGAGAATGGGCAACAAGTAGCGGCGAATTAGCAATCCTATCATGCTTT
TATGGCCGGCCAACTCTTGCCCGCGCATCTCAGTTCATCCGAAGCGGGACCAGGTCCAGGTTC
AAGTCGAGGTCCAGTACCCCTGCTATCCCGTCAACCCCTTTAGGGCGATAATCCTTCTAAATGT
TTGCATTAATTTCGAGGCGTGGACGGATTAGGGCGTGCTGGCTGGGCGGAACCCGCAGCAGAA
ACCGCCGAGGACACTGCACCGACTGACCTGCAGCCTACAGATCTCTGATCTTCGATCTCTAATC
CTTTCGCATTTGCAACTGACTTCTGCACTGGGTCCGCCCCTAATCCTTCCGCCGAGAAGGCGGC
AGAGTCGCGAGGTACTGGCCCGGGGTAATGGGATTATCTGCGATTACCCCAGATGATCCGCAG
AAAGTCAATCTGGTTCAGGGGCTAATTGTCAGCGAAGTCAACTAAATCCAATCCTTTCGCGCCC
CCTTCTGTTTATTTGTTTGTTTTCGTTTGTTTTGAGAATTTCTGGCAATTAAGTTGCCCGTTTTG

ATGCGCGGGGGCGGGTGCATCAAATCCTTTCGGCATACCTGTCCTGCACAAATGCTGAATTCC
GCATCCCATGGATACCCAGATATTCAGATATCCCAAGGC

>*eve2*, DePace stock #0216
TGGTTACCCGGTACTGCATAACAATGGAACCCGAACCGTAACTGGGACAGATCGAAAAGCTGG
CCTGGTTTCTCGCTGTGTGTGCCGTGTTAATCCGTTTGCCATCAGCGAGATTATTAGTCAATTG
CAGTTGCAGCGTTTCGCTTTCGTCCTCGTTTCACTTTCGAGTTAGACTTTATTGCAGCATCTTGA
ACAATCGTCGCAGTTTGGTAACACGCTGTGCCATACTTTCATTTAGACGGAATCGAGGGACCCT
GGACTATAATCGCACAACGAGACCGGGTTGCGAAGTCAGGGCATTCCGCCGATCTAGCCATCG
CCATCTTCTGCGGGCGTTTGTTTGTTTGTTTGCTGGGATTAGCCAAGGGCTTGACTTGGAATCC
AATCCCGATCCCTAGCCCGATCCCAATCCCAATCCCAATCCCTTGTCCTTTTCATTAGAAAGTCA
TAAAAACACATAATAATGATGTCGAAGGGATTAGGGG

>*eve4+6*, DePace stock #0214
GGATCCCTGGGCTCTGGGCTCTGGACTATCCGCCGACCCTCCATATCCATGATTTACAATTCTC
GTTTTTTTCGCGTTATTTTTTTAGGGGCTTTAATGACCGTCGTAAAGCCGCAGGAGGACCAGGA
CCAGGACTCTGCTCACATTTCGCGCACTGATTCTAAAAAATGAAATCATTTTTTCTTGAATTTCA
CGGCGCGCCTCGAGCAGGACTCTTTGTTCTCGGCCAGGCAATTGTCCTTTTTTGCGCTCAGCTC
TCAGTTTTTTCGTCCAGCGGGCATTACCTACACGGCGTTTTATGGCGGAGATGATATTCGCCTG
GGATCGGTTCCGTTTTTTAGGCCATAAAAATTAGGCGGCATAAAAAAACTGCATTGGAATTCTA
GTTCTAGTTTCAAGTTTTTAGGTTTCCAGGTTTCTGCCAGCCCGCCTAGATTCGCATTTCGCGG
AATTCGGAAGCGGAACAGAATGCCAGAATGGTCAGAATCCTGGCTGACCTTGCCTTTTGGCCA
GGGGCCGTAAAAAAATTGACTCGCTGCGGTGCGCGGAATATTTTTTAAATCTGACTTTCCAACA
ATCTCTGATCTGGGTTCGAATCGTAAAAAAAAAGCAGAACAAAAAGCGGGCATTTTCGTCGGC
AAATGATCTGTTAATGGGCCGGGCTAAAAAACTAAGTCACAAAGTCACAAGGTTGTCCGGTAA
ATTGACCCGGTTAAGAATGTCTGTCTGTACCGAGAAGGATGCAGGACATTCAGCACTTCAAAG
CTCCCACCGCTCGAAGGATTCCCCCGAAGATTCAC

>*eve4+6mini*, DePace stock #0333
TCGAGCAGGACTCTTTGTTCTCGGCCAGGCAATTGTCCTTTTTTGCGCTCAGCTCTCAGTTTTTT
CGTCCAGCGGGCATTACCTACACGGCGTTTTATGGCGGAGATGATATTCGCCTGGGATCGGTT
CCGTTTTTTAGGCCATAAAAATTAGGCGGCATAAAAAAACTGCATTGGAATTCTAGTTCTAGTT
TCAAGTTTTTAGGTTTCCAGGTTTCTGCCAGCCCGCCTAGATTCGCATTTCGCGGAATTCGGAA
GCGGAACAGAATGCCAGAATGGTCAGAATCCTGGCTGACCTTGCCTTTTGGCCAGGGGCCGTA
AAAAAATTGACTCGCTGCGGTGCGCGGAATATTTTTTAAATCTGACTTTCCAACAATCTCTGAT
CTGGGTT

>*eve5*, DePace stock #0215
ATATCCCAAGGCCGCAAAGTCAACAAGTCGGCAGCAAATTTCCCTTTGTCCGGCGATGTGTTTT
TTTTTTAGCCATAACTCGCTGCATTGTTTGGGCCAAGTTTTTCTTCTGCCAAATTGCGGAGATG
ATGCGGGGATTATGCGCTGATTGCGTGCAATTATGGACATCCTGCGAGGCCCCGAGGAACTTC
CTGCTAAATCCTTTCATCCGCCTACAGAACCCCTTTGTGTCCCGTTCGCCGGGAGTCCTTGACG
GGTCCTTCGACTATTCGCTTACAGCAGCTTGCGTAAAATTTCATAACCCTACGAGCGGCTCTTC
CGCGGAATCCCTGGCATTATCCTTTTTACCTCTTGCCAATCCGTTGGCTAAAAAACGGCTTCGA
CTTCCGCGTAACTGCTGGACAACAAAGACAAAAAACGGCGAAAGGACGGCGATTTCCAGGTAG

CATTGCGAATTCCGTCAAACTAAAGGACCGGTTATATAACGGGTTTATATGGCCAGAATCTCTG
CATCTCCACGACCGCCAGAAGCTGCGTAAAACTGCAGGCTCTGTTTTGATTTCTGCAACTTCAG
TTAATTGCCCGGGATGGCCAGCAATTGCCGGCAATTATAAAACAGCGCAGATGTGACTCAGCT
TCCATATCTAACTCTATATCTCATGCCGAAAATCTAGGGTGGGGAGCGGAGGGGCGGGGTGCG
TGGGTGACTTGCCTGCCAGGGAAAGGGGGCGGGGGTTCAGCGGGTGATAAATGTGCGTGATT
TGGAATGAATGCGCATCGATTAAAACCGCAGGGCAATCAATTT

References

Driever, W. & Nüsslein-Volhard, C., 1989. The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. Nature, 337(6203), pp.138–143.

Little, S.C., Tikhonov, M. & Gregor, T., 2013. Precise developmental gene expression arises from globally stochastic transcriptional activity. Cell, 154(4), pp.789–800.

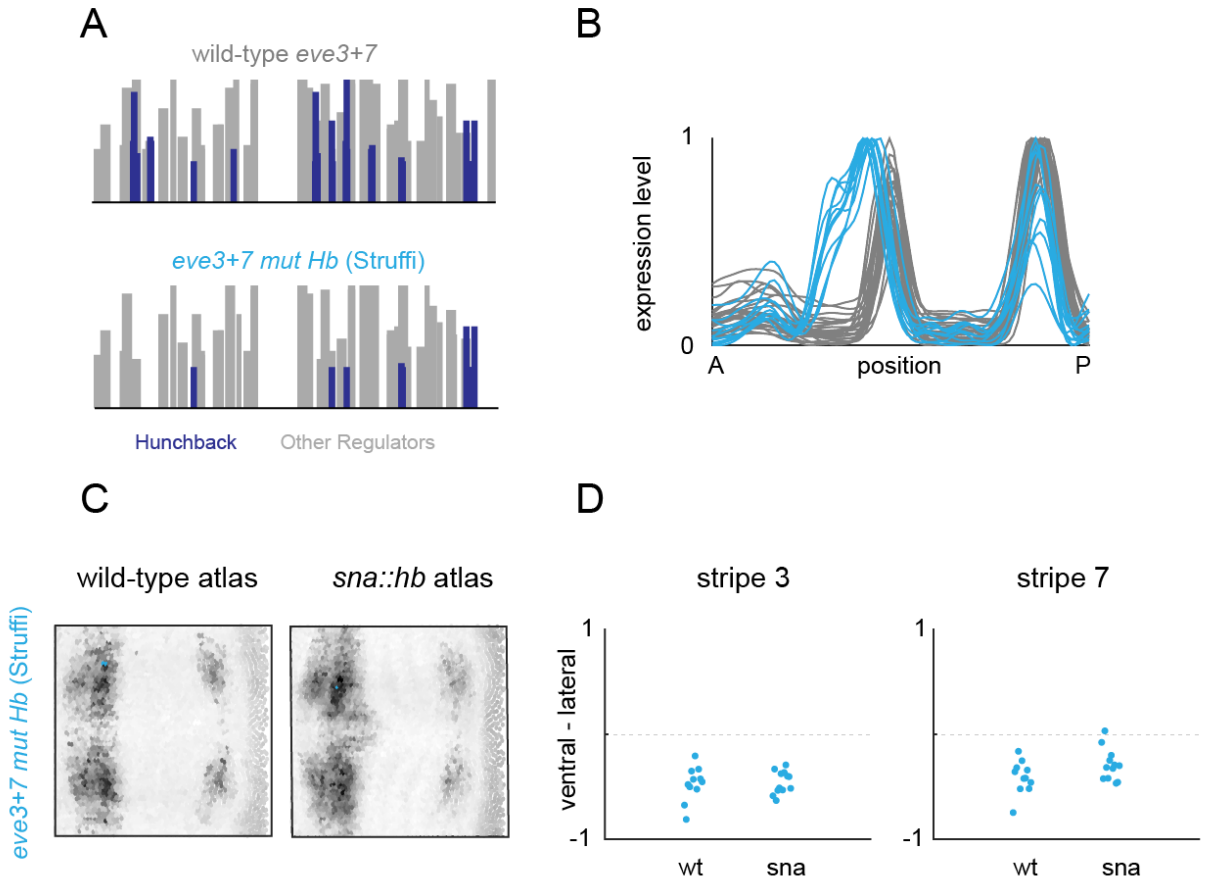
Noyes, M.B. et al., 2008. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nucleic acids research, 36(8), pp.2547–2560.

Staller, M.V. et al., 2015. A gene expression atlas of a bicoid-depleted Drosophila embryo reveals early canalization of cell fate. Development, 142(3), pp.587–596.

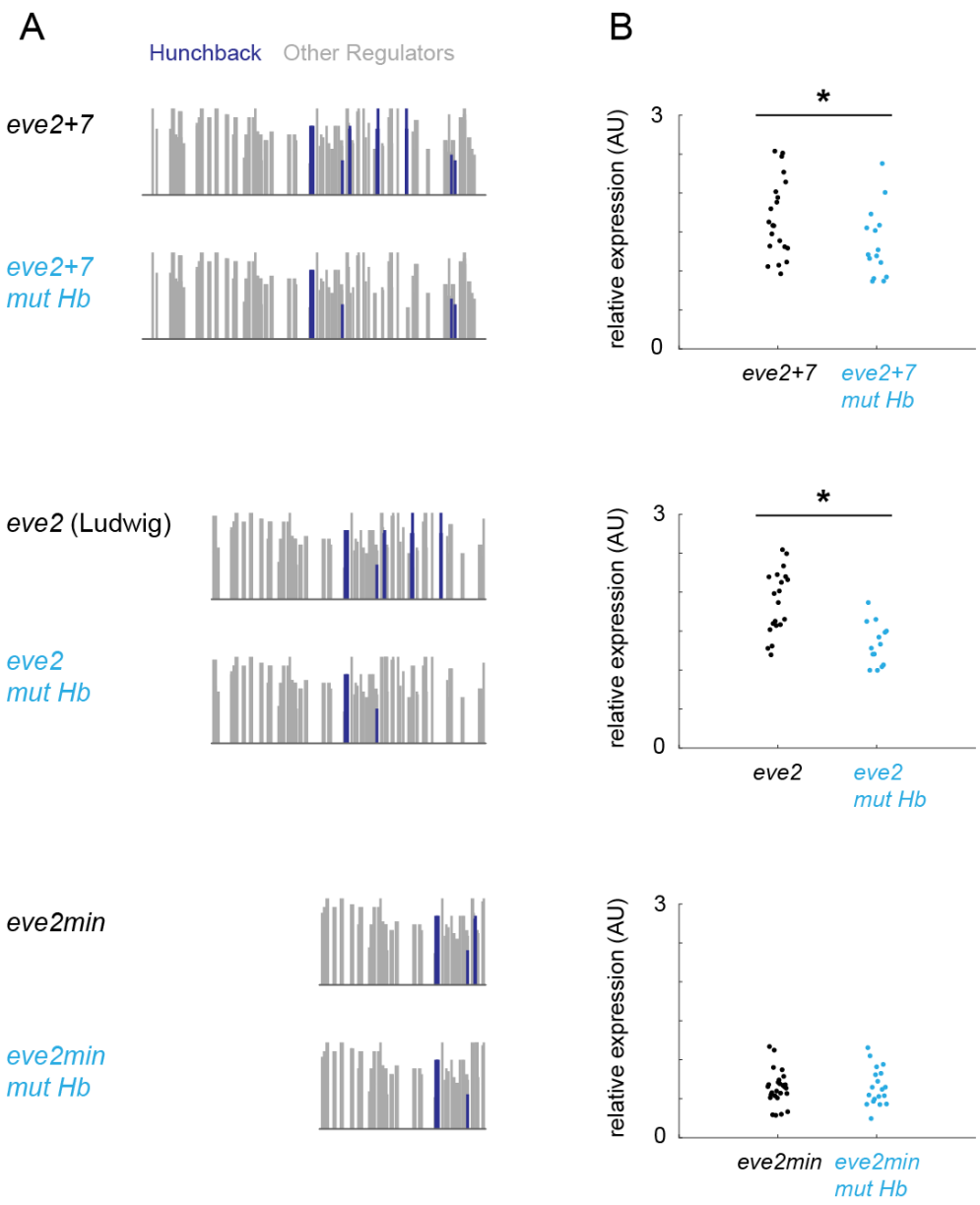
Tautz, D., 1988. Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres. Nature, 332(6161), pp.281–284.

APPENDIX C: SUPPLEMENTAL MATERIAL FOR CHAPTER 3

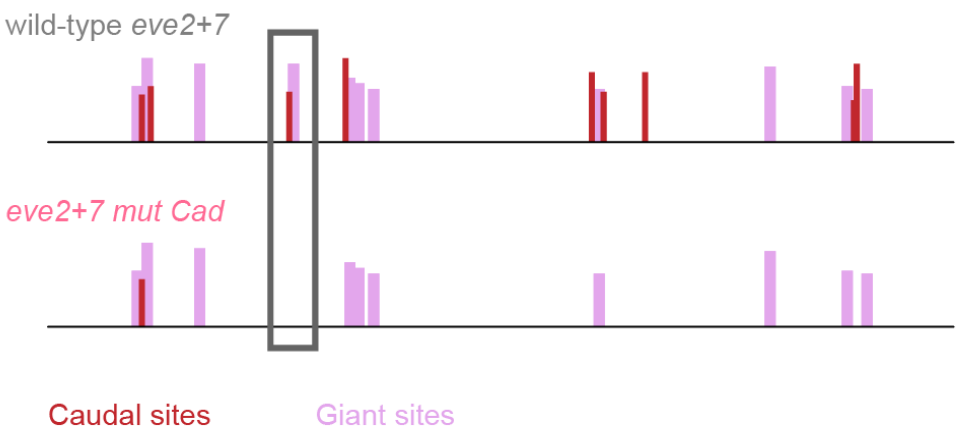
Supplemental Figures



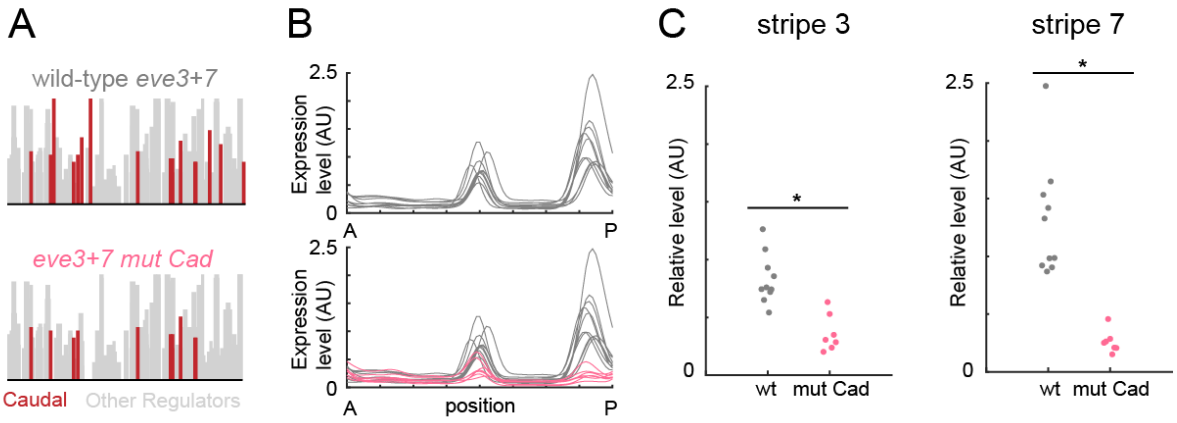
Supplemental Figure 3.1: Quantitative effects of *eve3+7* mutations designed by Struffi et al. (A) Predicted Hb binding sites (cobalt blue) in *eve3+7* and *eve3+7 mut Hb* plotted as in Figure 3.2. Sequence from Struffi et al. 2011. (B) Lateral line traces from individual WT embryos containing *eve3+7* reporter constructs (WT: grey, n= 26; m*ut Hb*: blue, n = 11). Each trace is normalized to its maximum value. Embryos are from all six time points in stage 5. (C) 2D projections of atlas data for reporter constructs expressed in WT or *sna::hb* embryos. Data is taken from time point 4. (D) Differences in the maximum values of ventral and lateral line traces are plotted for individual WT and *sna::hb* embryos containing *eve3+7 mut Hb i*n all 6 time points in stage 5. wt: n = 11; *sna::hb*: n = 12. Differences between WT and *sna::hb* embryos were not significant (Mann-Whitney U test, p-value > 0.1 for both stripes).



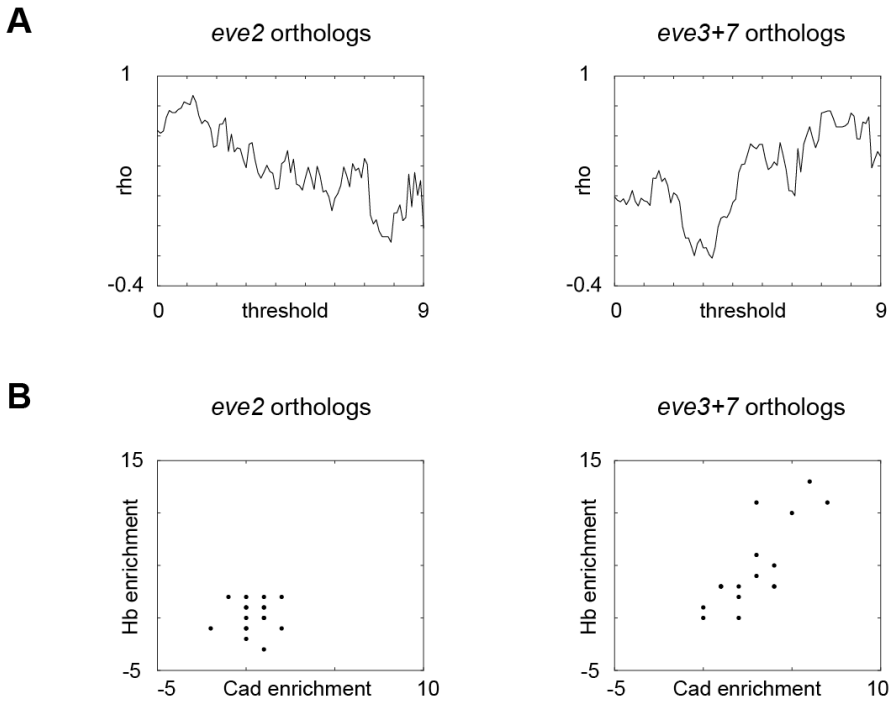
Supplemental Figure 3.2: Hunchback mutations have different effects depending on the enhancer context. (A) Predicted binding sites for Hb (cobalt blue) and other *eve2+7* regulators (grey) in different stripe 2 enhancer constructs (Staller et al. 2015; Ludwig et al. 1998; Small et al. 1992). (B) Peak stripe 2 expression levels for individual embryos from time points 2-4 were measured by normalizing *LacZ* expression levels using a *huckebein* co-stain (Wunderlich et al. 2014). Significance was assessed using the Mann-Whitney U test comparing wild-type enhancers with their counterparts with mutated Hb sites; asterisks indicate p-values < 0.05. Note that because each experiment was performed in separate hybridizations, comparisons cannot be made between different stripe 2 enhancers.



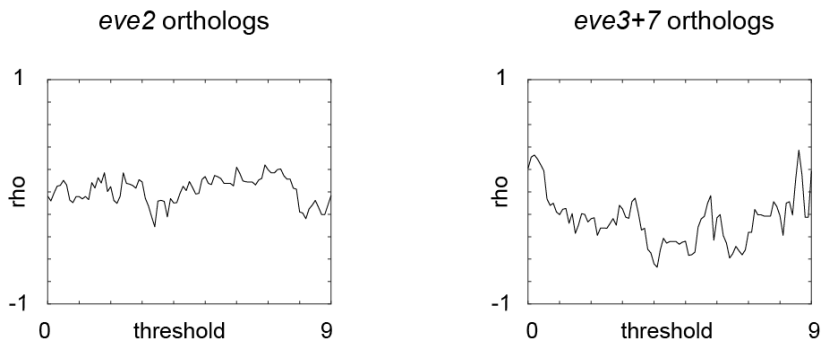
Supplemental Figure 3.3: Cad mutations may have disrupted one or more Giant binding sites. Predicted binding for Cad (carmine) and Gt (lilac) are shown in *eve2+7* and *eve2+7 mut Cad*. Many predicted Giant binding sites are near predicted Cad sites. One Cad binding site mutation in *eve2+7 mut Cad* (grey box) disrupts a predicted Giant binding site that also overlaps an annotated Giant binding site (Gt-2, Small et al. 1992). The Giant position weight matrix used from this analysis was published in Schroeder et al. 2011.



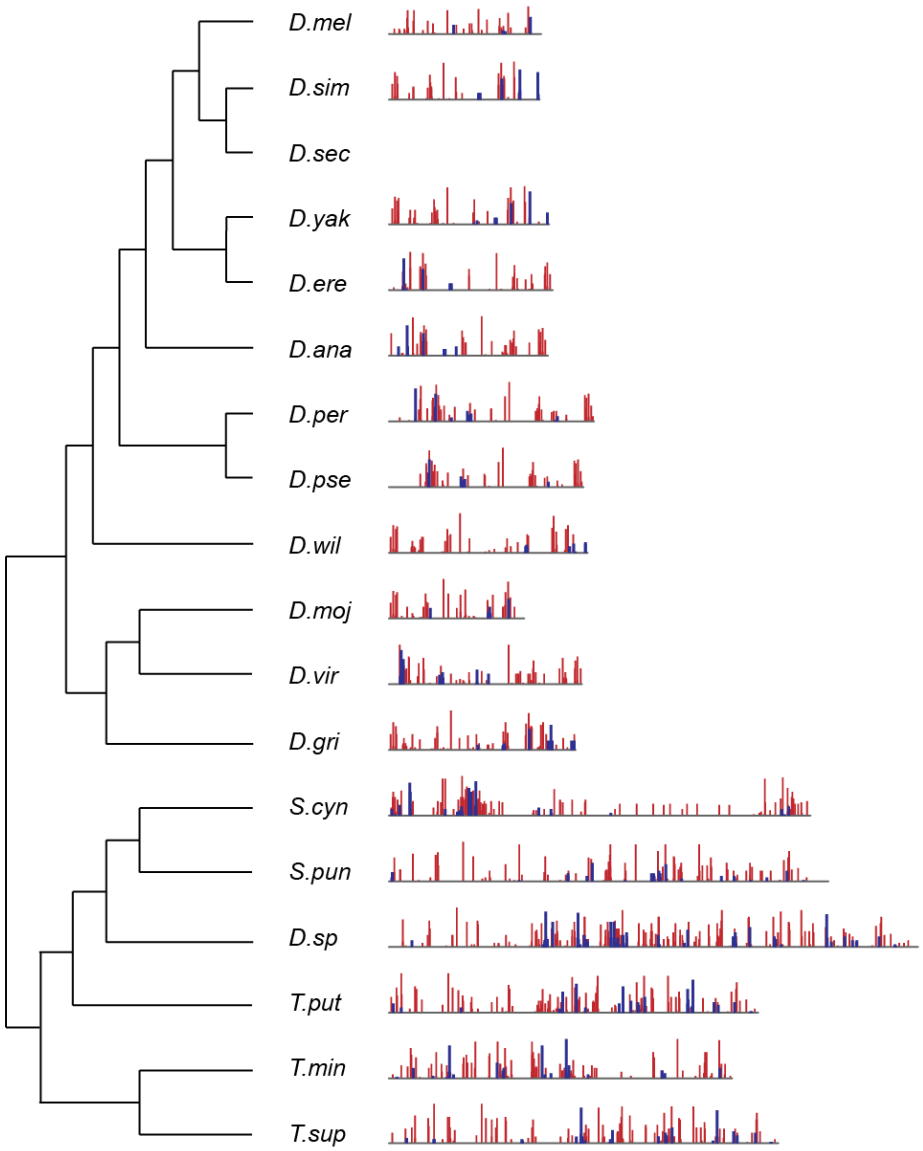
Supplemental Figure 3.4: Caudal directly activates *eve3+7*. (A) Predicted Cad binding sites in *eve3+7* and *eve3+7 mut Cad*. Sites were predicted and displayed as described in previous figures. (B) Lateral line traces from individual wild-type embryos containing reporter constructs for *eve3+7* (grey, n = 11) and *eve3+7 mut Hb* (carmine, n = 7). Traces were normalized using a *huckebein* co-stain; embryos are from time points 3 and 4. (C) Individual stripe peaks were found by taking local maxima from line traces in B. Asterisks indicate significant differences in stripe level (p-values < 0.001, Mann-Whitney U test).



Supplemental Figure 3.5: Sensitivity analyses for Caudal and Hunchback enrichment correlations. (A) Spearman correlation for Cad and Hb binding site enrichment is plotted as a function of binding site threshold for *eve2* and *eve3+7* orthologs. Binding site threshold refers to the minimum PATSER score for a predicted site to be counted in the analysis. Higher PATSER scores are assumed to reflect higher affinity sites. (B) Hb and Cad enrichment values are plotted for individual *eve2* and *eve3+7* orthologs at a binding site threshold of 7.2 – the threshold that maximizes rho in *eve3+7* orthologs.



Supplemental Figure 3.6: Sensitivity analyses for Bicoid and Hunchback enrichment correlations. (A) Spearman correlation for Bcd and Hb binding site enrichment is plotted as a function of binding site threshold for *eve2* and *eve3+7* orthologs. Enrichment for Bcd and Hb sites are not significantly correlated at any binding site threshold in either *eve2* or *eve3+7* orthologs.



Supplemental Figure 3.7: Predicted Hb and Cad binding sites in *eve2* orthologs. Sites predicted and displayed as in other figures. Cad sites are displayed in carmine, Hb sites are displayed in cobalt blue. Note that sequence data could not be extracted for *Drosophila sechellia* due to poor genomic quality.

Enhancer Sequences

>*eve3+7*, DePace stock #0204
GGATCCTCGAAATCGAGAGCGACCTCGCTGCATTAGAAAACTAGATCAGTTTTTTGTTTTGGCC
GACCGATTTTTGTGCCCGGTGCTCTCTTTACGGTTTATGGCCGCGTTCCCATTTCCCAGCTTCTT
TGTTCCGGGCTCAGAAATCTGTATGGAATTATGGTATATGCAGATTTTTATGGGTCCCGGCGAT
CCGGTTCGCGGAACGGGAGTGTCCTGCCGCGAGAGGTCCTCGCCGGCGATCCTTGTCGCCCGT
ATTAGGAAAGTAGATCACGTTTTTTGTTCCCATTGTGCGCTTTTTTCGCTGCGCTAGTTTTTTTC
CCCGAACCCAGCGAACTGCTCTAATTTTTTAATTCTTCACGGCTTTTCATTGGGCTCCTGGAAA
AACGCGGACAAGGTTATAACGCTCTACTTACCTGCAATTGTGGCCATAACTCGCACTGCTCTCG
TTTTTAAGATCCGTTTGTTTGTGTTTGTTTGTCCGCGATGGCATTCACGTTTTTACGAGCTC

>*eve3+7 mut Hb*, DePace stock #0506
CCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGA
ATTCGTTTAAACGGCCGGCCAGATCCAGGTCGCAGCGGCCGGATCCTCGAAATCGAGAGCGAC
CTCGCTGCATTAGAAAACTAGATCAGTTccTTGTTTTGGCCGACCGATccTTGTGCCCGGTGCTC
TCTTTACGGTTTATGGCCGCGTTCCCATTTCCCAGCTTCTTTGTTCCGGGCTCAGAAATCTGTAT
GGAATTATGGTATATGCAGAccTTTATGGGTCCCGGCGATCCGGTTCGCGGAACGGGAGTGTCC
TGCCGCGAGAGGTCCTCGCCGGCGATCCTTGTCGCCCGTATTAGGAAAGTAGATCACGTTccTT
GTTCCCATTGTGCGCTTccTTCGCTGCGCTAGTTccTTTCCCCGAACCCAGCGAACTGCTCTAATc
ggTcAATTCTTCACGGCTTTTCATTGGGCTCCTGGAAAAACGCGGACAAGGTTATAACGCTCTAC
TTACCTGCAATTGTGGCCATAACTCGCACTGCTCTCGTTTTTAAGATCCGTTTGTTTGTGTTTGT
TTGTCCGCGATGGCATTCACGTTTTTACGAGCTCGATCTAATCACTAGTGAATTCGAGCTCGCC
CGGGGATCGAGCGCAGCGGTATAAAAGGGCGCGGGGTGGCTGAGAGCAGCACACTCGAGCTG
TGACCGCC

>*eve3+7 mut Hb* (Struffi construct), DePace stock #0507
CCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGA
ATTCGTTTAAACGGCCGGCCAGATCCAGGTCGCAGCGGCCGGATCCTCGAAATCGAGAGCGAC
CTCGCTGCATTAGAAAACTAGATCAGTcTTcTaTTTTGGCCGACCGATTcTTaTaCCCGGTGCTCT
CTTTACGGTTTATGGCCGCGTTCCCATTTCCCAGCTTCTTTGTTCCGGGCTCAGAAATCTGTAT
GGAATTATGGTATATGCAGATTcTTATaGGTCCCGGCGATCCGGTTCGCGGAACGGGAGTGTCC
TGCCGCGAGAGGTCCTCGCCGGCGATCCTTGTCGCCCGTATTAGGAAAGTAGATCACGTTTTcc
GTTCCCATTGTGCGCTTcTTTCaCTGCGCTAGTTccTTTCCCCGAACCCAGCGAACTGCTCTAATT
TccTAATTCTTCACGGCTTTTCATTGGGCTCCTGGAAAAACGCGGACAAGGTTATAACGCTCTAC
TTACCTGCAATTGTGGCCATAACTCGCACTGCTCTCGTTcTcAAGATCCGTTTGTTTGTGTTTGT
TTGTCCGCGATGGCATTCACGTTcTTACaAGCTCGATCTAATCACTAGTGAATTCGAGCTCGCCC
GGGGATCGAGCGCAGCGGTATAAAAGGGCGCGGGGTGGCTGAGAGCAGCACACTCGAGCTGT
GACCGCC

>*eve3+7 mut Cad*, DePace stock #0472
GGATCCTCGAAATCGAGAGCGACCTCGCTGCATTAGAAAACTAGATCAGTTTTTTGTTTTGGCC
GACCGATTTTTGTGCCCGGTGCTCTCTTTACGGTgTcgGGCCGCGTTCCCATTTCCCAGCTTCTT
TGTTCCGGGCTCAGAAATCTGTATGGAAgTcgGGTATATGCAGATTTgTcgGGGTCCCGGCGATC

CGGTTCGCGGAACGGGAGTGTCCTGCCGCGAGAGGTCCTCGCCGGCGATCCTTGTCGCCCGTA
TTAGGAAAGTAGATCACGTTTTTTGTTCCCATTGTGCGCTTTTTTCGCTGCGCTAGTTTTTTTCC
CCGAACCCAGCGAACTGCTCTAATTTTTTAATTCTTCACGGCTTTTCATTGGGCTCCTGGAAAA
ACGCGGACAAGGTTATAACGCTCTACTTACCTGCAATTGTGGCCcgAcCTCGCACTGCTCTCGTT
TgTcgGATCCGTTTGTTTGTGTTTGTTTGTCCGCGATGGCATTCACGTTTcTgCGAGCTC

>*eve2+7*, DePace stock #0547
agaaggcttgcatgtgggcctttccaggtcggccagtaggtagagttgttgcgatgcggctatgccgggcgagttaatgccaatgcaaattg
cgggcgcaatataacccaataatttgaagtaactggcaggagcgaggtatccttcctggttacccggtactgcataacaatggaacccgaac
cgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaatccgtttgccatcagcgagattattagtcaattgcagt
tgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttgaacaatcgtcgcagtttggtaacacgctgtgccatac
tttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcgaagtcagggcattccgccgatctagccatcgc
catcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaatccaatcccgatccctagcccgatcccaatcccaat
cccaatcccttgtccttttcattagaaagtcataaaaacacataataatgatgtcgaagggattaggggcgcgcaggtccaggcaacgcaatt
aacggactagcgaactgggttatttttttgcgccgacttagccctgatccgcgagcttaacccgttttgagccgggcagcaggtagttgtgggt
ggaccccacgatttttttggccaaacctccaagctaacttgcgcaagtggcaagtggccggtttgctggcccaaaagaggaggcactatccc
ggtcctggtacagttggtacgctgggaatgattatatcatcataataaatgttttgcccaacgaaaccgaaaactttcaaattaagtcccggc
aactgggttcccattttccattttccatgttctgcgggcaggggcggccattatctcgct

>*eve2+7 mut Hb*, DePace stock #0535
agaaggcttgcatgtgggcctttccaggtcggccagtaggtagagttgttgcgatgcggctatgccgggcgagttaatgccaatgcaaattg
cgggcgcaatataacccaataatttgaagtaactggcaggagcgaggtatccttcctggttacccggtactgcataacaatggaacccgaac
cgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaatccgtttgccatcagcgagattattagtcaattgcagt
tgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttgaacaatcgtcgcagtttggtaacacgctgtgccatac
tttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcgaagtcagggcattccgccgatctagccatcgc
catcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaatccaatcccgatccctagcccgatcccaatcccaat
cccaatcccttgtccttttcattagaaagtcataaaGGcacataataatgatgtcgaagggattaggggcgcgcaggtccaggcaacgcaat
taacggactagcgaactgggttattCCtttgcgccgacttagccctgatccgcgagcttaacccgttttgagccgggcagcaggtagttgtgg
gtggaccccacgattCCtttggccaaacctccaagctaacttgcgcaagtggcaagtggccggtttgctggcccaaaagaggaggcactat
cccggtcctggtacagttggtacgctgggaatgattatatcatcataataaatgttttgcccaacgaaaccgaaaactttcaaattaagtccc
ggcaactgggttcccattttccattttccatgttctgcgggcaggggcggccattatctcgct

>*eve2+7 mut Cad*, DePace stock #0549
agaaggcttgcatgtgggcctttccaggtcggccagtaggtagagttgttgcgatgcggctatgccgggcgagttaatgccaatgcaaattg
cgggcgcaatataacccaataGGttgaagtaactggcaggagcgaggtatccttcctggttacccggtactgcataacaatggaacccgaa
ccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaatccgtttgccatcagcgagattatATgtcaattgca
gttgcagcgtttcgctttcgtcctcgtttcactttcgagttagacGCtattgcagcatcttgaacaatcgtcgcagtttggtaacacgctgtgcc
atactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcgaagtcagggcattccgccgatctagcca
tcgccatcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaatccaatcccgatccctagcccgatcccaatcc
caatcccaatcccttgtccttttcattagaaagtcGCaaaaacacataataGGgatgtcgaagggattaggggcgcgcaggtccaggcaa
cgcaGCtaacggactagcgaactgggttatttttttgcgccgacttagccctgatccgcgagcttaacccgttttgagccgggcagcaggtag
ttgtgggtggaccccacgatttttttggccaaacctccaagctaacttgcgcaagtggcaagtggccggtttgctggcccaaaagaggaggca
ctatcccggtcctggtacagttggtacgctgggaatgattatatcatcataGCaaatgttttgcccaacgaaaccgaaaactttcaaattaa
gtcccggcaactgggttcccattttccattttccatgttctgcgggcaggggcggccattatctcgct

>*eve2+7 mut Cad and Hb*, DePace stock #0634
agaaggcttgcatgtgggccttttccaggtcggccagtaggtagagttgttgcgatgcggctatgccgggcgagttaatgccaatgcaaattg
cgggcgcaatataacccaataGGttgaagtaactggcaggagcgaggtatccttcctggttacccggtactgcataacaatggaacccgaa
ccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaatccgtttgccatcagcgagattatATgtcaattgca
gttgcagcgtttcgctttcgtcctcgtttcactttcgagttagacGCtattgcagcatcttgaacaatcgtcgcagtttggtaacacgctgtgcc
atactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcgaagtcagggcattccgccgatctagcca
tcgccatcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaatccaatcccgatccctagcccgatcccaatcc
caatcccaatcccttgtccttttcattagaaagtcGCaaaGGcacataataGGgatgtcgaagggattaggggcgcgcaggtccaggca
acgcaGCtaacggactagcgaactgggttattCCtttgcgccgacttagccctgatccgcgagcttaacccgttttgagccgggcagcagg
tagttgtgggtggaccccacgattCCtttggccaaacctccaagctaacttgcgcaagtggcaagtggccggtttgctggcccaaaagagga
ggcactatcccggtcctggtacagttggtacgctgggaatgattatatcatcataGCaaatgttttgcccaacgaaaccgaaaacttttcaaa
ttaagtcccggcaactgggttcccattttccattttccatgttctgcgggcaggggcggccattatctcgct

>*eve2* (Ludwig), DePace stock #0464
aatataacccaataatttgaagtaactggcaggagcgaggtatccttcctggttacccggtactgcataacaatggaacccgaaccgtaactg
ggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaatccgtttgccatcagcgagattattagtcaattgcagttgcagcgt
ttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttgaacaatcgtcgcagtttggtaacacgctgtgccatactttcattta
gacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcgaagtcagggcattccgccgatctagccatcgccatcttctg
cgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaatccaatcccgatccctagcccgatcccaatcccaatcccaatcc
cttgtccttttcattagaaagtcataaaaacacataataatgatgtcgaagggattaggggcgcgcaggtccaggcaacgcaattaacggac
tagcgaactgggttatttttttgcgccgacttagccctgatccgcgagcttaacccgttttgagccgggcagcaggtagttgtgggtggacccc
acgatttttttggccaaacctccaagctaacttgcgcaagtggcaagtggccggtttgctggcccaaaagaggaggcactatcccggtcctgg
tacagttggtacgctgggaatgattatatcatcataataaatgttt

>*eve2* (Ludwig) *mut Hb*, DePace stock #0496
aatataacccaataatttgaagtaactggcaggagcgaggtatccttcctggttacccggtactgcataacaatggaacccgaaccgtaactg
ggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaatccgtttgccatcagcgagattattagtcaattgcagttgcagcgt
ttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttgaacaatcgtcgcagtttggtaacacgctgtgccatactttcattta
gacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcgaagtcagggcattccgccgatctagccatcgccatcttctg
cgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaatccaatcccgatccctagcccgatcccaatcccaatcccaatcc
cttgtccttttcattagaaagtcaGCaTCTcacataataatgatgtcgaagggattaggggcgcgcaggtccaggcaacgcaattaacgg
actagcgaactgggttatGCAtCtgcgccgacttagccctgatccgcgagcttaacccgttttgagccgggcagcaggtagttgtgggtgg
accccacgatGCAtCtggccaaacctccaagctaacttgcgcaagtggcaagtggccggtttgctggcccaaaagaggaggcactatccc
ggtcctggtacagttggtacgctgggaatgattatatcatcataataaatgttt

>*eve2min*, DePace stock #0216
ggttacccggtactgcataacaatggaacccgaaccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaat
ccgtttgccatcagcgagattattagtcaattgcagttgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttga
acaatcgtcgcagtttggtaacacgctgtgccatactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggtt
gcgaagtcagggcattccgccgatctagccatcgccatcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaa
tccaatcccgatccctagcccgatcccaatcccaatcccaatcccttgtccttttcattagaaagtcataaaaacacataataatgatgtcgaa
gggattagggg

>*eve2min mut Hb*, DePace stock #0463
ggttacccggtactgcataacaatggaacccgaaccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaat
ccgtttgccatcagcgagattattagtcaattgcagttgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttga
acaatcgtcgcagtttggtaacacgctgtgccatactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggtt
gcgaagtcagggcattccgccgatctagccatcgccatcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaa
tccaatcccgatccctagcccgatcccaatcccaatcccaatcccttgtccttttcattagaaagtcaGCaTCTcacataataatgatgtcga
agggattagggg

References

Ludwig, M.Z., Patel, N.H. & Kreitman, M., 1998. Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development, 125(5), pp.949–958.

Schroeder, M.D., Greer, C. & Gaul, U., 2011. How to make stripes: deciphering the transition from non-periodic to periodic patterns in Drosophila segmentation. Development, 138(14), pp.3067–3078.

Small, S., Blair, A. & Levine, M., 1992. Regulation of even-skipped stripe 2 in the Drosophila embryo. The EMBO journal, 11(11), pp.4047–4057.

Staller, M.V. et al., 2015. Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 112(3), pp.785–790.

Struffi, P. et al., 2011. Combinatorial activation and concentration-dependent repression of the Drosophila even skipped stripe 3+7 enhancer. Development, 138(19), pp.4291–4299.

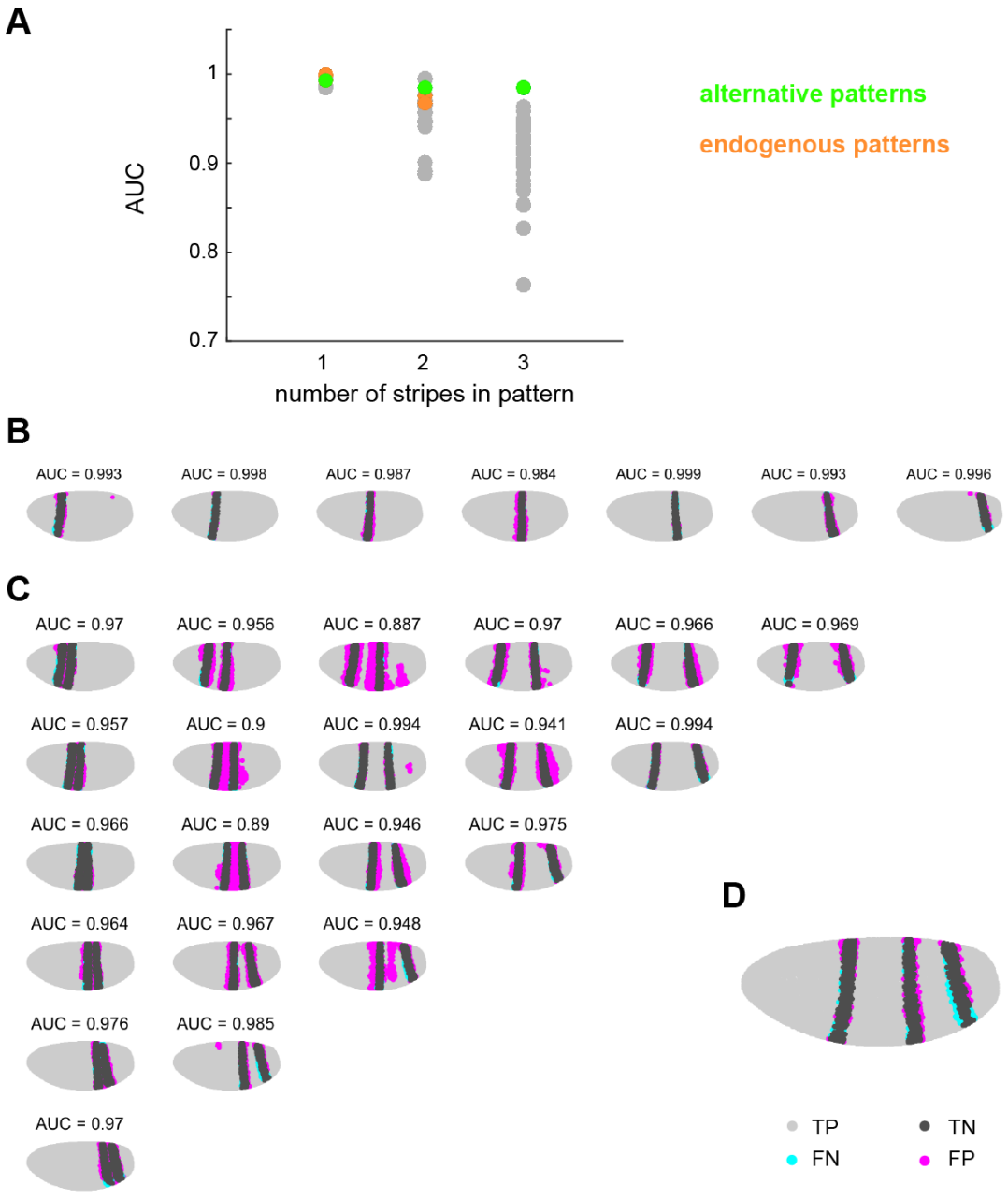
Wunderlich, Z., Bragdon, M.D. & DePace, A.H., 2014. Comparing mRNA levels using in situ hybridization of a target gene and co-stain. Methods, 68(1), pp.233–241.

Wunderlich, Z. et al., 2015. Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. Cell reports, 12(11), pp.1740–1747.

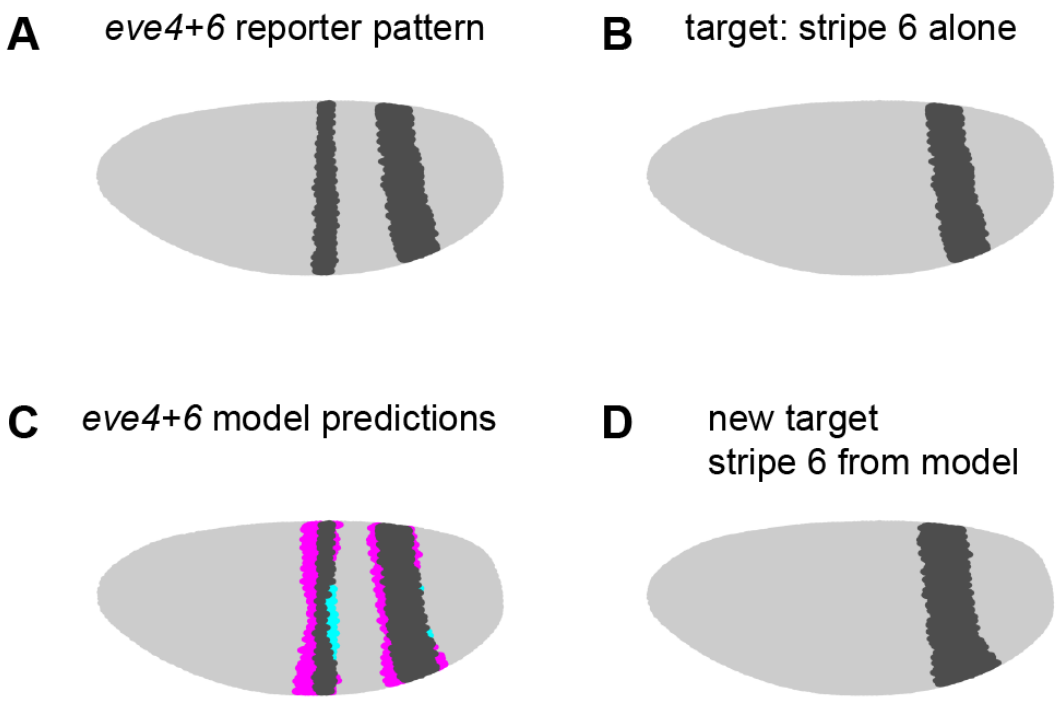
APPENDIX D: SUPPLEMENTAL MATERIAL FOR CHAPTER 4

Supplemental Figures

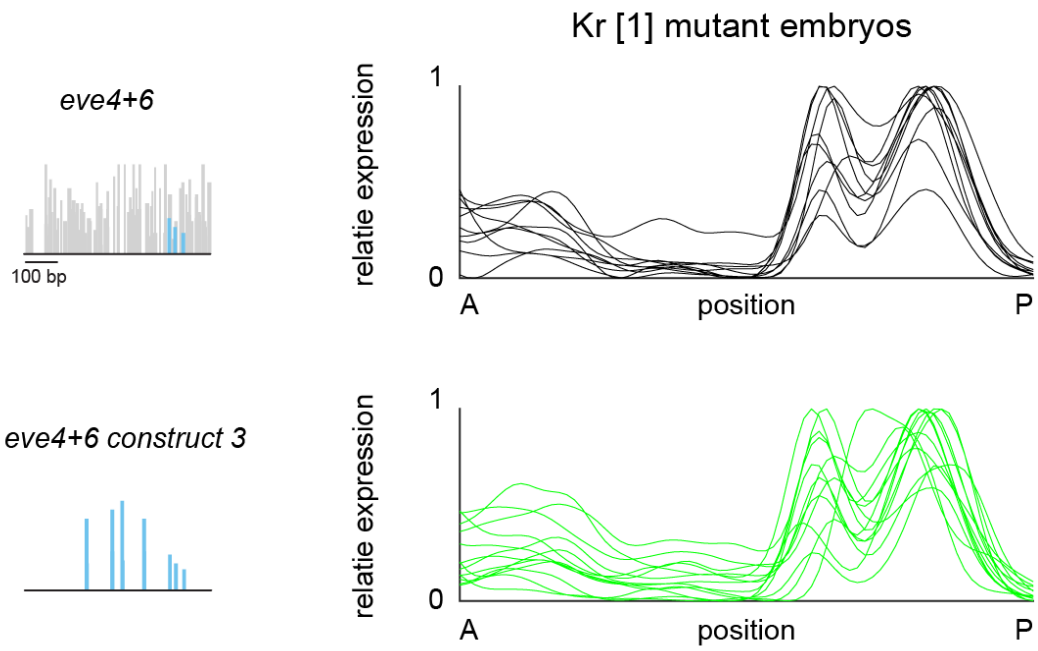
Supplemental Figure 4.1: Modeling predictions for all combinations of 1 and 2 *eve* stripes. (A) Model predictions for all single stripe patterns (stripe 1 through stripe 7 from left to right). Matthews correlation coefficient for predictions appears above each plot. Each nucleus in the virtual embryo is classified according to modeling predictions. True positives: dark grey; True negatives: light grey; False positives: magenta; False negatives: cyan. (B) Model predictions for all double stripe patterns (row 1 all include stripe 1, row 2 all include stripe 2, etc.). Plots displayed as in (A). Except for stripes 2, 5 and 7 (Figure 4.2), no combination of three or more stripes predicted expression in the same number of distinct stripes (data not shown).



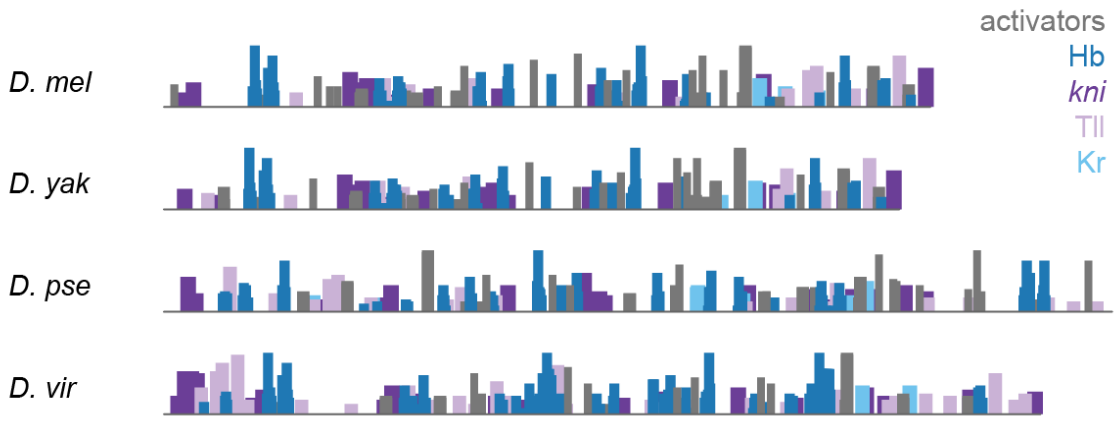
Supplemental Figure 4.2: Modeling predictions using the area under the receiver operating curve. (A) For each model, we plot the area under the receiver operating curve (AUC) as a function of the number of stripes in the pattern. Endogenous patterns are plotted in orange, alternative patterns are plotted in green, all other patterns are plotted in grey. (B) Model predictions for all single stripe patterns evaluated at thresholds that correspond to 80% sensitivity. Predictions are plotted as in Supplemental Figure 4.1. (C) Same as (B), for all double stripe patterns. (D) Same as (B) for the stripe 2, 5 and 7 combination. No other combination of three or more stripes predicted expression in the same number of distinct stripes (data not shown).



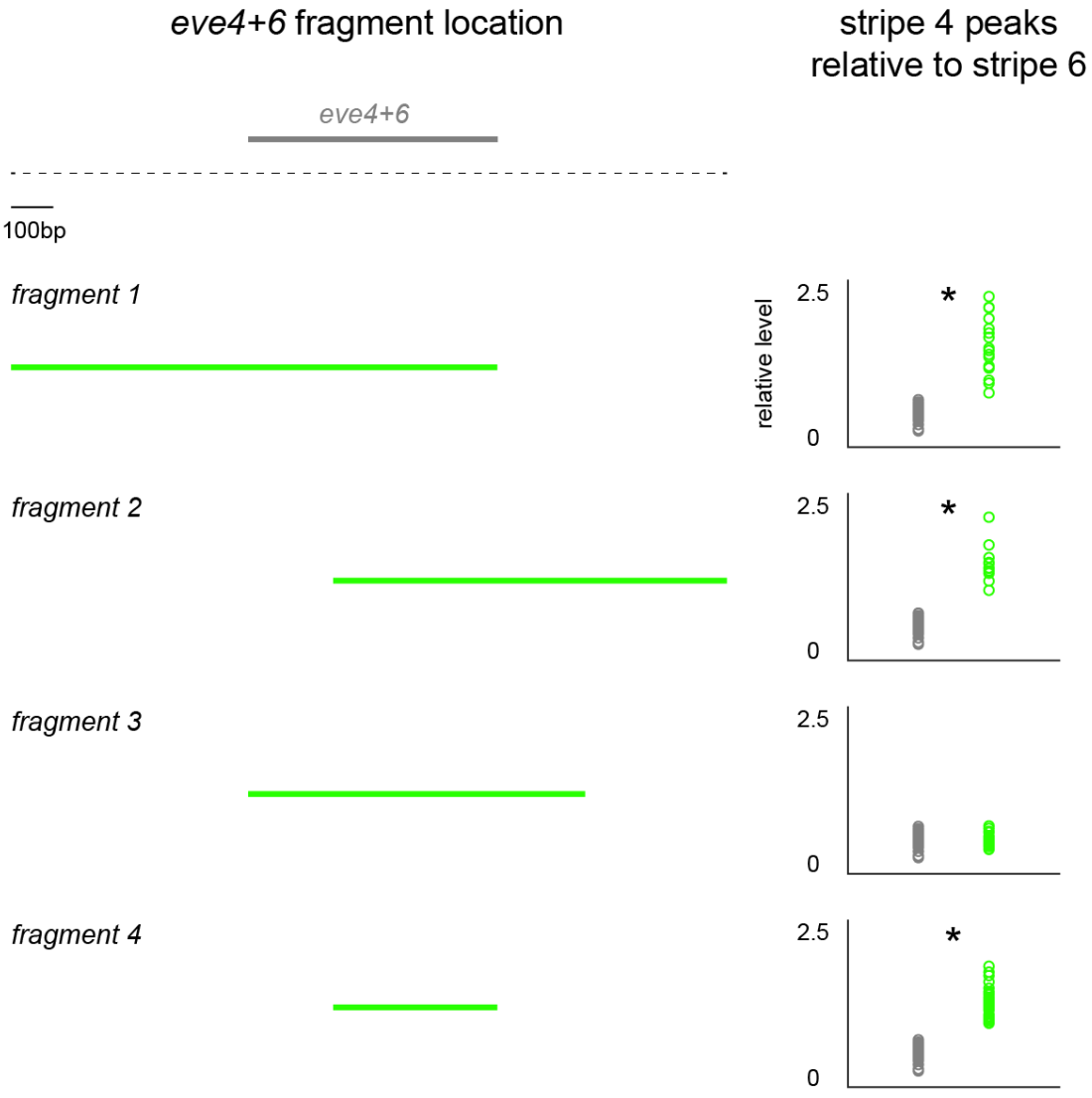
Supplemental Figure 4.3: Generating the stripe 6 target pattern for sensitivity analysis. Our goal is to generate stripe 6 alone by modifying *eve4+6*. (A) The *eve4+6* reporter pattern thresholded at a fixed value of 0.2. ON cells: dark grey; OFF cells: light grey. (B) a target pattern consisting of the stripe 6 portion of the reporter pattern. Note that this is not the target pattern used for evaluating in our sensitivity analysis. (C) Predictions from our *eve4+6* model. True positives: dark grey; true negatives: light grey; false positives: magenta; false negatives: cyan. Our model makes incorrect predictions for both stripes. If our goal is to predict how to turn off stripe 4 without affecting stripe 6, our target pattern should consist of all stripe 6 cells predicted to be ON by the *eve4+6* model. (D) the stripe 6 target used for sensitivity analysis. This pattern consists of true positives and false positives predicted by the *eve4+6* model in the area of stripe 6.



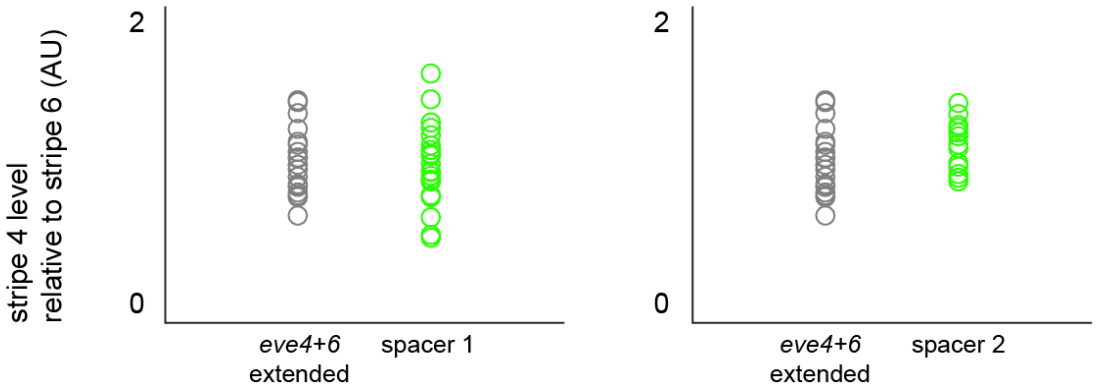
Supplemental Figure 4.4: *eve4+6* and *eve4+6 construct 3* drive similar expression patterns in *Krüppel* mutant embryos. Lateral line traces are shown for expression patterns driven by *eve4+6 (*black, top) and *eve4+6 construct 3 (*green, bottom) in Kr[1] mutant embryos. Line traces are normalized by the maximum value of the trace. While these embryos were co-stained with *hkb*, *LacZ* levels were uncorrelated with *hkb* levels in the hybridization. Therefore, comparisons to the *hkb* co-stain could not be made. Embryos are from all time points in stage 5. *eve4+6*, n = 10; *eve4+6 construct 3*, n = 13.



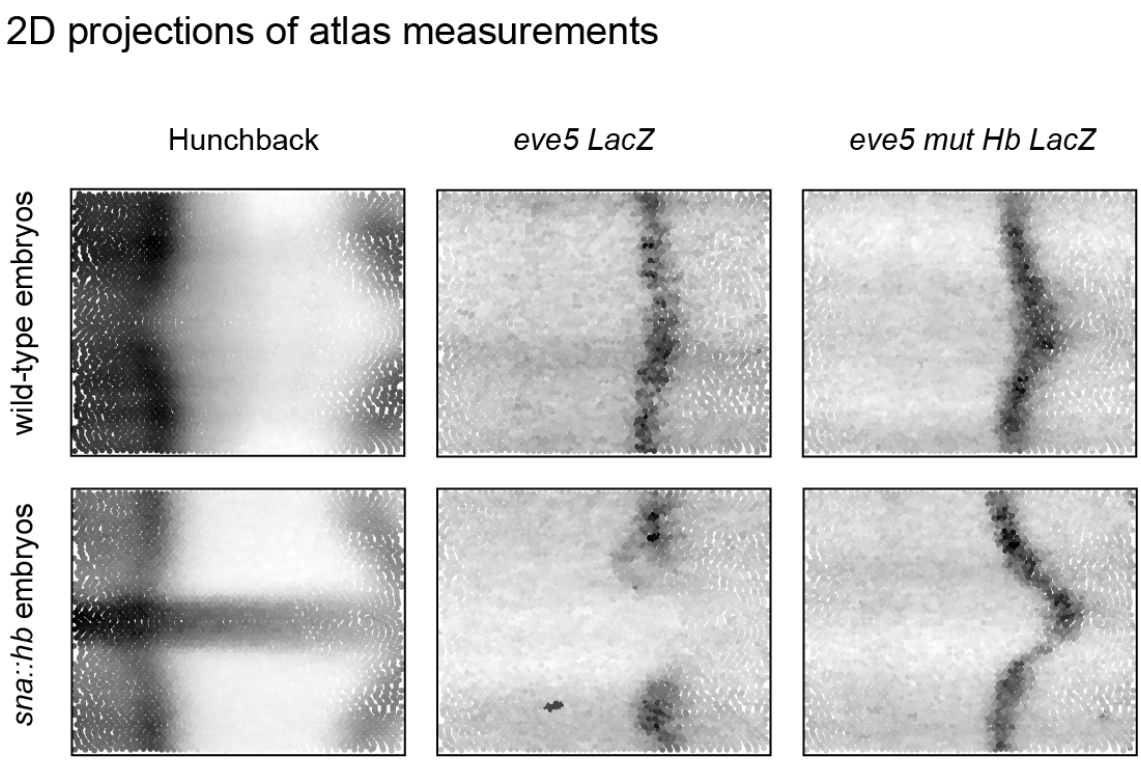
Supplemental Figure 4.5: Predicted binding sites in *eve4+6* orthologs. Binding sites for *eve4+6* regulators in *eve4+6* enhancers in *Drosophila melanogaster* (*D. mel*), *Drosophila yakuba* (*D. yak*), *Drosophila pseudoobscura* (*D. pse*) and *Drosophila virilis* (*D. vir*). Sites displayed as in Figure 4.1.



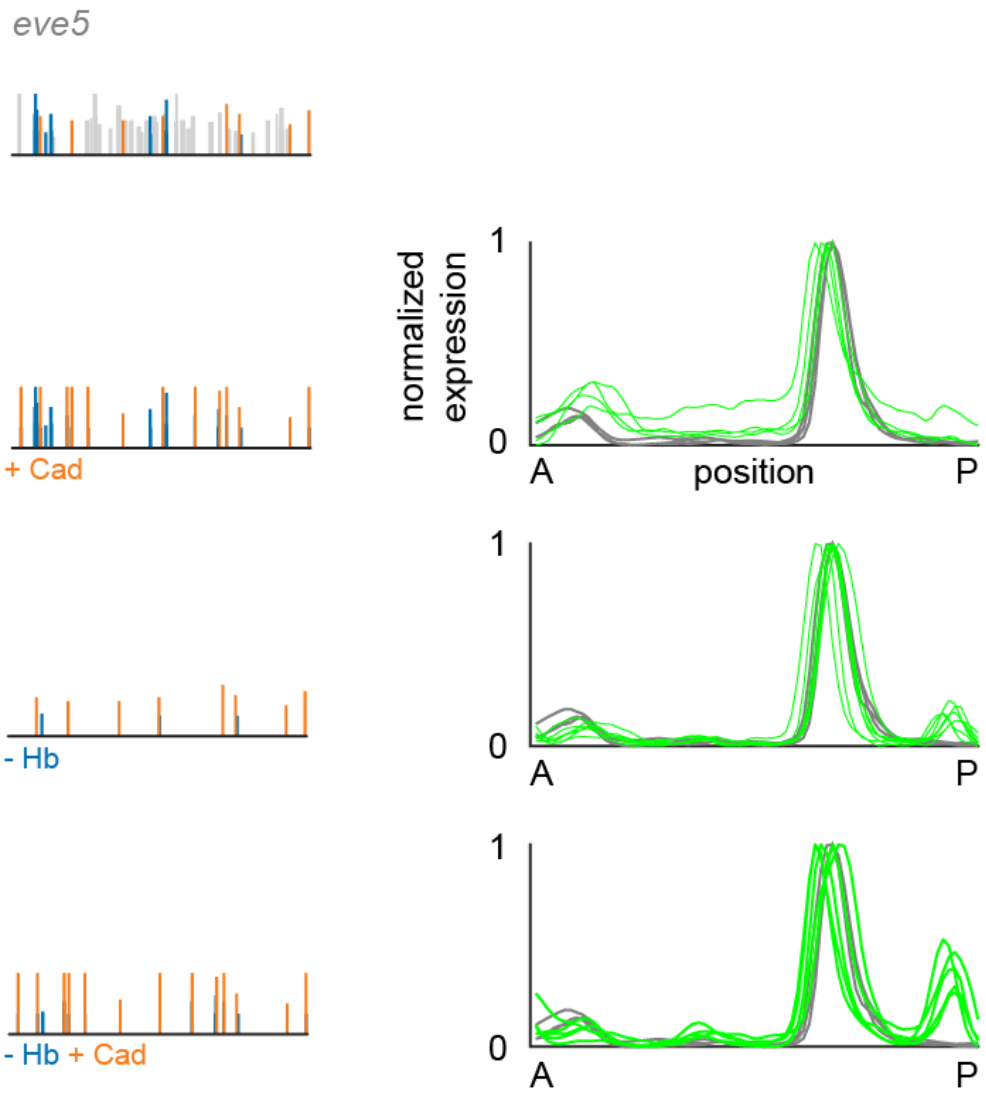
Supplemental Figure 4.6: Different *eve4+6* enhancer fragments drive different levels of stripe 4 relative to stripe 6. We measured expression levels of stripe 4 relative to stripe 6 in embryos containing reporter constructs with different fragments of the *eve4+6* enhancer region (Left). Relative stripe 4 levels driven by *eve4+6* (grey) were significantly different compared to other fragments (green). Asterisks indicate significant differences between *eve4+6* and other fragments (p-values < 10^{-6}, Mann-Whitney U test). Relative stripe 4 levels driven by *eve4+6* and *eve4+6 fragment 3* were not significantly different (p value > 0.8).



Supplemental Figure 4.7: *eve4+6 f*lanking sequence enriched in Kr binding sites does not affect stripe 4 expression relative to stripe 6 in *Drosophila melanogaster*. We measured expression driven by a reporter construct containing an extended fragment of *eve4+6*. We also measured expression driven by reporter constructs where sequence enriched in Kr binding sites was replaced with two different designed spacer sequence containing no binding sites for *eve4+6* regulators (Estrada et al. 2016). These spacer sequences should remove the influence of the flanking sequence while controlling for distance from the promoter. We measured expression levels of stripe 4 relative to stripe 6 in embryos containing these reporter constructs, but observed no significant effect of endogenous sequence replacement (p-values > 0.05, Mann-Whitney U test).



Supplemental Figure 4.8: Hunchback binding site mutations abolish stripe 5 retreat in embryos that express ventral *hunchback*. 2-dimensional projections of atlas data (see Figure 3.1) from WT embryos (top) and embryos expressing *hb* along the ventral surface of the embryo from the *snail* promoter (*sna::hb* embryos, bottom) in time point 5. Measurements were made for for Hb protein (left), *eve5 LacZ* mRNA (middle) and *eve5 mut Hb LacZ* mRNA (right) using immunofluorescence and *in situ* hybridization. *eve5* expression retreats from ventral Hb, but *eve5 mut Hb* does not.



Supplemental Figure 4.9: Caudal binding sites increase stripe 7 expression at late time points. Line traces from individual embryos containing reporter constructs for wild-type *eve5* (grey) or engineered *eve5* (green). Embryos were staged between 70-100% membrane invagination in stage 5. Line traces were background subtracted and normalized to the peak expression level of stripe 5. Stripe 7 peaks driven by *eve5 add Cad mut Hb* were significantly higher compared to peaks driven by *eve5 mut Hb* (p-value < 0.005, Mann-Whitney U test).

Enhancer Sequences

>*eve4+6 D. mel*, DePace stock #0332
aggatccctgggctctgggctctggactatccgccgaccctccatatccatgatttacaattctcgttttttcgcgttatttttttaggggctttaa
tgaccgtcgtaaagccgcaggaggaccaggaccaggactctgctcacatttcgcgcactgattctaaaaaatgaaatcatttttttcttgaattt
cacggcgcgcctcgagcaggactctttgttctcggccaggcaattgtccttttttgcgctcagctctcagttttttcgtccagcgggcattaccta
cacggcgttttatggcggagatgatattcgcctgggatcggttccgttttttaggccataaaaattaggcggcataaaaaaaactgcattggaat
tctagttctagtttcaagtttttaggtttccaggtttctgccagcccgcctagattcgcatttcgcggaattcggaagcggaacagaatgccaga
atggtcagaatcctggctgaccttgccttttggccaggggccgtaaaaaaaattgactcgctgcggtgcgcggaatatttttaaatctgactttc
caacaatctctgatctgggtt

>*eve4+6 construct 1*, DePace stock #0414
aggatccctgggctctgggctctggactatccgccgaccctccatatccatgatttacaattctcgttttttcgcgttatttttttaggggctttaa
tgaccgtcgtaaagccgcaggaggaccaggaccaggactctgctcacatttcgcgcactgattctaaaaaatgaaatcatttttttcttgaattt
AacggATTAActcgagcaggactctttgttctcggccaggcaattgtccttttttgcgctcagctctcagttttttcgtccagcgTgTTAta
cctacacggcgttttatggcggagaCgGAattcgcctgggatcggttccgttttttaggccataaaaattaggcggcataaaaaaaactgca
GGgTCCCtctagttctagtttcaagtttttaggtttccaggtttctgccagcccgcctagattcgcatttcgcggaattcggaagcggaacag
aatgccagaatggtcagaatcctggctgaccttgccttttggccaggggccgtaaaaaaaattgactcgctgcggtgcgcggaatatttttaaa
tctgactttccaacaatctctgatctgggtt

>*eve4+6 construct 2*, DePace stock #0413
aggatccctgggctctgggctctggactatccgTTAacccGGTatatccatgatttacaattctcgttttttcgcgttatttttttaggggctt
taTAACTGgGACAaaagccgcaggaggaccaggaccaggactctgctcacatttcgcgcactgattctaaaaaatgaaatcatttttt
cttgaatttAacggATTAActcgagcaggactctttgttctcggccaggcaattgtccttttttgcgctcagctctcagttttttcgtccagcg
TgTTAtacctacacggcgttttatggcggagaCgGAattcgcctgggatcggttccgttttttaggccataaaaattaggcggcataaaa
aaactgcaGGgTCCCtctagttctagtttcaagtttttaggtttccaggtttctgccagcccgcctagattcgcatttcgcggaattcggaag
cggaacagCaACccGgTatggtcagaatcctCTAATCccttCccttttggccaggggccgtaaaaaaattgactcgACCGggtAA
gcggaatatttttaTGtcCCaGttAccaacaatctctgatctgggtt

>*eve4+6 construct 3*, DePace stock #0412
aggatccctgggctctgggctctggactatccgccgaccctccatatccatgatttacaattctcgttttttcgcgttatttttttaggggctttaa
tgaccgtcgtaaagccgcaggaggaccaggaccaggactctgctcacatttcgcgcactgattctaaaaaatgaaatcatttttttcttgaattt
AaAggGTTAActcgagcaggactctttgttctcggccaggcaattgtccttttttgcgctcagctctcagttttttcgtccaAAgggTTA
AacctacacggcgttttatggcgAaAGGgTtaAtcgcctgggatcggttccgttttttaggccataaaaattaggcggcataaaaaaaact
gcaAAggGTtAAtagttctagtttcaagtttttaggtttccaggtttctgccagcccgcctagattcgcatttcgcggaattcggaagcgga
acagaatgccagaatggtcagaatcctggctgaccttgccttttggccaggggccgtaaaaaaattgactcgctgcggtgcgcggaatatttt
ttaaatctgactttccaacaatctctgatctgggtt

>*eve4+6 D. yak*, DePace stock #0440
aacccagatcaggattgttggaaagtcagatttaaaaaatattccgcgcaccgcagcgagtcaatttttttacggcccctggccaaaaggcaa
ggtcggccaggattgttccgcttccgaattccgcgaattgcgaatctttccgggctgtcggaaacctggaaacctaaaaacctgaaactagaa
ctagaattccaatgcagtttttttatgccgcctaatttttatggcctaaaaaacggaaccgatccctggcgaatatcatctccgccataaaacgc
cgtgtaggtaatgcccgctggccgaaaaaactgagagccgagcgcaaaaaaggacaactgcccggccgaaaacaaagagtcctgctcga
ggcgcgccgtgaaattcaagaaaaaatgatttcatttttagaatcagtgcgcgaaatgtgagcagagtgctggtcctagtcctcctgcggctt

tacgacggtcattaaagcccctaaaaaaataacgcgaaaaaaacgagaattgtaaatcatggaaatggagagtccagagtccggagtccg
gagccccgggttcct

>*eve4+6 D. pse*, DePace stock #0441
ggccaggatcagaatcgggatcgcgatcggcaaggaggctgcagaaaaagtcatttctaaaaaatatccaccgcaccgcagcgagtcaatt
tttttacggccctcagagaagaaggcggagaagaaggcaaggtcagccccggaatatcggatatttgccattctgttccgttccgcattccgc
attttgcttccgaattccgcgaatttcataaaaatctgcccaaagtccgaaaaactctggaaacctaaaaatctgaaactagaactagaattcc
aatgcagtttttttatgccgcctaatttttatggcctaaaaaacggaacagcaccgagagccgagccgagccgagcctcacagaatatcatct
ccgtcataaaaaccccggaggcctaggtagcactcgaagcgggggaaaaagaaacgatgccgagtgtgcaaaaaaggacgagcactggt
ccaggacaaagagcctttacacagtcgtctctgcgaaattcgagaaaaaatgatttcatttttagaatcagttcgcgaaatgggaaccacct
atgccacacctcctgctcgcactctcacatctgagggcaagagagagagggagaggagagagtagaggtctcaggtcctgcggctttacga
tggccattaaagcccctaaaaaaataacgcgaaaaaaatgagaattgtaaatcatggtcgaagcagaacctgggaccttgcatggtcct

>*eve4+6 D. vir*, DePace stock #0442
agcgcagctctgctctgctctgctctgtctttagttaaagtcgacttttggttgactttatgatttacagcgctcgtttttttcgcgttatttttttag
gagcgttaatggccatcgtaaagccgcagtgggccatgtcgcagtcgcagtcgccgtcgtcgtcgattcgcgcactgattctaaaaaatgaa
aacagtttttatgaatttctcacaacgcgcctcgactgcgacttggcctgttcctgtgcgttgtccttttcgccgttttctactattttcttgttgttg
ttttttttttgccttttacctagccgagtcgagcctggctcttttatgggcaaacaaatggcagcgattatattcggtttcggttttggttctgttttt
taggccataaaaattaggcggcataaaaaaaactgcattggaattctagttctagtttcagatttttaggttttcaggcttcgcgttgctgttttcta
tttacaatttttgttttttttttttttttaaattttcgcggaattcggaagcggaacaacgacggcccgaccttgccgtctctgaagagctattcaaa
gcgaaacatttggaaattgactcgctgcggtgcgcgaaatatttttagaactgactttcgaatcgtatgtcaattgtctctgatctgagct

>*eve4+6 fragment 1*, DePace stock #0331
cgacaatcaaactcgtgtttagccaacagtcgcagcatttccataccatgggggtggtctctgctgggtgttgcgaatgcgacgccaattggtc
agggttcactgtaaatggcccagccaaatgggaagcggcagttgaggagcgcccgaatcaattgccctgatggatgctgcggctgtccaaa
gttgcagcttttcgggtcactccgcgctggggctggagggctataaatccgcaggccagataatgaaactagaatgattgaggcaatcactg
gtgtggccagcagtctgctggcggagtttcctacgctctgcgcatgtcctggatacacacagctcggcgcacatagcgatagatacagatgca
gatacagatacagaggcagatagatacgtatgcggatacatagagcacgttacattatgtggcgaaggactggcgattaccgattccgcagt
tcaggacctctggatttgcgatcctgcgcacgacgtgtcaactttattgcggtttgactttgccgcggcccctcgaaactcacaaacgtatcctg
gaacccaggatccctgggctctgggctctggactatccgccgaccctccatatccatgatttacaattctcgttttttcgcgttatttttttaggg
gctttaatgaccgtcgtaaagccgcaggaggaccaggaccaggactctgctcacatttcgcgcactgattctaaaaaatgaaatcatttttct
tgaatttcacggcgcgcctcgagcaggactctttgttctcggccaggcaattgtccttttttgcgctcagctctcagttttttcgtccagcgggcat
tacctacacggcgttttatggcggagatgatattcgcctgggatcggttccgttttttaggccataaaaattaggcggcataaaaaaaactgcat
tggaattctagttctagtttcaagttttaggtttccaggtttctgccagcccgcctagattcgcatttcgcggaattcggaagcggaacagaat
gccagaatggtcagaatcctggctgaccttgccttttggccaggggccgtaaaaaaaattgactcgctgcggtgcgcggaatatttttaaatct
gactttccaacaatctctgatctgggtt

>*eve4+6 fragment 2*, DePace stock #0334
tcgagcaggactctttgttctcggccaggcaattgtccttttttgcgctcagctctcagtttttcgtccagcgggcattacctacacggcgtttta
tggcggagatgatattcgcctgggatcggttccgttttttaggccataaaaattaggcggcataaaaaaaactgcattggaattctagttctagtt
tcaagtttttaggtttccaggtttctgccagcccgcctagattcgcatttcgcggaattcggaagcggaacagaatgccagaatggtcagaatc
ctggctgaccttgccttttggccaggggccgtaaaaaaaattgactcgctgcggtgcgcggaatatttttttaaatctgactttccaacaatctctg
atctgggttcgaatcgtaaaaaaaaaagcagaacaaaaagcgggcattttcgtcggcaaatgatctgttaatgggccgggctaaaaaactaa
gtcacaaagtcacaaggttgtccggtaaattgacccggttaagaatgtctgtctgtaccgagaaggatgcaggacattcagcacttcaaagct
cccaccgctcgaaggattccccgaagattcacacggctggcgggggggatcaacgtttgatcagaaaccgccggcggctgttaacggcca

caaaaagcaagtgacatgcaaattaggcagattactgctggccagcaaggtcaggccgcccttaatgggcgtgaccgctggatctacctccc
tccccacccctgcccttcaacggcaaaggaaccgacaacggcagcaatacgatcaccacaaccaccgatcgatcgatcgagtgaccttcga
cctgccgcattgttatctgcgtgctgcgatttttttcttcatttttcacaggccatttgcgtgcgctcctgctgtccttggtgtgtcctttcagcatg

>*eve4+6 fragment 3*, DePace stock #0214
aggatccctgggctctgggctctggactatccgccgaccctccatatccatgatttacaattctcgtttttttcgcgttatttttttaggggctttaa
tgaccgtcgtaaagccgcaggaggaccaggaccaggactctgctcacatttcgcgcactgattctaaaaaatgaaatcatttttttcttgaattt
cacggcgcgcctcgagcaggactctttgttctcggccaggcaattgtccttttttgcgctcagctctcagttttttcgtccagcgggcattaccta
cacggcgttttatggcggagatgatattcgcctgggatcggttccgttttttaggccataaaaattaggcggcataaaaaaaactgcattggaat
tctagttctagtttcaagttttttaggtttccaggtttctgccagcccgcctagattcgcatttcgcggaattcggaagcggaacagaatgccaga
atggtcagaatcctggctgaccttgccttttggccaggggccgtaaaaaaaattgactcgctgcggtgcgcggaatatttttttaaatctgactttc
caacaatctctgatctgggttcgaatcgtaaaaaaaaaagcagaacaaaaagcgggcattttcgtcggcaaatgatctgttaatgggccgggc
taaaaaactaagtcacaaagtcacaaggttgtccggtaaattgacccggttaagaatgtctgtctgtaccgagaaggatgcaggacattcag
cacttcaaagctcccaccgctcgaaggattcccccgaagattcac

>*eve4+6 fragment 4*, DePace stock #0333
tcgagcaggactctttgttctcggccaggcaattgtccttttttgcgctcagctctcagttttttcgtccagcgggcattacctacacggcgtttta
tggcggagatgatattcgcctgggatcggttccgttttttaggccataaaaattaggcggcataaaaaaaactgcattggaattctagttctagtt
tcaagttttttaggtttccaggtttctgccagcccgcctagattcgcatttcgcggaattcggaagcggaacagaatgccagaatggtcagaatc
ctggctgaccttgccttttggccaggggccgtaaaaaaaattgactcgctgcggtgcgcggaatatttttttaaatctgactttccaacaatctctg
atctgggtt

>*eve4+6 extended*, DePace stock #0635
tacattatgtggcgaaggactggcgattaccgattccgcagttcaggacctctggatttgcgatcctgcgcacgacgtgtcaactttattgcgg
tttgactttgccgcggcccctcgaaactcacaaacgtatcctggaacccaggatccctgggctctggactatccgccgaccctccatatccatg
atttacaattctcgtttttttcgcgttatttttttaggggctttaatgaccgtcgtaaagccgcaggaggaccaggaccaggactctgctcacatt
tcgcgcactgattctaaaaaatgaaatcatttttttcttgaatttcacggcgcgcctcgagcaggactctttgttctcggccaggcaattgtccttt
tttgcgctcagctctcagttttttcgtccagcgggcattacctacacggcgttttatggcggagatgatattcgccttggatcggttccgttttttа
ggccataaaaaattaggcggcataaaaaaaactgcattggaattctagttctagtttcaagttttttaggtttccaggtttctgccagcccgcctaga
ttcgcatttcgcggaattcggaagcggaacagaatgccagaatggtcagaatcctggctgaccttgccttttggccaggggccgtaaaaaaaa
ttgactcgctgcggtgcgcggaatatttttttaaatctgactttccaacaatctctgatctgggttcgaatcgtaaaaaaaaaaagcagaacaaaa
agcgggcattttcgtcggcaaatgatctgttaatgggccgggctaaaaaaactaagtcacaaagtcacaaggttgtccggtaaattgacccgg
ttaagaatgtctgtctgtaccgagaaggatgcaggtcattctgcacttcaaagctcccaccgctcgaaggattcccccgaagattcacacggc
tggcgggggggatcaacgtttgatcagaaaccgccggcggctgttaacggccacaaaaagcaagtgacatgcaaattaggcagattactgc
tggccagcaaggtcaggccgcccttaatgggcgtgaccgctggatctacctccctccccacccctgcccttcaacggcaaaggaaccgacaa
cggcagcaatacgatcaccacaaccaccgatcgatcgatcgagtgaccttcgacctgccgcattgttatctgcgtgctgcgatttttttcttcatt
tttcacaggccatttgcgtgcgctcctgctgtccttggtgtgtccttttagcatgcgtgtccttttagcatgctaaaaatcaagcgatcaagaat
ctgcgcgatcttagccggaatcgggattagtcgtttatggcccacgccctcaatcctccataaaacactaagcgctttcgcctaatgtatgtatc
tcttcatatctggaaatctcac

>*eve4+6 spacer 1*, DePace stock #0636
tacattatgtggcgaaggactggcgattaccgattccgcagttcaggacctctggatttgcgatcctgcgcacgacgtgtcaactttattgcgg
tttgactttgccgcggcccctcgaaactcacaaacgtatcctggaacccaggatccctgggctctggactatccgccgaccctccatatccatg
atttacaattctcgtttttttcgcgttatttttttaggggctttaatgaccgtcgtaaagccgcaggaggaccaggaccaggactctgctcacatt
tcgcgcactgattctaaaaaatgaaatcatttttttcttgaatttcacggcgcgcctcgagcaggactctttgttctcggccaggcaattgtccttt

tttgcgctcagctctcagttttttcgtccagcgggcattacctacacggcgttttatggcggagatgatattcgccttggatcggttccgtttttta
ggccataaaaattaggcggcataaaaaaactgcattggaattctagttctagtttcaagtttttaggtttccaggtttctgccagcccgcctaga
ttcgcatttcgcggaattcggaagcggaacagaatgccagaatggtcagaatcctggctgaccttgccttttggccaggggccgtaaaaaaa
ttgactcgctgcggtgcgcggaatatttttttaaatctgactttccaacaatctctgatctgggttcgaatcgtaaaaaaaaaaagcagaacaaaa
agcgggcattttcgtcggcaaatgatctgttaatgggccgggctaaaaaactaagtcacaaagtcaccgaggtgcagccgactacgactgca
gactcgcacgctagccagccggaccgattccagtggcagtcttgagtggactttcgcgcgcgtcacgcgcgaccattcacgctagcgtacag
cggtaggctctagccgggatgcgaccggtgggaagatgcggtctgcgacttctagcgctctgggcggtgtactgctggacacgtgagcgttc
gccatccaacagccagcctcgcatacggctgccgatgaatgcggcggcacatgagaatcgtccgcgcgagggagctcgcggcgcagtggg
aacgagtcttcaaggtgtacgatccgtcgacggacgtttcaaagtcctatcatcgcctctacggaaactgtgaagacgctgtctgggtcgtgc
gagtatatgtcgaacgacgtttgcagagtgagagaagagaccctcgtagcggatactttcagtattgaagcatccccggctcctcggcgttcg
gtgctgagactgttcacaccgtgcctcgcgacgcgcgcgcagtctcaggaccgcgacagcccgaaacctcactctgagtacaagttagcgtg
ctgtgtgaaggccgtctgtggagtgggcgatt

>eve4+6 spacer 2, DePace stock #0637
tacattatgtggcgaaggactggcgattaccgattccgcagttcaggacctctggatttgcgatcctgcgcacgacgtgtcaactttattgcgg
tttgactttgccgcggcccctcgaaactcacaaacgtatcctggaacccaggatccctgggctctggactatccgccgaccctccatatccatg
atttacaattctcgtttttttcgcgttatttttttaggggctttaatgaccgtcgtaaagccgcaggaggaccaggaccaggactctgctcacatt
tcgcgcactgattctaaaaaatgaaatcatttttttcttgaatttcacggcgcgcctcgagcaggactctttgttctcggccaggcaattgtccttt
tttgcgctcagctctcagttttttcgtccagcgggcattacctacacggcgttttatggcggagatgatattcgccttggatcggttccgtttttta
ggccataaaaattaggcggcataaaaaaaactgcattggaattctagttctagtttcaagtttttaggtttccaggtttctgccagcccgcctaga
ttcgcatttcgcggaattcggaagcggaacagaatgccagaatggtcagaatcctggctgaccttgccttttggccaggggccgtaaaaaaa
ttgactcgctgcggtgcgcggaatatttttttaaatctgactttccaacaatctctgatctgggttcgaatcgtaaaaaaaaaaagcagaacaaaa
agcgggcattttcgtcggcaaatgatctgttaatgggccgggctaaaaaactaagtcacaaagtcatgtccaatgcagccgcgagtctctcca
cactcagactctcgtcgtgtcggctgtagctgttgagactatgatactttctcagtgtatacgcatctctggcgatggcagtgcccagctcggcg
aggggccgtctacgacatcttaacgccgattcggcacagcacacgtgtatacgcttgcagcagtcactcatagcgaccgcttggcatatgcat
gatcgcgatggtcgccgccgaagcagacgcatccggtgagtggaatgctcgtacgatatcggcgcctgggcgcaccgacatcagcagttgt
gccgagtgtacagctcgcaccgaagttcttacaacgttgcgagtcgtagcactgaatcagtaggcgtctcgtgcgactacagtatgatcgcca
ccgcgacattcagttctggccgtgtcagtgccatcggacgctgaagaccgccgagacagctgtggtatgagtcgactagactggatgtgaac
tatcccaagagcgtgtacagtgtcgctgacgaatgcctcccaggcgacaatcgtcttgagacagtagtagccatcgatggagacgtttaggc
gtcggtccgatccggcatacgaactca

>eve4+6 S. cyn, DePace stock #0659
ggactgcgaggatcccgcagtggagtcccacatctgaacctatgcgactgaggagacgaagtccgacgctcgtgggcatgcttcgacgcgc
tatgcgtcggcatgcgcatctcatgccgactttgcgcggccacagatcccatgtaccgctagacggtcctcgactgctcgcgaggaacgtccc
gatagatagacgtgctaacttgaggtcttacagacgtggacatcacagtccaactagcgtgcgatggcacagctcctgtaccgacgcattgg
actgttagctcgagtgagtctataccatgtcactcgcgaagtatccgacggcagatgccagtgcgtgtatcgcggctggagacttcgagatcg
acgtatcttcgggtccgttctgacggtcgagagccattcggcctcatactttcagcatctctagtgggagactcccgcagagtttccgcttgaca
gctacgaacattgtactgtaagattcccaaggagactgtcactttgacattctgtgtctccgcaagcatggtgagtacaggatctgtcgccagc
cagttggctcgaaagtgccgagttccgacgcgcagaacttcgctgtcaggcgctagcaccctcgtacactgcgctcaggggccgaagatcgc
tgccatgttggttttccgttgatttaagtggggtcgttgttgttgttgttgtcgtactatgagattggtgggtgagggcggcggtcgtggtggtga
taaatgcctgctacccgcttaaaattaataatacacttgaagactggttcttgttgttgtcgcgccacaccagagtcggaggctgagaccacaa
cccaaagacggagccgcaaaagataaattatttcattgattagtctattcttcgggccatatgcattcgttaaattaaaaaaagaggaaatcat
agaaaatttgaatgtaaaacgcaaagtggaatacaaaaatcaaaataaaaaacaaaataaatactaaaataaataaaatcgattaaaaat
ataatataatagattaaaaataaaaaatagaaattatattatacaatattaatttttttgtaattatttaaaggagtaaaattagaaacaaaat
ttaatccaacaaaatcgagcataaaattttccggccattactccacaccctccctttgcacccacacaattaaatggcccattaaatcaatttta

aaacgatgctcccgtcgcctcggcgaggagttctgggctctggtaaaccaccactggctaattgctgccatcaataaaatcgactccgatgcc
gaaacacgccaatcattctcataaaagcaccaattattcgtaaccataccaaagtgcgtttgtctgtgtgtctgtgtgagtgtgggtggttgcc
actatgtgtgagtctctccacactcagactctcgtcgtgtcggctgtagctgttgagactatgatactttctcagtgtatacgcatctctggcgat
ggcagtgcccagctcggcgaggggccgtctacgacatcttaacgccgattcggcacagcacacgtgtatacgcttgcagcagtcactcatag
cgaccgcttggcatatgcatgatcgcgatggtcgccgccgaagcagacgcatccggtgagtggaatgctcgtacgatatcggcgcctgggcg
caccgacatcagcagttgtgccgagtgtacagctcgcaccgaagttcttacaacgttgcgagtcgtagcactgaatcagtaggcgtctcgtgc
gactacagtatgatcgccaccgcgacattcagttctggccgtgtcagtgccatcggacgctgaagaccgccgagacagctgtggtatgagtc
gactagactggatgtgaactatcccaagagcgtgtacagtgtcgctgacgaatgcctcccaggcgacaatcgtcttgagacagtagtagcca
tcgatggagacgtttaggcgtcggtccgatccggcatacgaactca

>*eve5*, DePace stock #0215
atatcccaaggccgcaaagtcaacaagtcggcagcaaatttccctttgtccggcgatgtgtttttttttttagccataactcgctgcattgtttggg
ccaagtttttcttctgccaaattgcggagatgatgcggggattatgcgctgattgcgtgcaattatggacatcctgcgaggccccgaggaactt
cctgctaaatcctttcatccgcctacagaaccccctttgtgtcccgttcgccgggagtccttgacgggtccttcgactattcgcttacagcagcttg
cgtaaaatttcataaccctacgagcggctcttccgcggaatccctggcattatccttttacctcttgccaatccgttggctaaaaaacggcttc
gacttccgcgtaactgctggacaacaaagacaaaaaacggcgaaaggacggcgatttccaggtagcattgcgaattccgtcaaactaaag
gaccggttatataacgggtttatatggccagaatctctgcatctccacgaccgccagaagctgcgtaaaactgcaggctctgttttgatttctgc
aacttcagttaattgcccgggatggccagcaattgccggcaattataaaacagcgcagatgtgactcagcttccatatctaactctatatctca
tgccgaaaatctagggtggggagcggaggggcggggtgcgtgggtgacttgcctgccagggaaaggggcgggggttcagcgggtgata
aatgtgcgtgatttggaatgaatgcgcatcgattaaaaccgcagggcaatcaattt

>*eve5 add Cad*, DePace stock #0485
atatcccaaggccgcaaagtcaaTaaAtcggcagcaaatttccctttgtccggcgatgtgtttttttttttagcAataaAtcgctgcattgtttg
ggccaagtttttcttctgccaaattgcggagatgatgcggggattatgcgTtTattgcgtgcaTttatTgacatcctgcgaggccccgaggga
acttcctgctaaatcAAtAAatccgcctacagaaccccctttgtgtcccgttcgccgggagtccttgacgggtccttcgactattcgcttacagc
agcttgcgtaaaatttcataaccctacgagcggctcttccgcggaatccctggcattatccttttacctcttgccaatccgttggctaaaaaac
ggcttcgacttccgcgtaactgctggacaaTaaagacaaaaaacggcgaaaggacggcgatttccaggtagcattgcgaattccgtcaaac
taaaggaccggttatataacgggtttatTGggccagaatctctgcatctccacgaccgccagaagctgcgtaaaactgcaggctctgttttT
attActgcaacttcagttTattgcccgggatggccagcaattgccggcaattataaaacagcgcagatgtgactcagcttccatatctaactc
tatatctcatgccgaaaatctagggtggggagcggaggggcggggtgcgtgggtgacttgcctgccagggaaaggggcgggggttcagc
gggtgataaatgtgcgtgatttggaatgaatgcgcatcgattaaaaccgcagggcaatAaattt

>*eve5 mut Hb*, DePace stock #0489
atatcccaaggccgcaaagtcaacaagtcggcagcaaatttccctttgtccggcgatgtgtttCCCCtttagccataactcgctgcattgttt
gggccaagtttCCcttctgccaaattgcggagatgatgcggggattatgcgctgattgcgtgcaattatggacatcctgcgaggccccgagg
aacttcctgctaaatcctttcatccgcctacagaaccccctttgtgtcccgttcgccgggagtccttgacgggtccttcgactattcgcttacagca
gcttgcgtaaaatttcataaccctacgagcggctcttccgcggaatccctggcattatccttttacctcttgccaatccgttggctaaGGaac
ggcttcgacttccgcgtaactgctggacaacaaagacaaGGaacggcgaaaggacggcgatttccaggtagcattgcgaattccgtcaaa
ctaaaggaccggttatataacgggtttatatggccagaatctctgcatctccacgaccgccagaagctgcgtaaaactgcaggctctgttttga
tttctgcaacttcagttaattgcccgggatggccagcaattgccggcaattataaaacagcgcagatgtgactcagcttccatatctaactctat
atctcatgccgaaaatctagggtggggagcggaggggcggggtgcgtgggtgacttgcctgccagggaaaggggcgggggttcagcgg
gtgataaatgtgcgtgatttggaatgaatgcgcatcgattaaaaccgcagggcaatcaattt

>*eve5 mut Hb add Cad*, DePace stock #0487
atatcccaaggccgcaaagtcaaTaaAtcggcagcaaatttccctttgtccggcgatgtgtttCCCCtttagcAataaAtcgctgcattgtt
tgggccaagtttCCcttctgccaaattgcggagatgatgcggggattatgcgTtTattgcgtgcaTttatTgacatcctgcgaggccccga
ggaacttcctgctaaatcAAtAAatccgcctacagaaccccttgtgtcccgttcgccgggagtccttgacgggtccttcgactattcgcttac
agcagcttgcgtaaaatttcataaccctacgagcggctcttccgcggaatccctggcattatccttttacctcttgccaatccgttggctaaG
GaacggcttcgacttccgcgtaactgctggacaaTaaagacaaGGaacggcgaaaggacggcgatttccaggtagcattgcgaattccg
tcaaactaaaggaccggttatataacgggtttatTGggccagaatctctgcatctccacgaccgccagaagctgcgtaaaactgcaggctct
gttttTattActgcaacttcagttTattgcccgggatggccagcaattgccggcaattataaaacagcgcagatgtgactcagcttccatatct
aactctatatctcatgccgaaaatctagggtggggagcggaggggcggggtgcgtgggtgacttgcctgccagggaaaggggggcgggggt
tcagcgggtgataaatgtgcgtgatttggaatgaatgcgcatcgattaaaaccgcagggcaatAaattt

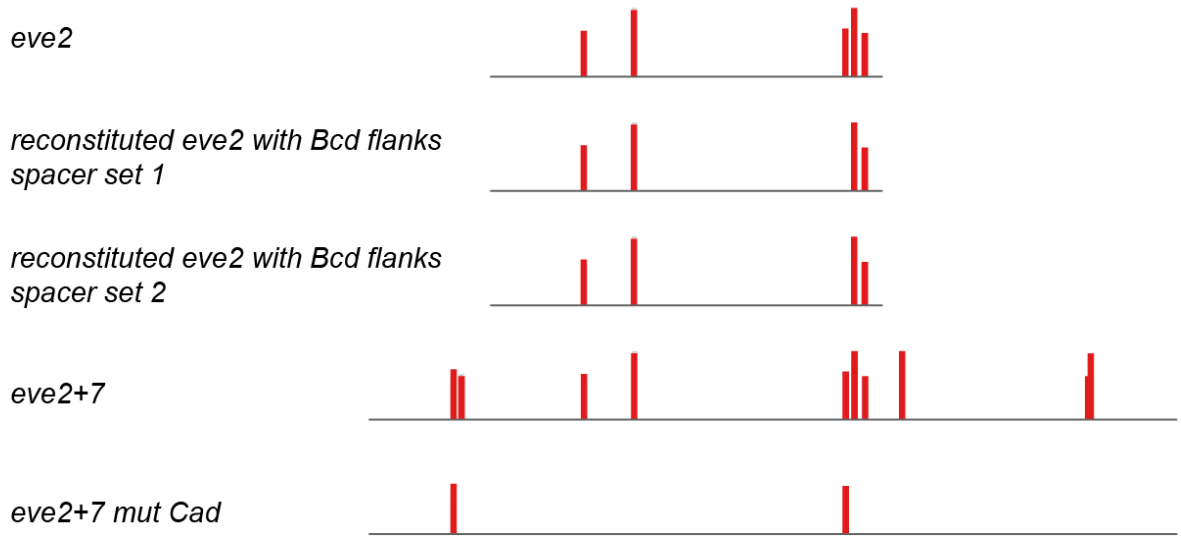
>*eve5 add Bcd*, DePace stock #0490
atatcccaaggccgcaaagtcaacaagtcggcagcTaatCtccctttgtccggcgatgtgttttttttttAaTccataactcgctgcattgtttg
ggccaagtttttcttctgccaaattgcggagatgatgcggggattatgcgctgattgcgtgcaattatggacatcctgcgaggccccgaggaa
cttcctgctaaatcctttAatccgcctacagaaccccttgtgtcccgttcgccgggagtccttgacgggtccttcgactattcgcttacagcGg
AttAcgtaaaatttcataaccctacgagcggctcttccgcggaatccctggcattatccttttacctcttgcTaatccgttggctaaaaaacg
gcttcgacttccgcgtaactgctggacaacaaagacaaaaaacggcgaaaggacggcgatttccaggtagcattgcgaattccgtTaaTc
CaaaggaccggttatataacgggtttatatggccagaatctctgcatctccacgaccgccagaagctgcgtTaaTcCgcaggctctgttttg
atttctgcaacttcagttaattgcccgggatggccagcaattgccggcaattataaaacagcgcagatgtgactAaTcCtccTAatcCaac
tctatatctcatgccgaTaatcCagggtggggagcggaggggcggggtgcgtgggtgacttgcctgccagggaaaggggggcgggggttca
gcgggtgataaatgtgcgtgatttggaatgaatgcgcatGgattaaaaccgcagggcaatcaattt

>*eve5 mut Hb add Bcd*, DePace stock #0480
atatcccaaggccgcaaagtcaacaagtcggcagcTaatCtccctttgtccggcgatgtgtttCCCCttAaTccataactcgctgcattgt
ttgggccaagtttCCcttctgccaaattgcggagatgatgcggggattatgcgctgattgcgtgcaattatggacatcctgcgaggccccgag
gaacttcctgctaaatcctttAatccgcctacagaaccccttgtgtcccgttcgccgggagtccttgacgggtccttcgactattcgcttacag
cGgAttAcgtaaaatttcataaccctacgagcggctcttccgcggaatccctggcattatccttttacctcttgcTaatccgttggctaaGG
aacggcttcgacttccgcgtaactgctggacaacaaagacaaGGaacggcgaaaggacggcgatttccaggtagcattgcgaattccgtT
aaTcCaaaggaccggttatataacgggtttatatggccagaatctctgcatctccacgaccgccagaagctgcgtTaaTcCgcaggctct
gttttgatttctgcaacttcagttaattgcccgggatggccagcaattgccggcaattataaaacagcgcagatgtgactAaTcCtccTAat
cCaactctatatctcatgccgaTaatcCagggtggggagcggaggggcggggtgcgtgggtgacttgcctgccagggaaaggggggcggg
ggttcagcgggtgataaatgtgcgtgatttggaatgaatgcgcatGgattaaaaccgcagggcaatcaattt

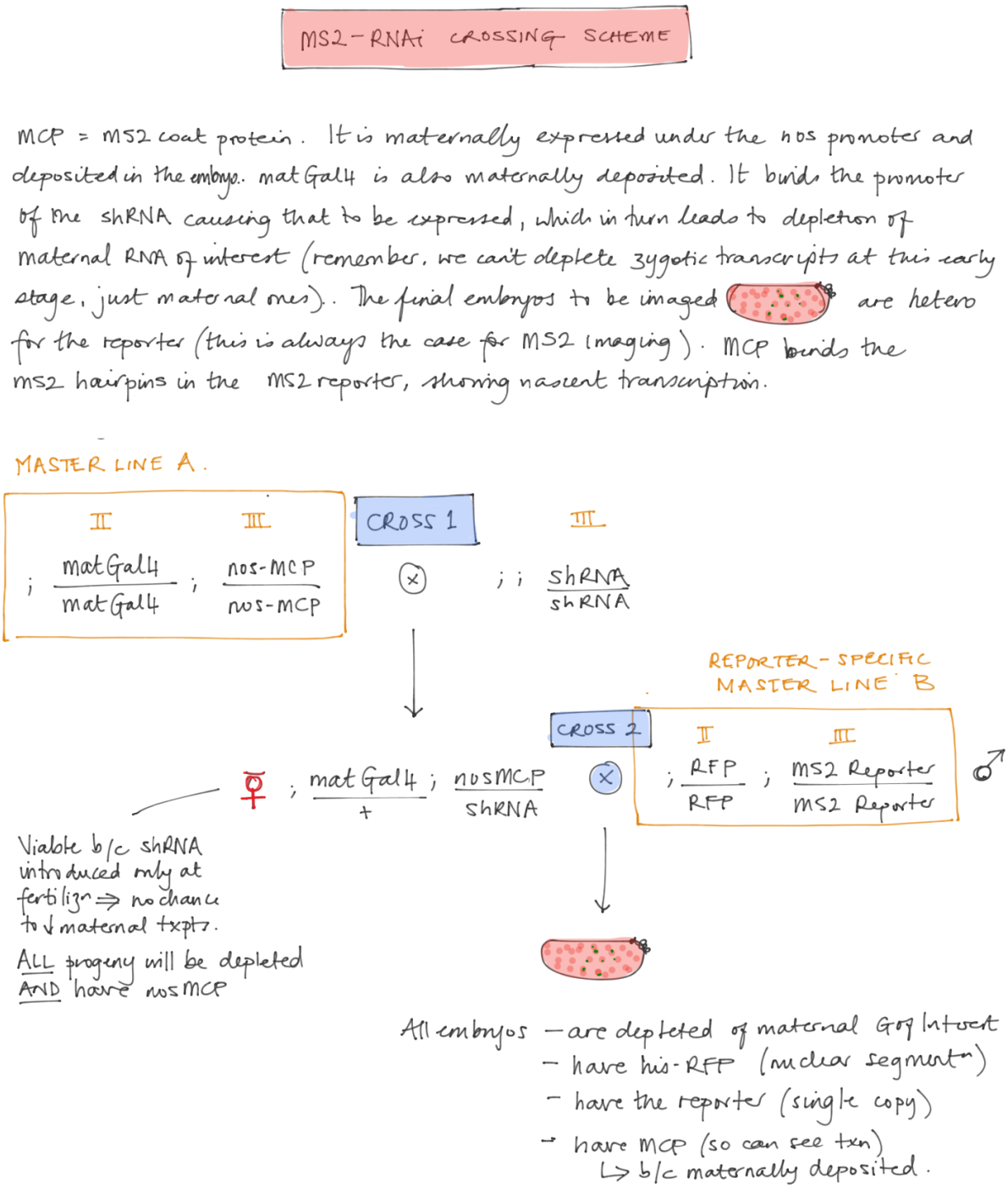
>*eve5 mut Hb add Cad add Bcd*, DePace stock #0634
agaaggcttgcatgtgggccttttccaggtcggccagtaggtagagttgttgcgatgcggctatgccgggcgagttaatgccaatgcaaattg
cgggcgcaatataacccaataGGttgaagtaactggcaggagcgaggtatccttcctggttacccggtactgcataacaatggaacccgaa
ccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaatccgtttgccatcagcgagattatATgtcaattgca
gttgcagcgtttcgctttcgtcctcgtttcactttcgagttagacGCtattgcagcatcttgaacaatcgtcgcagtttggtaacacgctgtgcc
atactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcgaagtcagggcattccgccgatctagcca
tcgccatcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaatccaatcccgatccctagcccgatcccaatcc
caatcccaatcccttgtccttttcattagaaagtcGCaaaGGcacataataGGgatgtcgaagggattaggggcgcgcaggtccaggca
acgcaGCtaacggactagcgaactgggttattCCtttgcgccgacttagccctgatccgcgagcttaacccgttttgagccgggcagcagg
tagttgtgggtggaccccacgattCCtttggccaaacctccaagctaacttgcgcaagtggcaagtggccggtttgctggcccaaaagagga
ggcactatcccggtcctggtacagttggtacgctgggaatgattatatcatcataGCaaatgttttgcccaacgaaaccgaaaactttcaaa
ttaagtcccggcaactgggttcccattttccattttccatgttctgcgggcaggggcggccattatctcgct

APPENDIX E: SUPPLEMENTAL MATERIAL FOR CHAPTER 5

Supplemental Figures

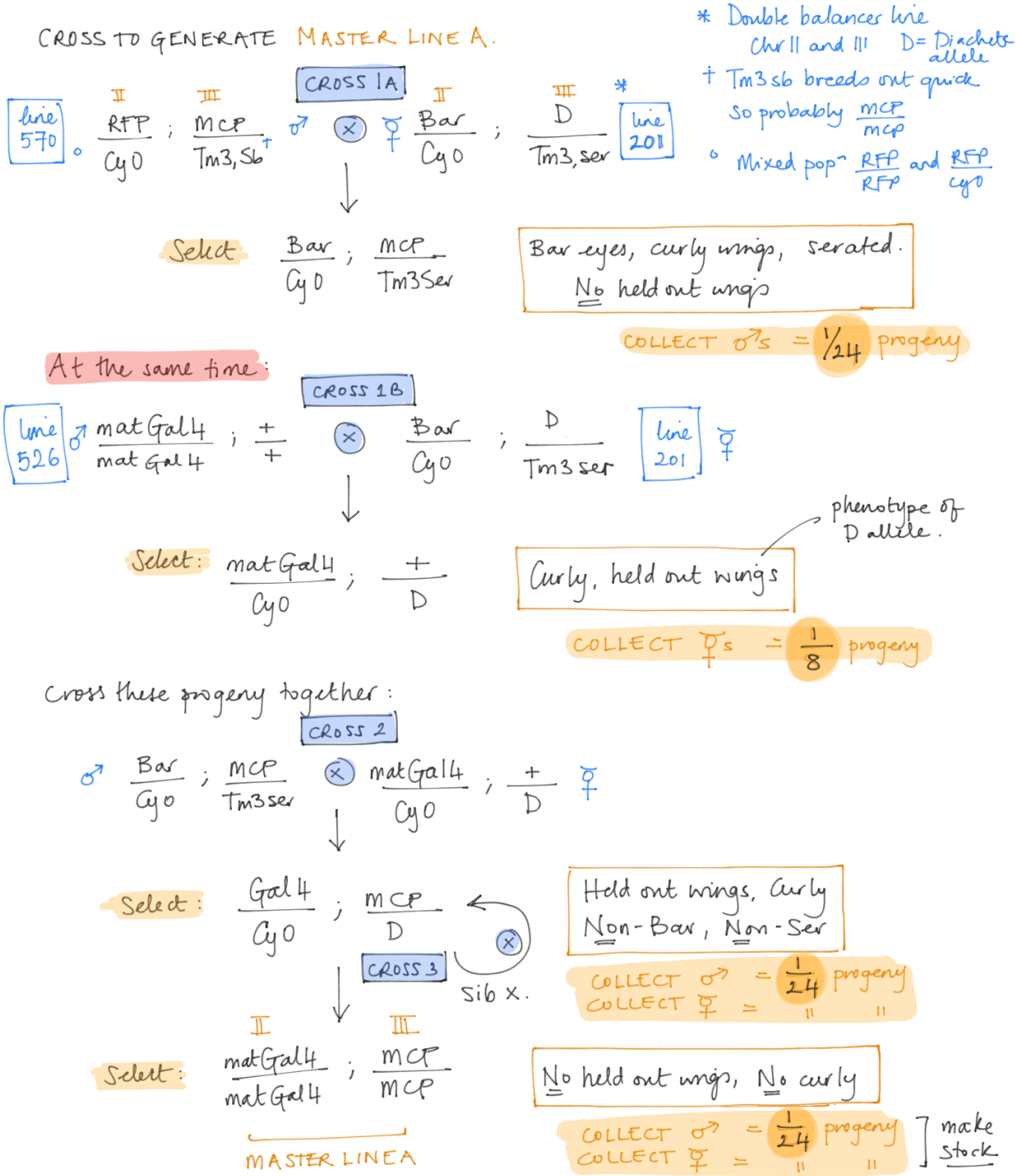


Supplemental Figure 5.1: Reconstituted *eve2* enhancers with Bcd flanking sequence include important Cad sites. PATSER-predicted Cad binding sites are plotted in red for WT and reconstituted *eve2* enhancers as well as for *eve2+7* and *eve2+7 mut Cad*. Reconstituted enhancers contain all predicted Cad sites mutated in *eve2+7 mut Cad*.

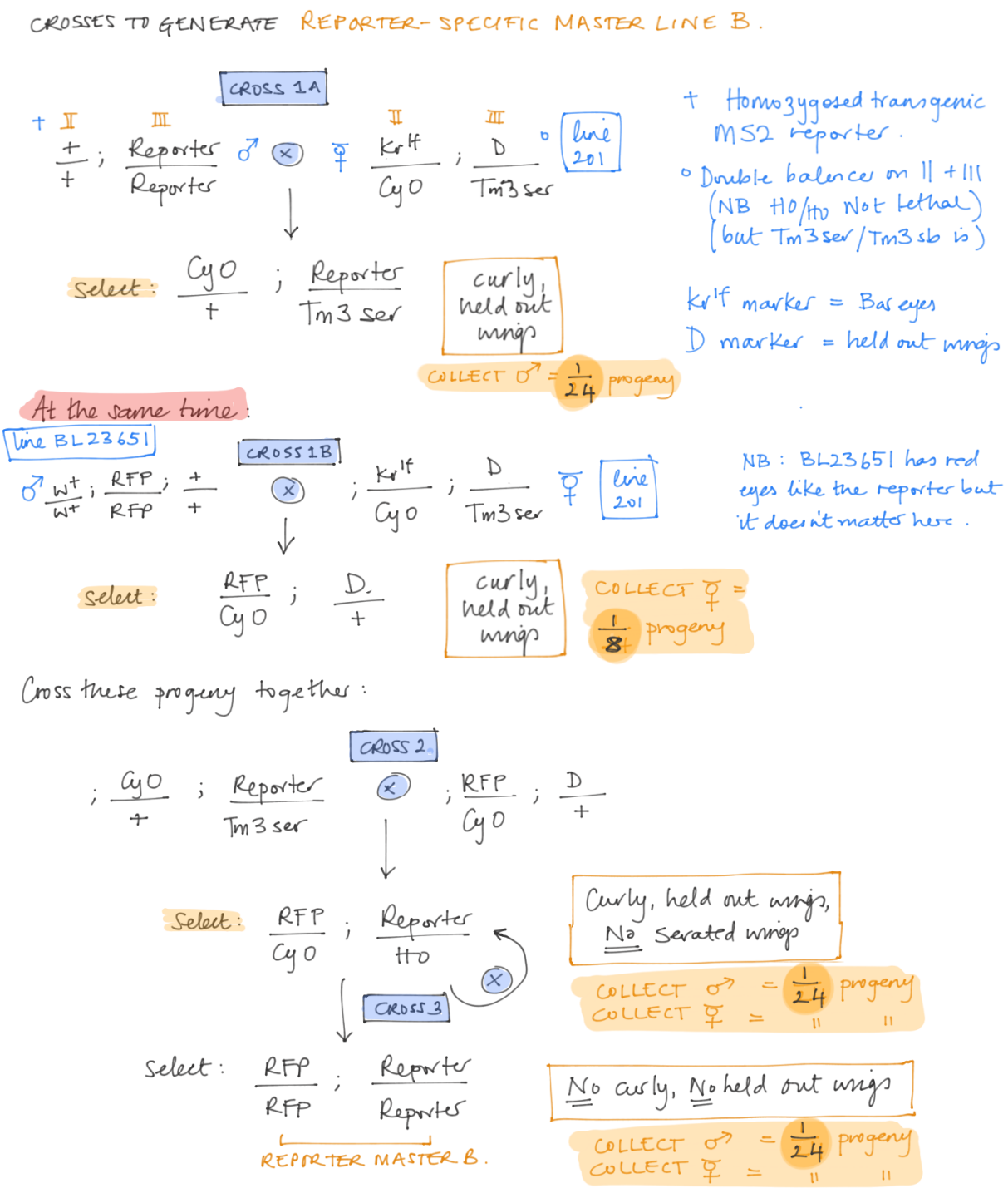


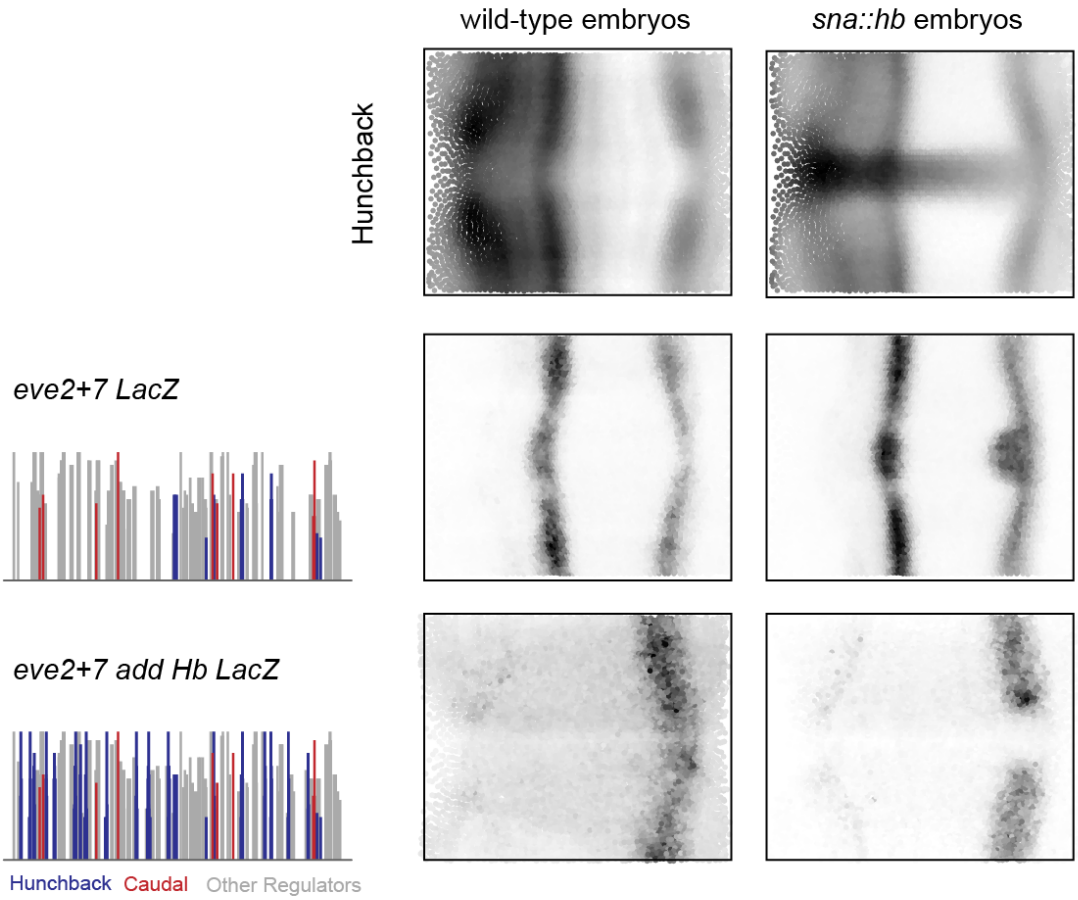
Supplemental Figure 5.2: Crossing scheme to generate maternal RNAi-depleted embryos with MS2 components. Figure was hand-drawn to perfection by Clarissa Scholes.

Supplemental Figure 5.2 (Continued).



Supplemental Figure 5.2 (Continued).





Supplemental Figure 5.3: Adding Hunchback sites to *eve2+7* allows for Hb repression. Left: predicted binding sites in *eve2+7* and *eve2+7 add Hb*. Binding sites plotted as in Figure 3.4. Right: projections of atlas data for Hb protein and *eve2+7 LacZ* mRNA in time point 4, as well as representative time point 4 pointclouds for *eve2+7 add Hb*.

Supplemental Methods

Construct design

We used SiteOut to design a new set of reconstituted *eve2* enhancers with flanks around annotated Bcd binding sites. We used the Spacer Designer tool with a p-value cutoff of 0.01 and a GC content of 48.35%. We used PWMs for Bcd, Cad, D, Gt, Hb, Kni, Kr, Stat92E and Zelda. For each factor, we included one or more PWMs from FlyFactorSurvey (Zhu et al. 2011) or the early bacterial 1-hybrid database (Noyes et al. 2008), and for each PWM, we generated 3 different frequency matrices with pseudocounts of 1, 0.1 and 0.01. We also used SiteOut to design proximal *Kr* shadow enhancers with mutations in Hb sites, Cad sites, or both Hb and Cad Sites. We used the Refine a Sequence tool with Hb and Cad PWMs from the set described above with a p-value cutoff of 0.05 and a GC content of 37.1%.

To design sequences for the *eve2* MS2 experiments, we introduced changes that created consensus motifs for Bcd (Noyes et al. 2008), Stat92E (Hou et al. 1995) and Zelda (Liang et al. 2008). We cloned these sequences into a plasmid that contains an MS2-LacZ cassette downstream of the *eve* promoter, with a 667 bp spacer sequence that was designed to minimize predicted binding sites for *eve* regulators (see below, *eve2 MS2 promoter spacer*). We designed this sequence using the Random Sequence tool, although the SiteOut parameters and PWM set used for this design have been lost to the sands of time.

Measuring nascent transcription in embryos

We used light sheet-based fluorescence microscopy (Wu et al. 2013) based on the MS2 reporter system (Garcia et al 2013) to collect data on nascent transcription. Males containing MS2 reporter constructs were crossed with females expressing MS2 Coat Protein-GFP and Histone-RFP (DePace stock #0570). Embryos from that cross were collected on molasses plates, dechorionated in 50% bleach, mounted using heptane glue, and submerged in water for imaging. Interested parties should contact Tim Harden for additional technical details on this protocol.

RNAi against maternal cofactors

We used the Gal4-UAS system to express shRNAs against *MED29* and *mor* (Staller et al. 2013). We crossed mat-tub-GAL4 females (DePace stock #0411) with UAS-shRNA males (*MED29*: DePace stock #0517; *mor*: DePace stock #0510) and crossed resulting siblings in bottles. F2 flies were corralled in cages. We collected, fixed, stained, imaged and analyzed embryos as described in previous chapters.

CRISPRi

For pilot studies, we obtained a transgenic line expressing *eve*-YFP from a BAC where the fusion is embedded in 16.4kb of eve regulatory sequence (a generous gift from Misha Ludwig and Marty Kreitman). gRNAs targeting *eve2* were designed by identifying all protospacer adjacent motifs in *eve2* and were synthesized using the Agilent SureGuide kit. We collected these embryos for 1 hour at 25˚C, bleached them to remove the chorion, and aligned them on a breathable membrane. We inject gRNAs at a concentration of 2000 ng/uL using an Eppendorff FemtoJet 4x microinjector. We imaged YFP in live embryos using a Zeiss 710 confocal microscope. Raw image stacks were analyzed in ImageJ.

Enhancer Sequences

>*reconstituted eve2 with Bcd flanks, spacer set 1*; DePace stock #0660
ggttacccggtacTGACTGGCAGCGCTTGACACATGCtaactgggacaTGTGACTGACACACActggtttCGA
ATGCGAAgtgccgtgttaatccgtttgccatcagcgagattattagtcaattgcagttgcGAGCTGTTGGCAAGCACGCCT
GGGATCCCAGCCCCGAgactttattgcagcatcttgaacaatcgtcgcagtttggtaacacgctGTCGGGCCTCGAAGT
CTagacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcAATTGCGGCCGATCTTGGAGGTGTA
CGAAGCACCGCTCCTGAGACAGgtttgtttgtttgtttgctgggattagccaagggcttgaATCGACGtccaatcccgatcc
ctagcccgatcccaatcccaatcccaatccctAGTAGCGAGCATGCgaaagtcataaaaacacataataatGATGTCGAAG
GGATTAGGGG

>*reconstituted eve2 with Bcd flanks, spacer set 2*; DePace stock #0661
ggttacccggtacTATCTTTCGACCGAGTGGCAAGCAtaactgggacaTTCGAATGACAGACActggtttCCGA
CGTATAgtgccgtgttaatccgtttgccatcagcgagattattagtcaattgcagttgcCTTGTCTCGGGCGACGATCCGA
CATAGTCGAGGTTCAgactttattgcagcatcttgaacaatcgtcgcagtttggtaacacgctGAGAATGACACATGTC

TagacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcAGACATCCGGCCGGGCCGTGGGTCGA
AGGCCCACACGTCTAGTCAGAgtttgtttgtttgtttgctgggattagccaagggcttgaCTACACGtccaatcccgatccct
agcccgatcccaatcccaatcccaatccctAGTAGGTGTCTATCgaaagtcataaaaacacataataatGATGTCGAAGG
GATTAGGGG

>*eve2 MS2*, DePace stock #0619 (same enhancer sequence as *eve2*, DePace stock #0216)
ggttacccggtactgcataacaatggaacccgaaccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaat
ccgtttgccatcagcgagattattagtcaattgcagttgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttga
acaatcgtcgcagtttggtaacacgctgtgccatactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggtt
gcgaagtcagggcattccgccgatctagccatcgccatcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaa
tccaatcccgatccctagcccgatcccaatcccaatcccaatcccttgtccttttcattagaaagtcataaaaacacataataatgatgtcgaa
gggattagggg

>*eve2 MS2 add Bcd*, DePace stock #0620
ggttacccggtactgcataacaatggaacccgaaccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaat
ccgtttgccatcagcgagattattagtcaattgcagttgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttga
acaatcgtcgcagtttggtaacacgctgtgccatactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggtt
gcgaagtcagggcattAATccgatctagccatcgccatcttAATcCggcgtttgtttgtttgtttgctgggattagccaagggcttgacttg
gaatccaatcccgatccctagcccgatcccaatcccaatcccaatcccttgtccttttcattagaaagtcataaaaacacataataatgatgtc
gaagggattagggg

>*eve2 MS2 add Zelda*, DePace stock #0621
ggttacccggtactgcataacaatggaacccgaaccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaat
ccgtttgccatcagcgagattattagtcaattgcagttgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttga
acaatcgtcgcagtttggtaacacgctgtgccatactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggtt
gcgaagtcagggcattcAgGTATtctagccatcgccatcttcAgGTAgcgtttgtttgtttgtttgctgggattagccaagggcttgactt
ggaatccaatcccgatccctagcccgatcccaatcccaatcccaatcccttgtccttttcattagaaagtcataaaaacacataataatgatgt
cgaagggattagggg

>*eve2 MS2 add Stat92E*, DePace stock #0622
ggttacccggtactgcataacaatggaacccgaaccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaat
ccgtttgccatcagcgagattattagtcaattgcagttgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttga
acaatcgtcgcagtttggtaacacgctgtgccatactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggtt
gcgaagtcagggcattTcCcGgaActagccatcgccatcttcCCGgAAcgtttgtttgtttgtttgctgggattagccaagggcttgactt
ggaatccaatcccgatccctagcccgatcccaatcccaatcccaatcccttgtccttttcattagaaagtcataaaaacacataataatgatgt
cgaagggattagggg

>*eve2 MS2 promoter spacer*
CTAGTTAGAAAATTTGTAGTAATATTCGTATATCTAGTATTGGTTCGGCATAGGAGATATAGTA
TCACCTTAGAGTTATATCTGTACACTGCGTTACTATTACTATTCTAGATCCTAAGTCTGGCTCAC
AAATATTGCTGTGGAGACACATATATTCGCTATAGAGATATTCTTCCTTCGTATGCTCTTCTAA
AATTTCTGAGTCACTGATTGTAGTTCTCAGCGAGCATTGGCTAGTACGATTGAAGAACTCTTGG
ATTCTAGCACTATTAGGAAGTCTTCATCTTCTTAGATTCTTACAGACTCATGTGAAACTCTGAAA
CACTCTTAGATCTGATAGATTCTATGATCAATACTACTAAGGTTGATATTCTAGAAACTCACTAC
AACGAGCCTGTATTCGTGTCTTGTACTCAACGCTCGTGAAGTGTCTAGCAGACTTCACTATACT

TTACACGTACATCACACTGTACAGAGAATACAGATTGTACTGTGAAACATACGACGTGATGAAC
GTTCTATCTAGTCTGTTAGAAGATACATTGTCGCACATACGATTGCTGAGACTACTCAAGAAGT
GAGGGTCTTAGAGTAAGATCGAATACGGTTCCACACAGAATATGATATCTAAGATCCACGATG
ATTCGTGAGATCGTTGTATCTACGA

>*eve2+7 add Hb*, DePace stock #0550
agaaggcttgcatgtgggccttttccaggtcggccagtaggtagagttTttTTgatgcggctatgccgggcgTTttTTtgccaaAAAaa
attgcgggcgcaatataacccaataattCAaaAAaactggcaggagcgaggtatTTttTTtggttacccggtactgcataacaatggaa
cccgaaccgtaactgggacagaAAAaaaaActggcctggtttTtTgctgtgtgtgTTTtTttaatccgtttgccatcagcgagattatta
gtcaattgcagttgcagcgtttcgTtttTTtcctcgtttcactttcgagttagactttattgcagcatcttgaacaatcgtcgcagtttggtaaca
cgctgtgccataGtttTTtttagacggaatcgagggaccctggacAaAaaAAAcacaacgagaccgggttgcgaagtcagggcattcc
gccgatctagccatcgTTTtTttTtgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaatccaatcccgatcccta
gcccgatcccaatcccaatcccaatcccttgtcctttttcattagaaagtcataaaaaAacataataatgatgtcgaagggattaggggcgcgc
aggtccaggcaacgcaattaacggactagcgaactgggttGttttttttgcgccgacttagccctgatccgcgagcttaacccgttttgagccg
ggcagcaggtagttTtTTTtggaccccacgTttttttttggccaaacctccaagctaacttgcgcaagtggcaagtggccgTtttTTtggcc
caaaagaggaggcactatcccggtcctggtacagttggtacgctAAAaaAAattatatcatcataataaatgttttgcccaacgaaaccga
aaactttcaaattaagtcccggcaactgggttcccattttccattttccatgttctgcgggcaggggcggccattatctcgct

>*Kr proximal mut Hb construct 1*, DePace stock #0663
GTAAGTTCAGGTGATTTGGACCCAAGCAACGGCCTCGACCATACGCTCTTTACACGCACTTCTG
CCCTATTCACTCGGAAATCTTTAGTTCGCTTTGGGATTAAAAATCTAGTCTATGTTTCATCTTCT
CGCTTTTTTGCTATTCTACTCTACTGCTTGCCGACAAATTCCATCTTTTCTAGGCCCGGATCTAT
GTAGCATCTGTTGACTTTTTGTGGATCTTTAGCTCAAAAGCCTCTGTATGTTAAGCAAACAACC
AAAAAAAATTTCCGACTAAGCGCCGAAACATAGGAATCGAATGAAGATTGCAACATTCCGGTC
AGCTCAGCCCGAAATTTCAGAATTGCAGTGATCTAGGGTCAGTTCGAAGCGCTTTGCTTGAATA
TTCAGTATTGCCTTCTGCGAATTTTGTTTTGGTGAAAAACTTGCATGGGGGTCGAACATTCGAT
GACAAAACCAGTACAACTTTCCCGTACGTGATGCAGAGCAAGGATATTTGTGTGCCTTTCTGCT
TTTGTTTGGTTTCACTTTCTTTAAATCCACGGTAGTTCAAATTTTGCACCCGGTGAGCTCTGAG
CGTCCTTGGCCTACCTACAAAACCAAAAATACCGATTTTCAACAAAGAACAGTGACAGCATGTC
CTTGCCAGAAAAATCTACCTTTAAATATTTAAAAGTCAATCCGCAAAAGTCACAGAAAACGTCT
GTACTAGCAAGAAAATTTCAACGATGCAGCTCTCGATTGTTGTCTAGGAACACTTCCTTTCTAC
CTGAATCCGTTTTTTTCTAGTTCCTTTTTAAGATTTCTTTTTTAGACGTGTGCCAAGATAAAGTT
GTCCATTCTGCTGAACTTCCACGTCTTTGAGACTTTGCTCAACAGGCTACAAGGAAGACAAAAA
ACTCACTCACTAGAGCTGGCAAAAAACCGTCGACCTTCTTTCTTACTTTCTGCACTCCTTGAGCC
GATGTTTTGTCTACTTCCGATTTTCTCGTACAGCGCACGCTTGAAATTCAACCCCGTCCGTGAC
ATCTCTCACAAACAATTCTGGCTCTTTTGCTCTCGCTGTTTCAAGTTAGCCCTGTTCACAATCGG
CGTTCATCCCTGGTATGCCAGCAGGGGATTTTCTATCGCTCAAATGTAAAAAGGGAAAACGAC
AAAGGCACATCTGAGTGAGTTTTTGTTGTGCGTCCGTTTGGTTCTACAAAAACTTCTACGTGAC
TTGAGCGTGGTGTACTATTCCTGCAAGTGCACAAAAAACTACGGCTTTAAAAAACACAAGCGG
AATCGATTCCGATGGTTCAAAAATGTCCCACAATGGTCAACACATAGTTCTGATTTTTTTTTTTG
TCTTGGGGTATAGAAACCTTAAGGTGAAAATACCCTCTACTCAGCGAATGATATGCGCATCGAC
AACCTATGTTAAAAAACCACATCGTACATATTTGACTCTAGGTGAGTATATCTCTTTTGCATTG
AAAGGAACCTCTAGCTGTCTCATTCGCACAAATTTGCCTATGCAAAACTAACACTTTTAGGTAC
AAAAAGGAATCGATGTACTCTCCTCACTCCACTCAGCGAAGTACCCA

>*Kr proximal mut Hb construct 2*, DePace stock #0664
GTAAGTTCCCATAGTTCGGACTTGTACAGCGGCCTCGACCATGCTGAGTAAGCACGCACTTCTG
CCCCATTCCCTCGGAAATCTTTAGTTCGCTTTGCCAAGCCTTATCTTGTGCTTCTTTCTATTCAT
CGCTTTGACACTATTCTCGATTGGTTTTAGCCGACAAAAGTATTCTTGAATATGCCCGGAACGT
AAGAGGATCTGTTGAGCATGAGTGGATCTTCCGTTGAGAAGCCTCTGCTTACCACGCAAACAA
CCAAAAAAAATTTCTTCCGAATCGCCGAAACATAGGAATCGCATGAAGATTAGAAAAAACCGGT
CAGCTCAGCCCGAACTTGAAGGATTTTTGACCAATGGGGTCAGTGTTAAGCGCTTTGCTTTCAG
TTGATCTGATGCCTTCTGCGAATTTTGTTTTGGTGAAAAACTTTCTTGGAAGTCGAACATTCGA
TGACAAAACATTGAGCAGGATATTTAAGACTATGCAGAGCAAGGATATTTGTGTGCCTTTCTGC
TTTTGTTTGGTTTCACTTTTTAAAAATCCACGGTAGTGGATAAATACCACCCGGTGAGCTCTGA
GCGTCCTTGGCCTACCTACAAAACCAAAAATACCGATTTTCAACATAGATTTAACATAGCATGT
CCTTGCCATCCAAATCTACATACACTTCTTTAAAAGTCAATCCGCAAAAGTCACAGAAAACGTC
AACACAAGCAAGAAAATTTCAACGATGCAGCTCTCTTTCAACTTTTTTCAAATATTCCTTTCTAC
CTGAATCGGTTTGGGTATTGTTCCTTCTCACAAGTTCTTTTTTAGACGTGTGCGAATACCCCAC
TGACCTTTAAGCTGAACTTCCACGTCTTTGAGACTTTGCTCAACAGGCTACAAGGAAGACAAAA
AACTCCTTGTGCATTGCAGGTAAAAAACCGTCGACCTTCTTTCTTACTTTCTGCACGCCTTAAGC
CGATGTTTTTTGTACTTCCTGTGTACTCGTACAGCGCACGCTTGAAATTCAACCCCGTCCGTGA
CATCTCTCGTAGTGCATTTTGGCTCTTTTGCTCTCGCTGTTTCAGCATTGCCCTGTTCTCCATTG
CCGTTCATCCCGCTTAGACCCGTTGTCTGTTGTCTAAAGCTCAAATGTAAAAAGGGGGAATGAC
AAAGGCACATCTGAGTGAGTTTTTGTTGTGCGTCCTCGAGTCACTCTACAAACATCTACGGGAT
TGAATCGTGGTGTAAAATACCTGCAAGTGCAATTCTCAAAAAGGCACTGTTTTACACAAGCGGA
AGCGGTTCATGCAACTCAAAAATGTCTTGTACCGGTCAACACATAGTTCTGATTCTAGATGTCT
GAAGAGTGACTAGAAACATACATGACAAATGCCCCTCTACTCACTCGTTTTTTTGTGCATCGAC
AACCTATGTTAAAAAACCACATCGTCCATGTTTGACTCTACCTGAATGTATATCTTTTGCATTGA
AAGGAACCTCTAGCTGTCTCATTCGCACCTGTTTGCATTTGAAAAATCAGCACTTTTAGGTATC
AAAAGGAATCGATGAAATCAGCTCCCCCAACTCAGCGAAGTACCCA

>*Kr proximal mut Cad construct 1*, DePace stock #0665
GTAAGTTCCCATATTTCGGACCTTATCAGCGGCCTCGACCATACGCACTTTACACGCACTTCTG
CCCCATTCACTCGGAAATCTTTAGTTCGCTTTGCCAATCAAAATCTAGACTATATGTATTTCACT
CGCTTTGATACTATTCTAGTTTAGTGATTGCCGACAAAGTCATTAACGATAAGGCCCGGATCTT
ATTAAAATCTGTTGAGCTTAAGTGGATCTTTCAATATCAAGCCTCTGCTTATCACGCTAACAAC
CAGAAAAATTTTCTGACTAATCGCCGAAACATAGGAATCGTATGAAGATTTTCTAAAACCGGTC
AGCTCAGCCCGAAATTTAAGAATTTCTTATCAATAGGGTCAGTTTTAAGCGCTTTGCTTGCATA
TTATTTAATGCCTTCTGCGAAATTTGTTTTGGTGAAAAACTTTCTATGAAGTCGAACATTCGAT
GACAAAACCATTATAATTTTACGTAACATTATGCAGAGCAAGGATATTTGTGTGCCTTTCTATT
TATGTTAGGTTTAACTTTCTTAAAATCCACGGTAGTTCAAATTTACCACCCGGTGAGCTCTGAG
CGTCCTTGGCCTACCTTCAAAACCCAATAATCCGATTTTCAACATAAATTTAACATAGCATGTCC
TTGCCCTAAATATCTACATAAATCCATTTAAAAGTCAATCCGCAAAAGTCACTGAAAACGTCTA
TACAAGCAAGAAAATTTCAACGATGCAGCTCTCTTTCAATTCTTAATTATATTTCCTTTCTACCT
GAATTCGGTCTTTACTAGTTTATATTTACGATTTCGTTCTAAGACGTGTGCCAATAATTAGTTGT
CCATTAAACTGAACTTCCACGTCTTTGAGACTTTGCTCAACAGGCTACAAGGAAGATTAGAAAC
ACACTCAGTATAGCTGGGAAAACGACGTCGACCATCTATCTTACTTTCTGCACTCCTTAAGCCG
AAGTTATGTTTACTTCCTGTTTACTCGTACAGCGCACGCTTGAAATTCAACCCCGTCCGTGACA
TCTCTTTTCATGAATTTAGGCTCTTTTGCTCTCGCTGTTTCATTTTCGCCCTGTTCTCCATTGGC

GTTCATCCCCCTTAGACACATTACTTATTCTTCATAGCTCAAATGTAATAAGGGGAAATGACAA
AGGCACATCTGAGTACGTTATTGTGGTGCGTCCCTTACTAACTATAAAGATCTCTACGTGATCT
TTATGTGGTGTAATATTCCTGCAAGTGACACTTATATGAAGGCTTACGATAACACAAGCGGAAG
TGGTCATAAGCATGAACAAATAGCCTTTAACGGTCAACACATAGTTCTGATATTGTAAGTTTTA
TTACTGACTAGAAACGTAAAGGTCATGAAACTCTCTACTCAGCGAAGATAATTCGCATCGACAA
CCTATGTTAAAAGTCCACATCGTCCATATTTGACTCTAACTGACTATATATCTTATGCATTGAAA
GGAACCTCTAGCTGTCTCATTCGCACCAATTTGCATATGAAATATTAGCACTTTTAGGTATAAA
GAGGAATCGATGAAATCAACTCACTCAACTCAGCGAAGTACCCA

>*Kr proximal mut Cad construct 2*, DePace stock #0666
GTAAGTTCCCATATTTCGGACCTTATCAGCGGCCTCGACCATACGCACTTTACACGCACTTCTG
CCCCATTCACTCGGAAATCTTTAGTTCGCTTTGCCAATCAAAATCTAATCTATATTTCGTTTACT
CGCTTTGATACTATTCTAGTTTAGTGATTGCCGACAAAGTCATTAACGATAAGGCCCGGATCTT
ATTAAAATCTGTTGAGCTTAAGTGGATCTTTCAATATCAAGCCTCTGCTTATCACGCATACAAC
CAAATAAAATTTCTGACTAATCGCCGAAACATAGGAATCGTATGAAGATTTTCTAAAACCGGTC
AGCTCAGCCCGAAATTTAAGAATTTCTTATCAATAGGGTCAGTTTTAAGCGCTTTGCTTGCATA
TTATTTAATGCCTTCTGCGAATTTTGCTTTGGTGAAAAACTTTCTATGAAGTCGAACATTCGAT
GACAAAACCATTATAATTTTACGTAACATTATGCAGAGCAAGGATATTTGTGTGCCTTTCTGCT
TATGTTTGGTTTGTGTTCATTAAAATCCACGGTAGTTCAAATTTACCACCCGGTGAGCTCTGAG
CGTCCTTGGCCTACCTACAATACATCACATACCGATTTTCAACATAAATTTAACATAGCATGTCC
TTGCCATACCAATCTACATAAATCCTTTTAAAAGTCAATCCGCAAAAGTCATACTAAAAGTCTAT
ACAAGCAAGAAAATTTCAACGATGCAGCTCTCTTACAATGTTTTATTATATTTCCTTTCTACCTG
AATCCTTATGTTACAAGTTTCTTCTTACGATTTATTGAATAGACGTGTGCCAATAATTAGTTGTC
CATTAAACTGAACTTCCACGTCTTTGAGACTTTGCTCAACAGGCTACAAGGAAGACTGAAAGGT
CACTCAGTATAGCTGGTAAATAACCTTCGACCTTCTTTCCTACTTTCTGCACTCCTTAAGCCGAT
GGTCTGTTTACTTCCTGTTTACTCGTACAGCGCACGCTTGAAATTCAACCCCGTCCGTGACATC
TCTTTTCATGAATTTGGGCCCTTTTGCTCTCGCTGTTTCATTTTCGCCCTGTTCTCCATTGGCGT
TCATCCCCCTTAGACACATTAATTATCTGCTATAGCTCAAATGTTAAAAGGGGAAATGACAAAG
GCACATCTGAGTGAATTATTGTTGTGCGTCCCTTACTAACTCTAAAATCTTCTACGTGATCTTTA
TGTGGTGTAATATTCCTGCAAGTGTAGTTCTTATGAAGGCTCTAAATCACACAAGCGGAAGTG
GTCATAAGCATTGAAAACTGTCCTTTAACGGTCAACACATAGTTATGTGTTATTAAGTTTTATT
ACTGACTAGAAACGTAAAGGTCATAATGACCTCTACTCAGCGAAGATAATTCGCATCGACAACC
TATGTTCTAAAGCCACATCGTCCATATTTGACTCTAACTGACTATATATCTTTAGCATTGAAAGG
AACCTCTAGCTGTCTCATTCGCACCAATTTGCATATGAAAACTTCGCACTTTTAGGTATAACATT
GAATCGATGAAATCAACTCACTCAACTCAGCGAAGTACCCA

>*Kr proximal mut Hb mut Cad construct 1*, DePace stock #0667
GTAAGTTCTCATACTTCGGACCTATTCAGCGGCCTCGACCATTCGGACTTTACACGCACTTCTG
CCCCATGCACTCGGAAATCTTTAGTTCGCTTTGCAAAGCAAAATCTAGTCGATATCTTGGTTAC
TCGCTTTGGTATGATTCTTGTGTATTTGAAGCCGACATAGAGAAGTGTAGAAAGGCCCGGAGC
TTAAGCAAGTAGAATGAGCATATGTGGATCTTTCATGTTCAAGCCTCTGCCTCACAGCATCACA
GCCAAAAGAAATTTCTGTCTTAACACCGAAACATAGGAATCGATTGAAGATTTTCTAACACCGG
TCAGCTCAGCCAACTAAAAGAGAATTTCTAACCGATAGGGTCCGTCTTAAGCGCTTTGCTTCTA
CCCTAGTTGGTGCCTTCTGCGAGATTTATCTTGGTGAAAAACTTTCTCAGAAGTCGAACATTCG
ATGAGTGAAATGTAAAACTTTTAAATACCAAGATGCAGAGCAAGGATATTTGTGTGCCTTTCTC

GGTCTGAGGAGTTGCCCTGTCTTCAAATCCACGGTAGTTCAGGTTGGCCACCCGGTGAGCTCT
GAGCGTCCTTGGCCTACCTTCTTCTGACAGACAACTGATTTTCAACACAAGTTTAACATAGCAT
GTCCTTGGAATAGAACTGTACATAGAACCGTCTAAAAGTCAATCCGCAAAAGTCACAGAAATCT
AAATACCAAGCAAGAAAATTTCAACGATGCAGCTATCTTAAAATTTTGCAGCTTAGTTCCTTTC
TACCTGAATACGTGGATCGCTATGTCCTTTGTACGATGTCGATTTTGGACGTGTGCAGATAAAG
AGTTCCGCGACTTCTCAAACTTCCACGTCTTTGAGACTTTGCTCAACAGGCTACAAGGAAGACA
AAATTCGCACTCAGTATTGGAGTTACGCAAACATCATGATTCGACGGTACTTTCAGAAATTTGT
TCGCCTATGTCTAGTGTACATCCTGTCTACTCGTACAGCGCACGCTTGAAATTCAACCCCGTCC
GTGACATCTCTCTTCATGCATTTTGGTATTCTTGCTCTCGCTGTTTGCTTTCCGCCCTGTTCTAA
GTTGACGTTCATCCCCCCGTAGACACATTGGCTGATGTCTATCGCTCACATGGATTTAACATTGT
ACTCTAAGGCACATCTGAGTGAGTTTTCGTTGTGCGATCCTTCTAGAATCTAGAAACTTCTTGT
AGCCAAAAATTTGGTGTAAAATGCCTGCAAGTGCAGTGACGATATGTGTTCATCGTGTATAGA
AAGAAAGTGGTCACAAGTATGAGTAGATGCCCTTCCACGGTCAACACATAGTCGAGTTTCACTC
TGTGATATCAGTGACTAGAAACCTAAAGCTCTACAAACTCTCTACTCAGCGAAGATAACTCGCA
TCGACAACCTATGTTAAAGACCTACATCGTCCATCTGTGTATACACTGGAAGCTAAAGGTTTTG
CATTGAAAGGAACCTCTAGCTGTCTCATTCGCACGAACTTGCACATGAGAGAATGGAACTTTTA
GGTAGACAAAGGAATCGATGAAATCAGCTGCCTCAACTCAGCGAAGTACCCA

>*Kr proximal mut Hb mut Cad construct 2*, DePace stock #0668
GTAAGTTCCCAAATGTCGGACCTTATCTGCGGCCTCGACCATACGCATTTTACACGCACTTCTG
CCCCAGTCACTCGGAAATCTTTAGTTCGCTTTGCCAATCCAAATCTAGTACACGGCTCGTCTATT
CGCTTTGAATCTATTCTGGTTTAGCGTTTGCCGACAAAGTCAAAGACAAAAGTGCCCGGATCTT
GTGAAAATCTGTTGAGCCTGTTGAAATCTTTCTAGAACAAGCCTCTGGTTATCTCGCAACGCAA
GCAAGTAACTCGATGACTAGTGGCCGAAACATAGGAAACGTATGAAGATTTTCTGAAACCGGT
CAGCTCAGCCCCAATACTTTTGCTAGATTTACCTTGAAGACAGTCTTAAGCGCTTTGCTTGCGG
CTTTAAATCTGCCTTCTGCGAATTTTGTTTCGGTGAAAAGATATCTATCGAGTCGAACATTCGA
TGAGGAAACATGAAATATCTTACGTAACATTCAGCAGAGCAAGGATATTTGTGTGCCTTTCTGC
CTATGTTTGGTTTTGGGTTGTGATCACCCACGGTAGTTCAAATATTCCACCCGGTGAGCTCTGA
GCGTCCTTGGCCTACCTACAAGATCTTACATTCTGATTTTCAGCCTGAATTTGACATAGCATGT
CCTTGCGTGTGATATCTACAGAACGACAGTGTATAGTCATTCCGCAAAAGTCACGGAAAACGT
GTACAAAGGCAAGAAAATTTCAACGATGCAGCTCTCGGATAACTTTCCACAAGATTTCCTTTCT
ACCTGAATTTGTGATACAGTAGTTCCTTGTCTCGATGTCGACCTTAGACGTGTGCATACAGTTG
GTTGTCACAAGCTATGGACTTCCACGTCTTTGAGACTTTGCTCAACAGGCTACAAGGAAGACCA
ACAACTCACTCAGAATGGCTGGTACTCAGCCGTCGACCATCTTACCTACTTTCTGCACACCTTA
AGCAGATGTTTTGCTTACTTCTGGTTTACACGTACAGCGCACGCTTGAAATTCAACCCCGTCCG
TGACATCTCTTTTCGTGAATTTTGGCTCTTTAGCTCTTCCTGGTTAGAATTCGCCCTGTTCTTCA
ATGTCGTTCATCCCCCCTACACACATTTGATATTTTCAAGAGCTCAAATGTAAGAAGGGGTGAT
GCCAAAGGCACATCTGAGTGTGTTCTTGTTGTGCGTCCCTACGAAAGTAAGAATACTTCTACGT
GATCGGAATGTGGTGTACTATTCCTGCAAGTGTAGTAGTCTTGAAGGCTCTGAAACAGACAAG
CGGAAGTGGCTAAAGTTACCTCGGCAGAAGCTTAAACCGGTAACACATAGTTCTGATGTAAGA
TGACACATTGTGGACTAGAAACGTGAAGGGCTTGATACCCTCTACTCAGCGAAGATAATACGC
ATCGACAACCTATGTGATCTAACCACATTGTACTAGTTCGACGCAGCTGAGTAGCTGGTCTTTC
GCATTGAAAGGAACCTCTAGCTGTCTCATTCGCACCCATTGGTATGTGCATTGCTAGCACCTTA
AAGAGTCATGCCAATTCGATGTACGAATTTCACTCACCTCAGCGAAGTACCCA

>*eve3+7 and eve2* (endogenous sequence between and including *eve3+7* and *eve2*); DePace stock #0202

ggatcctcgaaatcgagagcgacctcgctgcattagaaaactagatcagttttttgttttggccgaccgatttttgtgcccggtgctctctttacg
gtttatggccgcgttcccatttcccagcttctttgttccgggctcagaaatctgtatggaattatggtatatgcagatttttatgggtcccggcgat
ccggttcgcggaacgggagtgtcctgccgcgagaggtcctcgccggcgatccttgtcgcccgtattaggaaagtagatcacgttttttgttccc
attgtgcgctttttcgctgcgctagttttttccccgaacccagcgaactgctctaatttttaattcttcacggctttcattgggctcctggaaa
aacgcggacaaggttataacgctctacttacctgcaattgtggccataactcgcactgctctcgtttttaagatccgtttgtttgtgtttgtttgtc
cgcgatggcattcacgtttttacgagctcgttccttcgggtccaaaattatgccagtttgttttgtctctggcaattattggaaatttcattgggtc
gatttcgctgccttccttgctcttcccttgagaaaagtgaataggttgtgccataaaaatcgctgctcctgaagaccaaatgaaatggatttgtg
taagcattaaaaacgcgaggcaagccccaagattcctccactgctttttttatattgcccactgctaaatgcagctaattcgtcgattgtttaaa
aattaaattacttatgttgccattcatacatcccctcacattttatggccatttgagtgcggggtgcacagttctgtcttaagtggcggatggaaa
ccaccacatttactcgagggatgatgtgctctaatatctcctcatcaaatgggatggtttctatggaaaggcaaaatcgttgtaaagtgaggcg
gagttaaaaaataccttgttatagccttttaaaataacacaagatcgttcgaattgactagaaatatcaaagtctttttgtattgaagcgagtg
tagtctcaatttatgcttaattttaagaaatacatctctttattagccccaaaatgaaacaaatggtctactaattaagcaagtcaacagaatttt
tatgcaattattcaaaatgaaataatatatacataagatgtttttgggaatctgtcatggggtttctgaaataggtttgccaaacaaatttaagt
ataaatgtatacatatgtcaactaataaatttagcaaataaaatgtacctgcaagtatctataaatttattggaccaattttgtgtaaaaaactg
aactggcactcttcccaagaatgggacttcgaggactccttgctgaatcacttactcaacccattccaactcatccaatccgcgcaatcatcat
aaattttggccttttgttgtaattgttttatggcagaaattactcaatcatcaagcataattccctcgttttcgccgttttattgccaatttttgcact
gcctttgccttttcccgcccttcctcagcgttttgcgaatctttgccggcatttctattgcgcggacaatccggccagtgtgttggccatttactt
gccatgatgacgggcataatcagcgagatcggcgctttgtgagtgcagaatgtgcaataaagcggcaacaatcggcagggattcgccttcc
catattccgggtattgccggcccgggaaaatgcgaaagtgtttgcggatcgagatggaagatagaggattgagtattgaaacgaggaaggt
acttccgccggcggacactttcgcctaaccaagccaatccaacccatcccaatccaatccaacccacccgatcgccataaagggtatttactg
tcgctgccgcagagcctcgcttgacgacttaacccaagcggtcgtttcgcgtccattctccggacggagtcaaagacaaaggccggcggagg
tggacaataggcaaggttgttgcttgtgggtagggtttgagctatgagctatgagctgtgagctgttagccctgaaccccgaacctcgagaat
tgaacctttcccggggcaagaaggcttgcatgtgggccttttccaggtcggccagtaggtagagttgttgcgatgcggctatgccgggcgagt
taatgccaatgcaaattgcgggcgcaatataacccaataatttgaagtaactggcaggagcgaggtatccttcctggttacccggtactgcat
aacaatggaacccgaaccgtaactgggacagatcgaaaagctggcctggtttctcgctgtgtgtgccgtgttaatccgtttgccatcagcgag
attattagtcaattgcagttgcagcgtttcgctttcgtcctcgtttcactttcgagttagactttattgcagcatcttgaacaatcgtcgcagtttgg
taacacgctgtgccatactttcatttagacggaatcgagggaccctggactataatcgcacaacgagaccgggttgcgaagtcagggcattc
cgccgatctagccatcgccatcttctgcgggcgtttgtttgtttgtttgctgggattagccaagggcttgacttggaatccaatcccgatccctag
cccgatcccaatcccaatcccaatcccttgtccttttcattagaaagtcataaaaacacataataatgatgtcgaagggattagggg

Vector and Guide RNA Sequences

dCas9 expression plasmid insert (cloned into pBDP for integration into attP2, DePace stock #0662*)*

gatctGCTTCGACCGTTTTAACCTCGAAATATGCACATGTAAGGACGGATGTGAGCGAACGCCAG
TGATGACCGGGATCAGAGGTAACCTACCATGGTGGGGATTAGGTGACCGTTCGCAGGTAGTTT
GATCGGAGCGAATGTTCGGGGGGTCTGGCGTCAGAGGCTCTAAACTTTATGTAATTCCTGCCG
CGAAACACGCACGTATCAAGCAGTCAGCTGTTCTCTTCGTTCAGCGCGCGCCGGTGTTGCAAAA
CGAGCGCTCTTCGCCGGCGGTGGCTCGTGCGATAGTTCGTTTTGTCGGTAATCCGATGTTGCC
GCGCCGATATCATGTGATGTTGTCACAGTGCGCGAAATTCGAATGGTGGTGTGCAGTGATTGT
GTTGTGACGGCGAGTGGCGCGTGTGGGTGCTTAGTTTTGGGAGATGTTTTCGTATTTTTTTGT

TGATAACTCAGGCTTTGTTGCTGTGTTGTAGTACTATTTTCCATTGCGCGGTGTCCAGCTTTTA
ATTAGTGGCACATATTCTTAGCAAGTAAAAATTATTTTGCATACTATTAAATTTCTTATAAATTA
TTTTCTAAAATTAAGTTTACCTTTTCAATTTTACTAAAAATATCGATATATTTATTATCGCTGGA
AAACTACATTATTCCACCTCTAAGCAAGAACCGtctagaAAAAGGTAGGTTCAACCACTGATGCCT
AGGCACACCGAAACGACTAACCCTAATTCTTATCCTTTACTTCAGcctgcaggATGAAGCGTCCTGC
TGCTACTAAGAAAGCTGGTCAAGCTAAGAAAAAGAAAGACAAGAAGTACAGCATCGGCCTGGC
CATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAA
ATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGCGCCCTGCT
GTTCGACAGCGGAGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACA
CCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGG
ACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGC
GGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCT
ACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGG
CCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACA
ACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAA
ACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCA
GACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACC
TGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATG
CCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCG
GCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCG
ACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGA
AGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATCGATGGCGGAG
CCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGG
AACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCA
GCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTT
ACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTA
CGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAA
CCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCG
AGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGC
TGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGA
GAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCA
ACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGC
TGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATA
TCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATG
CCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCA
GGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATT
TCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGAC
CTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACAT
TGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGA
CGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAG
AGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAG
GGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAG
AACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTG

GACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGAC
GACTCCATCGATAACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGCTG
ATTACCCAGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGAT
AAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAG
ATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGGGAAGTGAAA
GTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGC
GCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCC
TGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACG
TGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCT
ACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGC
GGCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGTGGGATAAGGGCCGGGACTTT
GCCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAG
ACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGACAAGCTGATCGCCAGA
AAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTG
CTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCT
GGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAA
GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCT
GGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGG
CCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTC
CCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAACACTACCTGGACGAGAT
CATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAGGT
GCTGAGCGCCTACAACAAGCACAGAGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCA
CCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGAC
CGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACC
GGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACGGatcCGGTGGCGGAGGG
ATGGAAAAAGCGACACCGGAGGACGATGGTCCATTGGATTTGTCTGAAGATGGAGCCAGCTCT
GTGGATGGCCATTGCAGCAACATCGCACGGCGCAAGGCACAGGACATTCGTCGGGTTTTCCGG
CTGCCTCCACCGCAAATCCCTCACGTACCCAGTGATATGCCTGAGCAAACCGAGCCAGAGGATT
TGAGCATGCACTCTCCTCGCTCTATCGGATCTCACGAGCAAACCGATGATATTGACTTGTATGA
TTTAGATGATGCCCCGGCTTCTTATATGGGCCATCAACAACATtacccatacgacgtcccagactacgctTA
GaagcTTGGCATCAGgtaggcatcacacacgattaacaacccctaaaaatacactttgaaaatattgaaaatatgtttttgtatacatt
tttgatattttcaaacaatacgcagttataaaactcattagctaacccatttttctttgcttatgcttacaggctagcGATCTTTGTGAA
GGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAG
GTAAATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTA
GATTCCAACCTATGGAACTGATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAAACCT
GTTTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACT
CCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTT
TGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAA
AGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTTGATGTATAGTGCCTTGACTAGAGAT
CATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCC
TGAACCTGAAACATAAAATGAATGGAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGG
TTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTT
GTGGTTTGTCCAAACATCACCGGCCTGTACGAGACCTGGTTCCAgc

Features: *nos* promoter intervening sequence nuclear localization signal dCas9 coding sequence 5x glycine linker Kr repression domain HA tag stop codon *ftz* intron SV40 untranslated region

Primer sequences for eve stripe 2 gRNA synthesis

>tracrRNA template, no terminal complementary base
AAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATG
CTTTTCAGCATAGCTCTAAAAC

>tracrRNA template, complementary A
AAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATG
CTTTTCAGCATAGCTCTAAAACA

>tracrRNA template, complementary C
AAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATG
CTTTTCAGCATAGCTCTAAAACC

>tracrRNA template, complementary G
AAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATG
CTTTTCAGCATAGCTCTAAAACG

>tracrRNA template, complementary T
AAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATG
CTTTTCAGCATAGCTCTAAAACT

>gRNA1R
CGATGTAATACGACTCACTATAGGattgcgcccgcaatttgCATGTTTTAGAGCTATGCTGAAA

>gRNA2
CGATGTAATACGACTCACTATAGGacccaataatttgaagtaacGTTTTAGAGCTATGCTGAAA

>gRNA3
CGATGTAATACGACTCACTATAGGcaggagcgaggtatccttccGTTTTAGAGCTATGCTGAAA

>gRNA4R
CGATGTAATACGACTCACTATAGGttatgcagtaccgggtaaccGTTTTAGAGCTATGCTGAAA

>gRNA5
CGATGTAATACGACTCACTATAGGatggaacccgaaccgtaactGTTTTAGAGCTATGCTGAAA

>gRNA6
CGATGTAATACGACTCACTATAGGgacagatcgaaaagctggccGTTTTAGAGCTATGCTGAAA

>gRNA7R
CGATGTAATACGACTCACTATAGGgcacacacagcgagaaaccGTTTTAGAGCTATGCTGAAA

>gRNA8R
CGATGTAATACGACTCACTATAGGataatctcgctgatggcaaaGTTTTAGAGCTATGCTGAAA

>gRNA9R
CGATGTAATACGACTCACTATAGGtctaactcgaaagtgaaacgGTTTTAGAGCTATGCTGAAA

>gRNA10
CGATGTAATACGACTCACTATAGGcttgaacaatcgtcgcagttGTTTTAGAGCTATGCTGAAA

>gRNA11
CGATGTAATACGACTCACTATAGGgtgccatactttcatttagaGTTTTAGAGCTATGCTGAAA

>gRNA12
CGATGTAATACGACTCACTATAGGtagacggaatcgagggacccGTTTTAGAGCTATGCTGAAA

>gRNA13
CGATGTAATACGACTCACTATAGGctataatcgcacaacgagacGTTTTAGAGCTATGCTGAAA

>gRNA14R
CGATGTAATACGACTCACTATAGGaatgccctgacttcgcaaccGTTTTAGAGCTATGCTGAAA

>gRNA15R
CGATGTAATACGACTCACTATAGGgatggcgatggctagatcggGTTTTAGAGCTATGCTGAAA

>gRNA16
CGATGTAATACGACTCACTATAGGgcgtttgtttgtttgtttgcGTTTTAGAGCTATGCTGAAA

>gRNA17
CGATGTAATACGACTCACTATAGGgattagccaagggcttgactGTTTTAGAGCTATGCTGAAA

>gRNA18R
CGATGTAATACGACTCACTATAGGatcgggctagggatcgggatGTTTTAGAGCTATGCTGAAA

>gRNA19R
CGATGTAATACGACTCACTATAGGaagggattgggattgggattGTTTTAGAGCTATGCTGAAA

>gRNA20R
CGATGTAATACGACTCACTATAGGtttatgactttctaatgaaaGTTTTAGAGCTATGCTGAAA

>gRNA21
CGATGTAATACGACTCACTATAGGacacataataatgatgtcgaGTTTTAGAGCTATGCTGAAA

>gRNA22
CGATGTAATACGACTCACTATAGGgattaggggcgcgcAGGTCCGTTTTAGAGCTATGCTGAAA

References

Estrada, J. et al., 2016. SiteOut: An Online Tool to Design Binding Site-Free DNA Sequences. PloS one, 11(3), p.e0151740.

Garcia, H.G. et al., 2013. Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Current biology, 23(21), pp.2140–2145.

Hou, X.S., Melnick, M.B. & Perrimon, N., 1996. Marelle acts downstream of the Drosophila HOP/JAK kinase and encodes a protein similar to the mammalian STATs. Cell, 84(3), pp.411–419.

Liang, H.-L. et al., 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature, 456(7220), pp.400–403.

Noyes, M.B. et al., 2008. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nucleic acids research, 36(8), pp.2547–2560.

Wu, Y. et al., 2013. Spatially isotropic four-dimensional imaging with dual-view plane illumination microscopy. Nature biotechnology, 31(11), pp.1032–1038.

APPENDIX F: YEARLY PLANNING MEETINGS: INDIVIDUALIZED DEVELOPMENT PLANS AREN'T JUST MORE PAPERWORK

Ben J. Vincent, Clarissa Scholes, Max V. Staller, Zeba Wunderlich, Javier Estrada, Jeehae Park, Meghan D.J. Bragdon, Francheska Lopez Rivera, Kelly M. Biette, and Angela H. DePace

This work was published in 2015 as the equal contribution of all authors. While its subject matter is distinct from the rest of the dissertation, the article articulates scientific approaches and values that made possible all work described in previous chapters. It is reprinted from Molecular Cell, 58(5), Vincent et al., Yearly planning meetings: individualized development plans aren't just more paperwork, pp.718–721, Copyright 2015 with permission from Elsevier.

Molecular Cell
Forum

Yearly Planning Meetings: Individualized Development Plans Aren't Just More Paperwork

Ben J. Vincent,[1,2] Clarissa Scholes,[1,2] Max V. Staller,[1,2,3] Zeba Wunderlich,[1,2,4] Javier Estrada,[1,2] Jeehae Park,[1,2] Meghan D.J. Bragdon,[1,2] Francheska Lopez Rivera,[1,2] Kelly M. Biette,[1,2] and Angela H. DePace[1,2,*]
[1]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[2]These authors contributed equally to this manuscript
[3]Present address: Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108, USA
[4]Present address: Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92697, USA
*Correspondence: angela_depace@hms.harvard.edu
http://dx.doi.org/10.1016/j.molcel.2015.04.025

The National Institutes of Health (NIH) encourages trainees to make Individualized Development Plans to help them prepare for academic and nonacademic careers. We describe our approach to building an Individualized Development Plan, the reasons we find them useful and empowering for both PIs and trainees, and resources to help other labs implement them constructively.

In the current research climate, there are too many trainees for very few academic positions, and ignoring this fact may lead to many missed opportunities to productively shape the training experience (Polka, 2014a, 2014b; NIH, 2012; National Academy of Science, 2014; Sauermann and Roach, 2012; Bourne, 2013). Having a career development plan in place will help trainees to prepare themselves for fulfilling careers both within and outside academia. In our lab, we have been making career development plans for all trainees, including students, postdocs, and staff, for the last 7 years; we all enjoy the process and see it as valuable. We see the new requirement from the National Institutes of Health (2014a) as an opportunity to discuss how to mentor trainees and provide leadership skills that positively influence lab culture.

The Goals

Every year, we have a one-on-one meeting between the PI and each trainee to accomplish the following:

1. Motivate people by celebrating their accomplishments
2. Set short-term and long-term research and career goals
3. Help people make rapid progress by prioritizing projects and identifying barriers
4. Clarify and solidify relationships by giving honest constructive criticism
5. Clarify expectations in both directions and address any disagreements

In the process, we generate documents that are consistent with the requirements for an individualized development plan (IDP)—a defined career goal and steps to attain it. It's worth noting that we have never called the meetings or the resulting documents an IDP; giving it an acronym makes it feel uncomfortably bureaucratic to us. We just call them yearly planning meetings.

Conducting yearly planning meetings helps the lab run smoothly. People are excited about their projects and feel supported in meeting their goals. The meetings also model a way to give constructive feedback and resolve conflicts, which has far-reaching benefits in the lab. In Figure 1 we describe our guidelines for giving constructive criticism in detail.

The Process

Our yearly planning meetings take about an hour and are organized around two very simple worksheets (blank and completed examples are available in the Supplemental Information, available online). But here's the key: the worksheets are not the point. They help to organize the conversation by allowing people to gather their thoughts beforehand, remember what to say during the meeting, and have a record of what was discussed. The point is to have a real conversation, not to check boxes to satisfy the NIH.

Before the meeting, both the PI and the trainee fill out the Goals and Planning Worksheet; this gives both parties an opportunity to think through all of the categories on their own so that together they can be complete, and identify any discrepancies. During the meeting, the PI and the trainee go through the Goals and Planning Worksheet together and then jointly fill out the Calendar. The goal is to have a rough and flexible road map for the year and to come to a reasonable estimate for how long things will take. The major benefit of filling out the calendar together is that both parties agree on the basic plan. At the end of the meeting, the PI and the trainee will have two amended copies of the Goals and Planning Worksheet and one filled-out version of the Calendar. All completed worksheets are given to the trainee at the end to copy or scan and return right away. Everything is then totally transparent, because neither has had a chance to edit a sheet privately. It also allows the trainee to leave with something tangible. Both trainee and PI keep the full set of completed worksheets.

Before the Meeting: Tips for Trainees

Before the meeting, think carefully about your future and not just your next

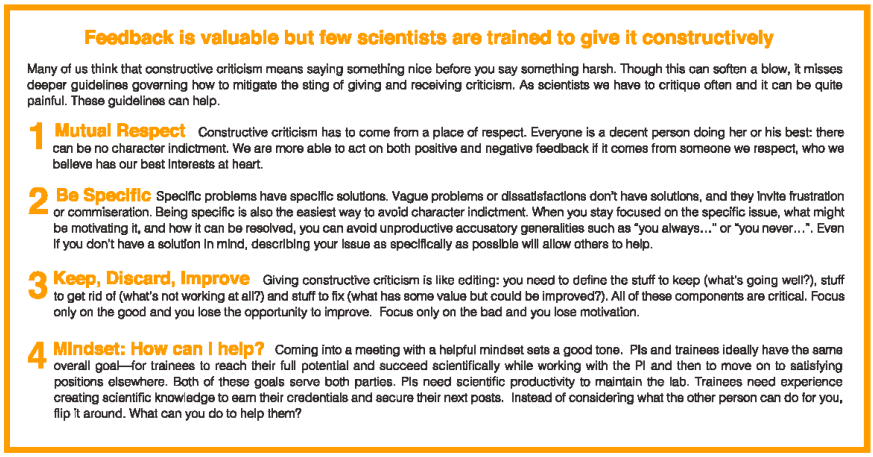


Figure 1. Guidelines for Delivering and Receiving Critical Feedback

experiment. These meetings are an opportunity to articulate your long-term goals and then make a plan to attain them, in terms of both your research and developing necessary skills. It can be helpful to seek feedback from other people in the lab about your strengths and areas that you could improve upon, and to brainstorm about specific steps to take to realize your goals. The more specific your plan before the meeting, the more productive the meeting will be. Trainees may not have yet decided on their long-term career goals; in this case the goal of the meeting can be to explore different options in a concrete way. Using online tools for exploring careers (NIH, 2014b), identifying a mentor in your field of interest, attending career seminars, or engaging a career coach (Kamens, 2015) can all be good places to start.

Before the Meeting: Tips for PIs

Before the meeting, think about your own goals for your trainee, as well as their goals for themselves. Think about the personal and professional circumstances of your trainee and how you can help their progress and motivation to reach their goals. For example, is someone in the midst of a technically challenging set of experiments? Do they need additional expertise? Can you help them find a collaborator or other resources to fill the gap? Do they need encouragement to persist or permission to move on? Or, if someone is considering a job in industry or teaching, do they have all the information they need to decide if this career direction is a good fit? Are they concerned about your opinion? For example, might they fear that if they admit they're not aiming for an academic career you will reduce your support for them? This is also an opportunity to address whether the trainee's goals are realistic. Do they have the qualifications that they need to succeed in their chosen career? Is their plan to achieve those qualifications realistic, or is it time to adjust their goals? Having a meeting every year allows both the PI and trainee to discuss both goals and the paths to attain them over time; this provides multiple opportunities to develop and execute a plan to attain a satisfying career. Remember that everyone needs different things at different times; tailoring your advice and help to their particular circumstances is useful and appreciated.

Filling Out the Goals and Planning Worksheet

Here are explanations of what should go in each category of the Goals and Planning Worksheet. Be brief. This is a bullet point list to guide your conversation and serve as a reminder later.

What to Include: Accomplishments

Most scientists are terrible perfectionists and never give themselves enough credit for the things they've achieved. Remember to be broad and generous in what you consider an accomplishment. Don't just include things that would go on a CV, such as publishing papers or giving talks at a conferences. Include progress toward goals (e.g., drafting or submitting a paper, getting a tough experiment to work), important exams (e.g., qualifying exams, GREs), applications (e.g., for jobs, fellowships, or graduate school), development of transferrable skills (e.g., learning a new technique, organizing a workshop, reviewing a paper or grant), and milestones (e.g., choosing a postdoc lab, having a thesis committee meeting, organizing an industry internship). Remember that science happens in the context of someone's whole life—you can acknowledge things outside the lab that impact work (e.g., family commitments, moving from a foreign country and getting settled, choosing a thesis lab).

What to Include: Research Goals

Focus on major milestones for getting projects accomplished, rather than nitty gritty weekly or daily goals. Goals on the

1–3 month timescale are about right, but the precise timing isn't so important, since we all know how unpredictable research can be. The point is to prioritize research goals and make an initial estimate for how long they will take so that obstacles can be clearly identified through the year. For this to work, your estimates have to be realistic, both in terms of how long things take and how many things you can accomplish during the year. If there too many goals or the estimates are unrealistic, the plan will become a recipe for disappointment. Also, consider the goals from the year before and whether they've been accomplished. If yes, fantastic! They should be listed in the accomplishments section. If not, why not? Should those goals be restated for the upcoming year, or should the project change direction or be jettisoned?

What to Include: Professional/ Personal Goals

First, this section is for first articulating your long-term career goal. Second, it is for articulating which professional skills you'd like to develop to attain that goal. Examples include getting more teaching experience, networking with people outside academia, or improving communication skills. If your career goal is still unclear, you can use this section to think about how to build on current strengths and improve on weaknesses.

What to Include: Feedback

This section is for constructive criticism about how goals are being met, both at the level of the individual and at the level of the lab. Articulate appreciation for things that are going well and specific issues that could be improved. For trainees, think about how the lab and your interactions with your PI are working for you. For example, are you meeting too little or too much with your PI? Are you worried about the trajectory of your project or someone else's in the lab or in the field? Do you have the balance of projects and free time that you're looking for? If there are general issues in the lab, even if they don't pertain to you in particular, it can be helpful to let the PI know. Remember, you're not tattling, you're pointing out conflicts so that they can be resolved. For PIs, articulate the trainee's strengths and areas where they could improve. Ideally these should be related to their long-term goals; it is most useful for them to build on relevant strengths and improve on relevant skills. Make sure you say thank you to your trainees for specific things (e.g., getting slides ready for your big talk at the last minute, bringing a new technique to the lab, mentoring a rotation student) (Riordan, 2013). It would be a disaster if you only put negative things in this section.

Tips on Conducting the Meeting

It Is Critical for the Trainee to Lead the Conversation

This simply means that the trainee shares their content first. The PI can pipe in, but must be careful not to interrupt. The conversation proceeds section by section. When the trainee is finished, the PI may add any additional comments, or summarize the discussion. If the PI shares their content first, the whole meeting will have a tone of the trainee being assessed. Allowing the trainee to lead the meeting gives them a sense of ownership over their goals.

Work from Printed Copies of the Worksheet during the Meeting

Having an iPad or laptop screen between you will keep you from making eye contact, and for this meeting to succeed, having a good interaction is more important than taking tidy notes! Electronic copies are also less transparent, because you can alter them during the meeting. Printed copies can and should be annotated by both people during the meeting and will show everything that the PI and trainee recorded individually.

Starting with Accomplishments Helps to Set a Positive Tone

Pausing to reflect on accomplishments from the previous year together helps people feel appreciated and competent and motivates everyone to accomplish in the next year.

The PI's Responsibility Is to Make Sure that Everything Has Actionable Outcomes

Trainees may bring up issues that they need help solving. PIs can offer lots of different types of help. For example, they can connect trainees to people with needed expertise, either inside or outside of the lab. They can offer funding for conferences, workshops, or courses. They can offer more or less oversight, or an extra set of hands for a tough experiment. Finally, PIs have the authority and responsibility to resolve interpersonal conflicts.

Feedback Comes Last in the Meeting, and the PI Receives Feedback from the Trainee First

The PI is in the power position; modeling how to productively seek and receive feedback helps the trainee to give and receive feedback themselves. Remember that it's hard to tell people negative things! It's a real emotional risk—it would be much easier to just say everything is fine. Try to see critical feedback as an opportunity to improve. You don't have to incorporate all suggestions immediately. Have some phrases ready in case you're faced with a critique you don't know how to respond to (e.g., "Thanks for letting me know. I'll definitely think about that more.").

Advantages for PIs

Motivated Individuals Make a Productive Lab

When you engage in this process with your trainees, they will be motivated and empowered to pursue their work. Moreover, the benefits extend to the lab as a whole. You've taken time to connect people with relevant expertise and resolve potential conflicts. Think of it like yearly preventative maintenance.

A Way to Check In on Progress through the Year

Discrepancies between the calendar and actual progress should trigger one-on-one meetings. This happens all the time. Things get delayed for all kinds of reasons. The calendar helps to frame these conversations and make them specific. For example, "We thought that by April you'd be on to part X of your project, but that hasn't happened yet. Let's chat about why—maybe we underestimated, or another priority has come up. But if there's an obstacle, let's figure out what it is."

Paying it Forward

Every scientist owes something to a fantastic mentor. Spending time with your trainees, focused on how you can help them reach their goals, can be immensely rewarding and is a way to repay your own mentors (Murray, 2011).

CellPress

Advantages for Trainees

Strengths and Accomplishments Are Formally Acknowledged

This happens fairly rarely in the critical culture of science, particularly for trainees, and can be a huge confidence booster.

Concrete Goals Are Motivating

Reaffirming research priorities can relieve pressure and provide clear ways forward when you're juggling many projects.

Personal Goals Are Formally Discussed

This is important in the current climate, in which academic positions are scarce and trainees are often looking for opportunities in other sectors. An honest conversation with your PI can help dispel anxiety and formulate a tractable training plan to help you secure your next post.

Personal/Interpersonal Issues Can Be Discussed and Ironed Out

This meeting provides the opportunity to give feedback to your PI, as well as to discuss any tensions that may make life in the lab less than optimal. Remember, your PI may not be aware of issues with the lab atmosphere, but finding solutions to problems is possible and can make a huge difference to everyone's happiness and productivity.

Research Expectations and Priorities Are Clarified

This enables both parties to be on the same page in terms of priorities and timelines for the coming year.

Conclusion

We hope that this new requirement for IDPs inspires real conversations between trainees and PIs. It undoubtedly will look different in every lab. We hope these guidelines are a useful starting point to develop a concrete, rewarding process for you.

SUPPLEMENTAL INFORMATION

Supplemental Information includes blank and completed examples of the Goals and Planning Worksheet and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2015.04.025.

ACKNOWLEDGMENTS

The authors gratefully acknowledge all past DePace lab members for participating in yearly planning meetings and helping to refine the process; the Department of Systems Biology for encouraging strong mentorship; support from the Research Scholar Initiative, the PhD Program in Biological and Biomedical Sciences, the Molecules, Cells and Organisms Graduate Program and the Systems Biology PhD Program; the Harvard Office for Postdoctoral Fellows for embracing this process; Galit Lahav, Becky Ward, Kristin Krukenberg, Jessica Polka, Jim Gould, and Edward Pym for their thoughtful comments on the manuscript; Sašo Kočevar for his insightful leadership training for scientists; Uri Alon for his early encouragement to positively change lab culture; and all of our own inspirational mentors.

REFERENCES

Bourne, H. (2013). eLife *2*, e01139, http://dx.doi.org/10.7554/eLife.01139.

Kamens, J. (2015). http://blog.addgene.org/career-coaching-for-scientists.

Murray, A. (2011). https://www.youtube.com/watch?v=WiK0nm19_D0.

National Academy of Science (2014). http://www.nap.edu/catalog/18982/the-postdoctoral-experience-revisited.

National Institutes of Health (2012). http://report.nih.gov/investigators_and_trainees/acd_bwf/.

National Institutes of Health (2014a). http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-113.html.

National Institutes of Health (2014b). http://www.nigms.nih.gov/training/strategicplanimplementationblueprint/pages/IndividualDevelopmentPlans.aspx.

Polka, J. (2014a). http://ascb.org/where-will-a-biology-phd-take-you/.

Polka, J. (2014b). Nature *511*, 256.

Riordan, C. (2013). http://hbr.org/2013/04/foster-a-culture-of-gratitude/.

Sauermann, H., and Roach, M. (2012). PLoS ONE *7*, e36307, http://dx.doi.org/10.1371/journal.pone.0036307.

GOALS and PLANNING

Name & date

Accomplishments (from previous timeframe)

Research Goals (for upcoming timeframe)

Professional & Personal Goals (for upcoming timeframe)

Feedback

GOALS and PLANNING Name & date

January

February

March

April

May

June

July

August

September

October

November

December

GOALS and PLANNING
FROM JANE

June 2015
Name & date

Accomplishments (from previous timeframe)

Published paper
Drafted main paper including new experiments
Genome editing experiment in progress
Wrote NSF research plan; funded!
Followup transgenics in progress
Tried CRISPRi - need to troubleshoot
CSHL poster
Fly meeting abstract submitted
Supervised George - expt in evolution paper
Supervised Ringo - sufficiency experiment
Supervised Paul - cofactor screen + followup
Talks at recruitment and retreat
Organized group meeting and journal club
Scheduled DAC #3

From Angela - expts on defining regulators

Research Goals (for upcoming timeframe)

Continue rescue experiment w/ genome editing
Measure followup constructs
cis/trans experiments for bifunctionality project
} high priority

Continue to support CRISPRi
Followup on cofactor screen
} low priority

From Angela -
- Think about kinetic synergy angle for second project
- Incorporate some followup in cell culture / bioinformatics

Professional & Personal Goals (for upcoming timeframe)

Apply for communication award
Submit main paper (think about where)
DAC #3
Present at a national meeting
Outline cofactor screen project
Start thinking about postdoc labs
Department talk?
Graduate late 2016

From Angela -
- If undergraduate-focused teaching + research is goal, think about system cost in postdoc lab
- Contact info for colleagues who have focused on undergraduate education.

Feedback TO ANGELA

New system has really helped with communication with you and others in lab
As always, you provide excellent support both scientifically and personally and help us develop as well-rounded scientists! :)
Writing the NSF grant was a great experience! 3-person team writing is a good template for the future.
We've gotten better at setting appropriate expectations for rotation students.
Lack of clear timetables on paper drafts have been a source of frustration for a couple lab members
I've mentored 7 people in 3 years and often feel like the only person with rotation projects in place.

GOALS and PLANNING

Jane Smith 2015

FROM ANGELA

Accomplishments (from previous timeframe)

- ✓ Helped to plan and write NSF grant.
- ✓ Published paper!
- Mentored ✓ George, ✓ Ringo and ✓ Paul
- ✓ Organized and executed Genetics bootcamp course
- ✓ Executed all CRISPRi cloning, got back transgenic flies
- ✓ Finished data collection for all synthetic enhancers, have found the narrative for the paper
- ✓ Half-way done with identifying all the regulators for 2nd enhancer project

Poster @ CSHL
Poster @ Fly Meeting

Grad program recruitment & retreat
DAC Meetings
Took over organizing journal club.

Research Goals (for upcoming timeframe)

- ✓ For synthetic enhancer paper, need final piece of data on transgenic rescue
- ✓ For second enhancer project, finish identifying all regulators, consider their functional role
- ✓ Transcription factor bifunctionality project, complete cis and trans experiments outlined in the grant
- [Revisit the cell culture and biochemistry experiments suggested by your committee]

→ Not super interested in this, haven't taken any practical steps yet.

longer term, think about coordinating this project with John.

Professional & Personal Goals (for upcoming timeframe)

- ✓ Work from paper drafts for two manuscripts above
- ✓ Begin thinking of next steps at end of 2015
- ✓ Presentation at conference?

iBio presentation
DAC #3
Departmental talk
Graduate 2016 - interviews for postdocs late 2015/early 2016.

Aiming for research & teaching at undergrad institution

submit #1 (eLife, PLoS Bio, MSB, or PLoS Genetics)

Feedback

STRENGTHS - Science is coming along at a good pace - you may wrap up 3 separate papers before graduating. Such excellent communication skills (with me and everyone else in lab). Really appreciate how you keep things running and take the initiative to get things done (while I was on maternity leave, and since getting back). Writing the grant and the paper was fun for me, largely because it's fun to work together. Spectacular mentoring and teaching skills. Will be looking to you and John to help with the transition to a new set of people after the big turnover this year because of graduation and new jobs for multiple people.

AREAS FOR MORE FOCUS - You may need to triage some experimental directions if your goal is to graduate in 2016. It would be helpful to diversify your experiments (you're doing a lot of molecular biology and imaging). It's also time to start preparing for next steps (what can I do to help?).

FEEDBACK FOR ANGELA (from Jane)

- Writing the grant was a good experience
  - helped with writing skills
  - good opportunity to think about future directions
- I've gotten better at setting expectations for rotation students.
- There's a lack of clear time tables when writing papers. It's frustrating when the PI is the bottleneck. Maybe let people know what order they are in line if lots of people are in the queue for PI attention?
- Jane's done too much rotation student mentoring recently — she really needs a break.

GOALS and PLANNING

Jane Smith 2015
Name & date

January — GRANT CONSTRUCT CLONING

February — REDO COMPUTATIONAL ANALYSIS

March — FLY MEETING — poster, look @ postdoc labs
DAC

April — SUBMIT PAPER #1?

May

June — SEND IN REVISION OF R01

July — DATA COLLECTION ON TF CONSTRUCTS

August

September

DAC?

October — SUBMIT PAPER REVISION
DECIDE WHERE TO APPLY FOR POSTDOCS
OUTLINE THESIS.

November

December