Patched Together: *cis*-Regulatory Logic of the Hedgehog Response

by

David Samuel Lorberbaum

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2016

Doctoral Committee:

Associate Professor Scott E. Barolo, Chair
Assistant Professor Benjamin Allen
Professor Kenneth M. Cadigan
Associate Professor Anthony Antonellis

Dedication

To my dad and brother for their unwavering support, encouragement, and love on this incredible journey.

To my mom for reminding me to be happy, even in the face of despair.

To the rest of my family and all of my friends for keeping a smile on my face. I could not have done it without you.

And to Dazzle for her unconditional love.

Table of Contents

List of Figures

List of Tables

Abstract

Understanding the processes that control how we develop from a fertilized embryo to a functional adult is paramount for treating the diseases that result when these processes are disrupted at any stage of life. My dissertation investigates the *cis*-regulatory logic underlying how cell signaling pathways utilize the genome to create and maintain the wide variety of cell types and tissues needed for proper development and survival.

Surprisingly few cell signaling pathways are used throughout embryonic development; I have chosen to focus on Hedgehog (Hh) signaling, a pathway used in such diverse cellular contexts as digit specification, brain development, lung function, and reproductive maintenance (Ingham and McMahon, 2001). Disruption of this pathway results in developmental defects (Lipinski et al., 2010; Schachter and Krauss, 2008) and cancer (Barakat et al., 2010; Hahn et al., 1996; Johnson et al., 1996; McMahon et al., 2003). It is essential to understand the mechanisms by which Hh signaling functions to treat these diseases more effectively. One relatively unexplored mechanism of Hh function is how its signal is transduced at the level of DNA, specifically through the regulation of gene expression. In this thesis, I explore the mechanisms that mediate tissue-specific, Hh-dependent gene regulation, and uncover an ancient *cis*-regulatory logic shared between flies and mice that has significant implications for the maintenance and evolution of cellular communication. I experimentally demonstrate that multiple enhancer elements, which control tissue-

specific gene expression, rely on sub-optimal DNA sequences for binding of GLI proteins, the transcriptional effectors of Hh signaling. These sequences are essential to control gene expression in response to Hh and can influence the function of the pathway in a variety of cellular contexts. I also characterize several new transcriptional regulators of Hh signaling and introduce new tools to the field that allow for in depth analysis of the regulatory landscape of Hh target genes at any stage of development.

My work presented here addresses a significant gap in our knowledge of how the Hh signaling pathway functions at the *cis*-regulatory level and describes a framework by which new advances can be made on this topic in the future.

Chapter 1

Introduction

1.1 Non-coding DNA and a brief history of enhancers

DNA is the genetic material that holds a code which determine the characteristics defining who we are; from the color of our eyes to the number of toes on our feet (Avery et al., 1944). Unfortunately, it is also the culprit behind many inherited genetic disorders, developmental defects, and environmentally acquired diseases, like cancer. Similar to how a set of blueprints can be read by architects, biologists have the ability to read DNA, although the code has not yet been completely cracked. We know that at the heart of these blueprints there are four molecules: Adenine (A), Thymine (T), Guanine (G), and Cytosine (C). These molecules are called nucleotides and are arranged in a particular order (Watson and Crick, 1953). Under the correct conditions, certain stretches of DNA, called genes, are processed into RNA via transcription, which then can then be translated into proteins. It is critical that transcription occurs at the correct time, at the appropriate levels, and in the right contexts throughout development. The field of gene regulation is dedicated to understanding how genes are expressed in this way, but we are far from a complete understanding of how the genome functions (Dahm, 2005). My dissertation focuses on how a single genome regulates gene

expression so that that we develop normally and maintain healthy cellular processes well into adulthood.

The ability to quickly and efficiently sequence entire genomes has provided a wealth of data that has been instrumental for understanding how the genome works (Shendure and Ji, 2008). Prior to sequencing the genome, estimates based on the (false) premise that the average human gene was roughly 30,000bp long and that genes covered the entire genome led scientists to believe there could be more than one hundred thousand genes encoded in our DNA (Pertea and Salzberg, 2010). After completion, the human genome project revealed that there are only between 22,000 and 23,000 genes in the human genome (Lander et al., 2001). Strikingly, this means that only ~2% of our total DNA consists of coding genes, leaving the other ~98% as non-coding DNA (Elgar and Vavouri, 2008). That 98% of human DNA not making genes serves many purposes and one of its most important functions is directing the proper temporal- and tissue-specific expression of genes, a process known as gene regulation (Levine, 2010; Levine et al., 2014). This is largely achieved through regions of DNA called *cis*-regulatory modules (CRMs). These modules are made up of patterns of nucleotides that can be bound by proteins called transcription factors (TFs) at the right time and place in development. Once bound, TFs recruit the basal transcriptional machinery to the coding region of a gene to make RNA, via transcription (Villicaña et al., 2014). One type of well-studied CRM is an enhancer, named because of its ability to "enhance" expression of its target gene. Pioneering work in the early 1980's characterized a 72 bp fragment of viral DNA that increased the transcriptional activation of the rabbit *β-Globin* gene more than 200-fold (Banerji et al., 1981). This was also a

significant discovery because of the physical location of this regulatory DNA relative to the *β-Globin* gene: this enhancer was found distal to the coding region, demonstrating that transcriptionally relevant sequences can be located at a distance from the genes they regulate. This landmark discovery demonstrated that non-coding regions of the genome play a functional role in biology. Similar work was published around this time demonstrating other genes were regulated by enhancers in a similar manner as *β-Globin* (Benoist and Chambon, 1981; Fromm and Berg, 1983). All of these findings helped pave the way for enhancer biologists to better understand how genes are regulated at the level of DNA sequence.

Today, we have many examples of similar kinds of complex gene regulation by enhancers in different contexts. While Banerji and colleagues showed that enhancers can work at a distance from the gene they regulate, a more striking example of this was later shown for the *Sonic Hedgehog* (*Shh*) gene in mice and humans. This gene is absolutely essential for the normal development of many different developmental tissues, like the spinal cord, limbs, and several organs required for survival (Bellusci et al., 1997; Chiang et al., 1996; Ingham and McMahon, 2001; Niswander et al., 1994). In the early 2000's, a great amount of effort was dedicated to identifying a distally located enhancer of *Shh*, which resides more than 1Mb away from its coding region (Lettice et al., 2003). This enhancer specifically regulates *Shh* expression even though it is found in the intron of another gene completely unrelated to *Shh.* It is also tissue-specific, contributing to the regulation of *Shh* in just the limb for digit specification, which introduces yet another layer of complexity to the function of enhancers; not only do genes require enhancers that can be located far away from the coding region of the

gene, enhancers are often tissue specific. Importantly, deletion of this enhancer, while leaving the *Shh* coding region intact, abrogates normal limb development, demonstrating the functional contribution of non-coding sequence to developmental processes (Sagai et al., 2005). Follow-up studies on this distal *Shh* enhancer also uncovered similar mutations in polydactylous felines known as Hemingway's cats (Lettice et al., 2008). These cats have been celebrated for having extra digits, and the driving force behind this malformation were single nucleotide substitutions in this conserved *Shh* enhancer, demonstrating both strong conservation between species and functional importance of regulatory elements in key developmental processes (Lettice et al., 2008).

1.2 Enhancer structure and function

Enhancers are able to regulate gene expression in specific tissues using information encoded within the sequence of DNA itself, but also rely on the available transcription factors in different developmental contexts. Enhancers can range in size and there is no limit to how large an enhancer might be (Hnisz et al., 2013; Parker et al., 2013). The content of enhancers is an assortment of transcription factor binding sites arranged in a particular order to facilitate the binding of their cognate TFs (Johnson et al., 2008). An enhancer for the *Drosophila Pax2* (*dPax2*) gene has been essential for learning more about how enhancers work at the level of DNA sequence. This relatively short 362bp enhancer termed *sparkling, or spa*, requires input from Notch, EGFR, Runx, Mapk, and ETS factors, making it an ideal example of a well-characterized signal-regulated enhancer (Flores et al., 2000; Fu and Noll, 1997; Fu et al., 1998). Interestingly, a synthetic version of this enhancer containing just the known binding sites

for the required inputs of *spa* failed to recapitulate normal expression, thus demonstrating that *spa* required even more regulatory information than previously thought to activate *dPax2.* This again highlighted the need for multiple transcriptional inputs into a single regulatory element (Swanson et al., 2010a; Swanson et al., 2011)**.** The other part of enhancer tissue specificity is the availability of TFs in each developmental context. Since all of our cells, regardless of cell type or function, have the same DNA, they also have the same enhancers for every gene that will ever be activated or repressed throughout development. There will often be high quality TF binding sites in regions of the genome that will remain unbound because there is no protein to recognize that sequence. For example, even if there is a perfect binding site for a leg specific TF in the eye-specific *spa* enhancer, that TF will not be present in the developing eye and that binding site will remain unoccupied (or at least unoccupied by that particular TF). These constraints further increase the specificity by which enhancers can regulate gene expression.

1.3 Enhancer identification

One of the big, open questions in the field of transcriptional biology is how to identify functional binding sites within enhancers. There are many different tools available to accomplish this and each has their own advantages and disadvantages. Since there is ample sequence data freely available across many different species, we can use evolutionary conservation to identify regions of DNA that are shared between different species, like mouse and human, for example. This comparative genomic bioinformatic technique has been used with varying degrees of success (Berman et al., 2004; Gurdziel et al., 2015; Pennacchio et al., 2006)**;** see Chapter 4 for more details).

By aligning genomic sequences from different organisms, we can find which specific sequences of DNA are conserved, suggesting that they are functionally relevant. We can then use any number of transcription factor binding site prediction algorithms to predict if TFs might be important for these enhancers (just a handful of examples are CIS-BP, (http://cisbp.ccbr.utoronto.ca/), (Weirauch and Hughes, 2010)**;** Weeder-Web (Pavesi et al., 2004)**;** and databases like Fly Factor Survey - http://mccb.umassmed.edu/ffs/ - curated by Michael Brodsky and Scot Wolfe). While these methods have the advantage of being able to rely on sequence data alone, using only this prediction tool is limited. The first limitation is that not all transcription factors are expressed in all tissues. This gets to the point of context-specific transcription factors, like in the case of *spa* discussed above. Looking at just the sequence information does not give us clues as to if the TF is actually expressed in the relevant tissue where that enhancer is required for development. The second limitation is that we are still learning what a transcription factor binding site (TFBS) actually is and how it works. Most of the data available for TFBS prediction comes from in vitro detection methods like SELEX, EMSAs, and Bacterial 1 Hybrid analysis (Christensen et al., 2012; Hallikas et al., 2006; Tuerk and Gold, 1990; Zhu et al., 2011). These methods generally provide a prediction of what sequence a TF might recognize, but lack information about the required co-factors or context restricted availability of the TFs. Furthermore, these binding sites that are strongly predicted for certain TFs in vitro have been shown to not always be favored during development (Crocker et al., 2015) as discussed in Chapter 2. For these reasons, additional analyses are required to more accurately predict and characterize enhancers.

There are also methods by which enhancers can be identified on a genome-wide basis. DNAse hypersensitivity and FAIRE-seq analyses, for example, both identify regions of open chromatin that predict where certain TFs can bind to DNA. DNAse hypersensitivity, or footprinting, takes advantage of protein binding to DNA (Galas and Schmitz, 1978). In this assay, protein is bound to DNA, the DNA is then sheared using DNAse enzyme that will digest free DNA, which is not bound by protein, leaving all regions bound by protein in tact. Once sequencing technology became available, these regions bound by protein could then be dissociated from the bound protein and sequenced using high throughput analysis to identify a genome's worth of binding data. FAIRE-seq, or Formaldehyde-Assisted Isolation of Regulatory Elements uses the same principles as DNAse hypersensitivity, but with the added benefit of being able to look more closely at cell-type specific data (Giresi et al., 2007). These techniques can accurately predict enhancers, but are not able to assign function to enhancers and cannot delineate specific TFs that are bound to these regions of DNA.

To more specifically identify the DNA regions bound by specific TFs, ChIP-seq is a very popular and efficient tool (Johnson et al., 2007). This technique relies on similar principles of detecting protein-DNA complexes, but takes it a step further by using a protein-specific antibody. By using this antibody, the protein-DNA complex can be isolated from specific cell-types and DNA fragments sequenced, revealing the in vivo TF occupancy of specific TFs. While this technique allows for more controlled identification of enhancers, it also relies on the specificity of the antibody, or how well it can recognize it's epitope on the protein trying to be isolated (Marx, 2013). Some TFs, like the Hox genes as one example, are notoriously difficult to target because of protein redundancy

(Mann et al., 2009). Furthermore, ChIP-seq, along with DNAse and FAIRE-seq only provide a snapshot of what is going on in the cell at one particular timepoint and does not provide direct evidence for the function of these enhancers for the regulation of gene expression.

To assess enhancer function on a genome-wide scale, a technique has recently been described called STARR-seq (Self-Transcribing Active Regulatory Region - sequencing) to detect enhancers by allowing them to actively transcribe themselves using a minimal heterologous promoter (Arnold et al., 2013). This technique takes advantage of enhancer function (i.e. their ability to up-regulate transcription) to demonstrate enhancer-promoter specificity and to drive transcription (Lorberbaum and Barolo, 2015; Zabidi et al., 2014). While this technique faithfully identifies many tissue-specific enhancers and highlights that they are functional, it is limited by the requirement to conduct experiments in cell culture, rather than in their native developmental context.

Another method used to study regulatory elements is an enhancer-reporter transgene that can be integrated into the chromosome of a model organism. This technology relies on easily detectable genes that are not present in the host genome being studied and are visualized either by a chemical reaction (*lacZ*) or by emitting detectable light (GFP, or any variety of fluorescent protein). For a brief review and relevant examples, see (Barolo et al., 2000; Shaner et al., 2005). To test the function of an enhancer in vivo, it can be cloned upstream of a promoter, which will drive the expression of a reporter gene. This construct can then be integrated into the genome using one of many transfection techniques that vary by organism. Depending on when and where that enhancer is normally active, it will drive expression of the reporter gene

at the right time and place during development. Furthermore, because the enhancer-reporter construct is integrated into the host chromosome, it will also experience everything DNA normally experiences in that cell type -- including all the relevant TFs, co-factors, or any epigenetic regulation that might be important. Downstream processing using immunofluorescence, for example, can also provide important clues about enhancer activity. One of the major drawbacks of this technique, however, is that in order to be tested, an enhancer must already be identified. It is also a relatively low throughput method of characterizing enhancer function, but does have the advantage of being an in vivo readout of activity.

In addition to those already mentioned above, there are still other methods as well that rely on histone marks such as methylation and acetylation to provide clues on a genomic scale for enhancer mapping (see ENCODE project, review here: Stamatoyannopoulos, 2012). Taken together, however, there is no single, simple method for identifying enhancers. All of these resources must be used in combination to identify and validate enhancers that are important for the regulation of gene expression.

1.4 Hedgehog signaling in development

In order to better understand the role of enhancers in gene regulation and developmental biology, I have opted to study these CRMs alongside cell signaling pathways. Amazingly, there are only a handful of cell signaling pathways used repeatedly throughout development (Barolo, 2002). Rather than examine all of these pathways, I have dedicated my efforts to understanding transcriptional regulation controlled by the Hedgehog signaling pathway (Hh). First identified in a forward genetic screen in *Drosophila melanogaster* (Nüsslein-Volhard and Wieschaus, 1980), Hh has

been extensively studied and shown to regulate amny cellular responses to signaling and shape morphological development (Ingham, 2016). Regardless of the tissue or context, Hh signaling works similarly: in the absence of Hh ligand, the 12-pass transmembrane receptor Patched (Ptc in flies; PTCH1 in mammals) inhibits the function of Smoothened (Smo) (Chen and Struhl, 1996; Ingham et al., 1991). This inhibition leads to the proteolytic cleavage of GLI, known as Cubitus Interruptus or Ci in flies, which removes the activation domain leaving GLI to act as a repressor (GLI-Repressor, GLI-R;) so that when it enters the nucleus GLI-R will be able to repress the transcription of target genes (Aza-Blanc et al., 1997; Méthot and Basler, 2001; Orenic et al., 1990). In the presence of the Hh ligand the inhibition of Smo by Ptc is removed, protecting the activation domain of GLI, allowing GLI-Activator (GLI-A) to enter the nucleus, recognize its binding site(s) in the appropriate enhancer, and activate transcription of target genes (Figure 1.1; (Briscoe and Thérond, 2013; Hui and Angers, 2011; Infante et al., 2015)). In all cases known from flies to mice, GLI proteins are the main transcriptional effectors of the pathway (Buttitta et al., 2003; Méthot and Basler, 2001; Motoyama et al., 2003).

Two well-studied contexts that require Hh signaling in *Drosophila* development are the wing imaginal disc and the embryonic ectoderm. Imaginal discs are specialized structures found in the larval stages of developing *Drosophila* that are destined to develop into a specific adult appendage (Basler and Struhl, 1994; Mohler, 1988; Tabata and Kornberg, 1994). In order for this development to occur, imaginal discs must receive signaling input from many signaling pathways, including Hh. Normally, cells in the posterior produce the Hh ligand, which is then secreted as a morphogen across the anterior-posterior (A-P) axis to regulate target genes (Figure 1.2, left; (Aza-Blanc et al.,

1997; Motzny and Holmgren, 1995)). Importantly, the transcription factor Engrailed (En) permits Hh expression and represses Ci in those cells, so that the pathway is not constitutively active, and it eventually helps to transduce the signal across the axis (Blair and Ralston, 1997; Eaton and Kornberg, 1990; Rodriguez and Basler, 1997). Cells sitting in the anterior region near that A-P axis will experience high levels of Hh signaling and activate the pathway, turning on Hh target genes (Figure 1.2, left hand panel; (Hooper and Scott, 1989; Nakano et al., 1989)). The embryonic ectoderm is regulated in a similar manner as the wing imaginal disc, but instead of having just one posterior and anterior region, there are segmentally repeated stripes that produce Hh where Ci is similarly repressed by En (Eaton and Kornberg, 1990; Forbes et al., 1993; Perrimon, 1994). Hh ligands are secreted both anteriorly and posteriorly and cells receiving high levels of signal in each segment will activate the pathway, turning on the appropriate target genes ((Briscoe and Thérond, 2013; Fietz et al., 1995; Hidalgo and Ingham, 1990); depicted in the right hand panel of Figure 1.2). Importantly, in developmental contexts regulated by Hh signaling, including the wing and embryonic ectoderm, one of those key target genes that must be activated is *patched* (*ptc*) (Alexandre et al., 1996; Forbes et al., 1993; Goodrich et al., 1996; Hooper and Scott, 1989). Since *ptc* codes for the transmembrane receptor the of the pathway, it introduces a negative feedback mechanism that is essential to regulate the pathway in all Hh-responsive developmental contexts (Ingham et al., 1991; Tabata and Kornberg, 1994; Vokes et al., 2007).

We know that in order for Hh target genes to respond to their signal, GLI must recognize a particular DNA sequence, or binding site, in the target gene's enhancer(s).

Much work has been done on the sequences GLI prefers to bind dating back to the early 1990s, when it was demonstrated that human GLI prefers the sequence GACCACCCA as its consensus binding site (Kinzler and Vogelstein, 1990). While these data were generated in human cell lines with human proteins, it was later shown that other mammalian GLIs and Ci recognize nearly identical sequences in both mice and flies (Hallikas et al., 2006), perhaps not too surprisingly, since this pathway is highly evolutionarily conserved (Ingham et al., 2011). Three of these consensus GLI binding sites (GBS) were identified in the *Drosophila* locus of the Hh target gene *ptc*. When a promoter proximal *ptc* enhancer was cloned into a reporter construct, it responded in a Hh-like pattern (Alexandre et al., 1996). When the region of this enhancer containing those three consensus GBS was deleted in a parallel experiment expression was greatly reduced, demonstrating their requirement for activation in the wing imaginal disc (an updated version of this experiment is described in Chapter 2, specifically mutagenizing those sites and using a GFP reporter instead of *lacZ*). This seemingly simple experiment characterized the first Hh responsive *ptc* enhancer, and also provided strong evidence that consensus GLI binding sites are critical for the regulation of Hh target genes.

1.5 The importance of GLI binding site affinity

While GLI binding sites in the promoter-proximal *ptc* enhancer were sufficient to respond to Hh signaling in the wing imaginal disc, they were not sufficient to respond in all Hh-regulated developmental contexts (see Chapter 2). In fact, most known Hh target genes in flies do not require high affinity GLI motifs to respond to Hh signaling (Ramos and Barolo, 2013). Rather, they require binding sites that slightly vary from the

consensus sequence recognized by GLI, referred to as low affinity binding sites, since they are recognized by GLI with a lower predicted affinity (Hallikas et al., 2006). Two examples of well-characterized enhancers that require low affinity GBS to respond to Hh signaling belong to *wingless* (*wg*) and *stripe* (*sr*). First, *wg*, which encodes the ligand for the WNT signaling pathway, is directly activated by Hh signaling in the *Drosophila* embryo via an enhancer that contains four low affinity GLI binding sites (Ohlen and Hooper, 1997; Ohlen et al., 1997). While no transgenic reporter animals carrying this enhancer were examined in these studies, cell culture experiments in embryo-derived cells were used with luciferase reporters to demonstrate the importance of these low affinity GBS. Additionally, these sites were confirmed with EMSAs to show that GLI could recognize those specific binding sites. This enhancer was later examined as a transgenic reporter construct in flies and again required low affinity GBS to properly respond to Hh signaling (Ramos and Barolo, 2013). Second, *stripe* (*sr*), a regulator of muscle development in late *Drosophila* embryogenesis, also requires low affinity GBS in its enhancer to respond to Hh signaling. A GFP-reporter transgene carrying a wildtype enhancer responded to Hh signaling, but when its low affinity GBS were mutagenized, the enhancer could no longer respond (Piepenburg et al., 2000). The Barolo lab further characterized both the *wg* and *sr* enhancers by improving the affinity of their low affinity GBS so that they match the consensus predicted GLI motif (GACCACCCA). Even with an affinity upgrade, we saw less activation in these contexts further demonstrating the importance of low affinity GBS for Hh response (Ramos and Barolo, 2013).

There is also strong evidence supporting the requirement of low affinity GBS for Hh target gene regulation in mammals. For instance, work published by Oosterveen and

colleagues identifies a series of Hh target gene enhancers active in the neural tube that are regulated by GLI-repressor acting through low affinity GBS. Interestingly, GLI-activator was not affected by changing affinities of GBS (Oosterveen et al., 2012). While this group found an interesting link between low affinity GBS and GLI-R activity, another group demonstrated that both GLI-A and -R depend upon low affinity binding sites. Hh-responsive enhancers for *FoxA2* and *Nkx2.2*, direct Hh transcriptional targets in the vertebrate neural tube, are both influenced by GBS affinity for maintaining normal patterning of the neural tube and preventing unwanted expansion of the floor plate in this Hh-regulated context (Peterson et al., 2012). Taken together, these studies strongly suggest that low affinity GBS are critical for generating a normal Hh response in a wide array of developmental contexts, tissues and animals.

In this dissertation, I identify an array of new enhancers, describe how they are able to respond to signaling in different contexts, and provide several key pieces of evidence further supporting the importance of GBS affinity in Hh target gene regulation. I address the problem of enhancer identification from a bioinformatics and genomics standpoint and offer a solution that combines both methods to better identify Hh target gene enhancers. Finally, I introduce new tools to better understand how genes are regulated in their native genomic contexts using CRISPR/Cas9 gene editing technology. The chapters all work towards the same goal: to better understand how a single genome can be used to pattern diverse tissues to ultimately create a normal, healthy adult.

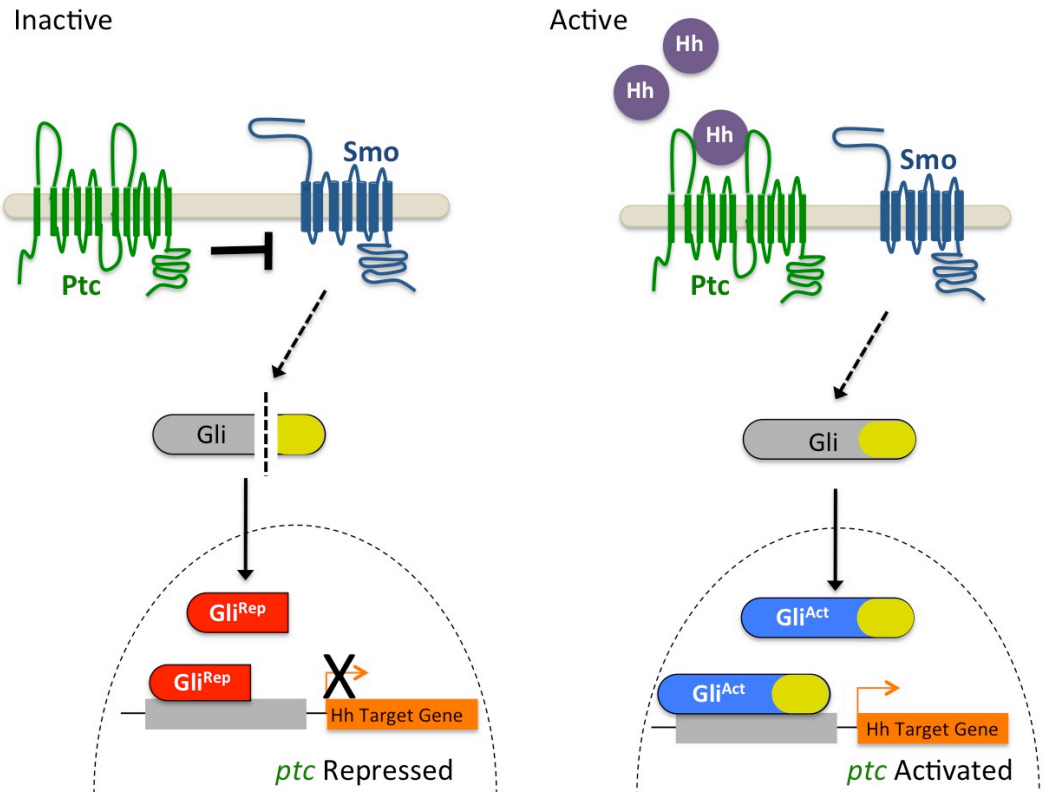


Figure 1.1 The Hedgehog (Hh) signaling pathway

Starting in the left hand panel, in the absence of Hh ligand, Ptc inhibits Smo. This leads the proteolytic cleavage of Gli (or Ci in flies). This cleavage event removed the activation domain of Gli (yellow), generating a Gli-Repressor that enters the nucleus, and upon binding to its *cis*-regulatory element, will repress target gene transcription, of which *ptc* is a target gene. On the right hand panel, active Hh signaling is depicted. When Hh ligand is present, it can bind to Ptc, relieving the inhibition of Smo. This protects the activation domain of Gli, rendering it a transcriptional activator, so when it enters the nucleus, it binds to DNA to activate Hh target genes. Once again, *ptc* is a direct target of the pathway.

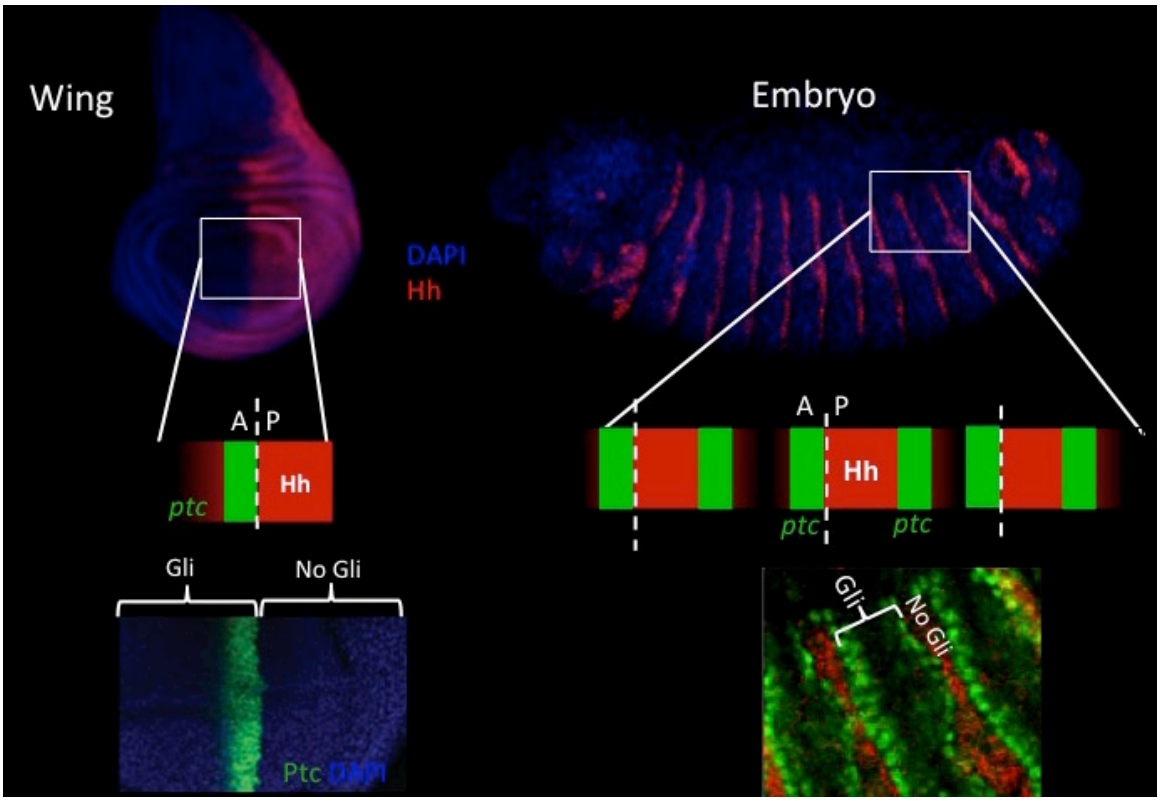


Figure 1.2 Hh signaling in the wing imaginal disc and embryonic ectoderm

In the wing, left, Hh is expressed in posterior cells that do not express Gli. The Hh ligand is secreted into the anterior cells where it acts as a morphogen to activate the pathway (Figure 1.1) turning on target genes, of which *ptc* is again used as an example. In the embryonic ectoderm, right, Hh is expressed in the posterior of each segment in the embryo. It is secreted both anteriorly and posteriorly to cells that express Gli and are able to activate the pathway. These cells receiving high levels of Hh signal will then turn on Hh target genes, like *ptc*.

Chapter 2

Fine Tuning Hedgehog Signaling: A *cis*-regulatory Strategy Conserved from Flies to Mice

2.1 Abstract

The Hedgehog signaling pathway is part of the ancient developmental-evolutionary animal toolkit. Frequently co-opted to pattern new structures, the pathway is conserved among eumetazoans yet flexible and pleiotropic in its effects. The Hedgehog receptor, Patched, is transcriptionally activated by Hedgehog, providing essential negative feedback in all tissues. Our locus-wide dissections of the *cis*-regulatory landscapes of fly *patched* and mouse *Ptch1* reveal abundant, diverse enhancers with stage- and tissue-specific expression patterns. The seemingly simple, constitutive Hedgehog response of *patched/Ptch1* is driven by a complex regulatory architecture, with batteries of context-specific enhancers engaged in promoter-specific interactions to tune signaling individually in each tissue, without disturbing patterning elsewhere. This structure—one of the oldest *cis*-regulatory features discovered in animal genomes—explains how *patched/Ptch1* can drive dramatic adaptations in animal morphology while maintaining its essential core function. It may also suggest a general model for the evolutionary flexibility of conserved regulators and pathways.

2.2 Introduction

Like other major developmental signaling pathways, the Hedgehog (Hh) pathway is broadly conserved among bilaterians, from its basic signal transduction mechanism to the DNA-binding specificity of its effector molecules, the zinc finger transcription factors of the GLI family (Briscoe and Thérond, 2013; Hallikas et al., 2006). Hh signaling is used for cell-cell communication in many contexts during development, and is maintained into adulthood to control tissue homeostasis. Disruption of signaling has been directly linked to several human cancers such as medulloblastoma and basal cell carcinoma in addition to developmental disorders including spina bifida, exencephaly and cleft lip/palate (Barakat et al., 2010; Scales and de Sauvage, 2009; Teglund and Toftgård, 2010). The Hh pathway itself is highly pleiotropic, regulating many different cell fate decisions in different cellular contexts, but the vast majority of the direct transcriptional targets of Hh/GLI respond to signaling in a strictly limited, tissue- and stage-specific pattern (Barolo, 2002). A very small number of target genes, such as *patched* in flies and *Ptch1*, *GLI1*, and perhaps *Hhip1* in vertebrates, all of which encode components or modifiers of the Hh pathway itself, seem to respond to Hh signaling in a universal, constitutive manner (or nearly so), to the extent that the expression of these target genes is sometimes considered diagnostic for the presence of Hh signaling (Epstein, 2008; Shahi et al., 2015). Although the regulation of these "universal" target genes, which respond to Hh/GLI in most or all signaling contexts, has been studied to some extent—for example, an enhancer of *Drosophila patched* was the first direct Hh/GLI target discovered in any organism (Alexandre et al., 1996)—the basis of a constitutive response to Hedgehog/GLI, or to any pleiotropic signaling pathway, is not

understood. One could imagine a simple "master" response element that is universally activated by a given signal in all developmental and adult stem-niche contexts where that signal is present. But while efforts have been made to synthesize such generic response elements by multimerizing binding sites, with varying degrees of success (Barolo, 2006), so far no such element has been found in nature.

Here we examine the regulation of the *patched* gene, a universal transcriptional target of Hh/GLI signaling in *Drosophila*, and its mouse ortholog *Ptch1*, which has an equally universal response to signaling. The Hh receptor Patched (Ptc) normally inhibits the function of Smoothened, blocking signal transduction and favoring the production of the repressor isoform of the Ci/GLI transcription factor (GLIR). This inhibition is relieved upon binding of Hh ligand to Ptc, allowing the activator isoform (GLIA) to accumulate. Ptc is expressed broadly, keeping the Hh pathway silent in the absence of ligand, but it is also directly transcriptionally activated by GLI in all Hh-regulated tissues, where it moderates signal levels (Figure 2.1A). This constitutive negative feedback circuit, which is conserved from flies to mammals and is essential for normal development, has been extensively studied in *Drosophila*, especially in the contexts of the developing wing and embryonic ectoderm (Alexandre et al., 1996; Briscoe and Thérond, 2013; Chen and Struhl, 1996; Milenkovic et al., 1999). A previously characterized promoter-proximal enhancer of *Drosophila patched* (Alexandre et al., 1996), referred to here as *ptcprox* (Figure 2.1B), contains three optimal consensus GLI binding sites which are required to activate expression in Hh-responsive wing cells (Figure 2.1C-E), but was not examined in other developmental contexts.

Since *patched* is activated in all Hh-responsive tissues, we wondered whether this promoter-proximal element, with its cluster of optimal GLI sites, is capable of reproducing the universal Hh response pattern of its parent gene. Our previous work has demonstrated that high-affinity GLI motifs identical to those in *ptcprox* (GACCACCCA) can repress enhancer activity in embryos, even in Hh-responding cells (Ramos and Barolo, 2013; White et al., 2012). These results presented a potential paradox: the same GLI motifs that have been shown to repress transcription in embryonic stripes are also found in an enhancer of *patched*, a gene which is directly activated by Hh/GLI in those same cells. Two explanations that could reconcile these results were (1) that optimal GLI motifs can either activate or repress transcription in the same cells, depending on their *cis*-context, or (2) that the *ptcprox* element is not a universally Hh/GLI-responsive enhancer. To distinguish between these possibilities, we examined the *ptcprox* enhancer in the embryonic ectoderm, and found that it fails to respond to Hh/GLI in this context (Figure 2.1F,G). This is also consistent with reports that a larger 3.2 kb fragment including *ptcprox* is also insufficient for a complete embryonic response, but larger fragments that include extensive 5' sequences are sufficient to drive *ptc*-like embryo patterns (Forbes et al., 1993; Millard and Martin, 2008). Because the promoter-adjacent cluster of conserved, optimal GLI motifs produces a tissue-restricted expression pattern, it is not sufficient to explain the universal Hh response of the *patched* gene (Figure 2.1C-G), suggesting that the control region for embryonic stripe expression is located elsewhere in the locus. We then set out to identify the enhancer module responsible for *ptc* expression in embryonic

ectodermal stripes and to characterize its Hh/GLI response, hypothesizing that it may depend on lower-affinity GLI inputs than the *ptcprox* element.

Our findings show that the seemingly simple constitutive Hh response of *patched* is controlled by an unexpectedly complex multi-modular system of enhancers spread across the *ptc* locus. The mouse *Ptch1* locus seems to be regulated by a very similar overall structure, making this an uncommonly ancient *cis*-regulatory strategy in the animal genome, perhaps as old as the Hh-GLI-Ptc negative feedback circuit itself. We propose that this regulatory structure, in which many context-specific Hh-responsive enhancers together produce a "simple" constitutive response pattern, can explain why *patched* is an important locus for morphological divergence: it allows Hh signaling levels to be adjusted independently in each tissue and stage, without disturbing pathway activity in other developmental and stem cell contexts.

2.3 Results

2.3.1 The universal Hh response of *patched* is mediated by a large array of stage- and tissue-specific enhancers

For this study, we created a nomenclature of sequence fragments by dividing the 5’ intergenic sequences into regions A through Z (with A containing the *ptc* transcription start site) and the first intron into regions 1A through 1I, and then naming each sequence fragment after the 5’- and 3’-most regions included in that fragment (Figure 2.3A): for example, the *ptcprox* enhancer, which spans fragments D, C, and B, will be referred to in the rest of this report as module DB. To reduce bias in our search and to account for the possibility of multiple embryonic enhancers, we independently tested

overlapping fragments of the large 5' intergenic region and the large first intron of *ptc* as single-copy, targeted-insertion reporter transgenes in which GFP was driven by the *ptc* promoter (fragment BA) (Figure 2.3A). We focused on the 5' intergenic and first-intron sequences of *ptc* in this functional analysis because the other four introns of *ptc* are short (<350 bp) and poorly conserved and the 3' neighboring gene is located less than 700 bp downstream of *ptc*.

The BA promoter fragment alone drives broad, low-level gene expression in most or all tissues—suggesting that the broad, non-Hh-regulated aspect of *patched* expression may be controlled by this region—but it is not able, on its own, to drive Hh-responsive patterns in any tissue (Figures 2.3J,J''). FAIRE-seq and DNAse-seq datasets (McKay and Lieb, 2013; Thomas et al., 2011) show that chromatin states in non-coding regions of the *patched* locus are highly dynamic in the *Drosophila* embryo and larval wing (Figure 2.3B; Figure 2.3.). Furthermore, the promoter-proximal consensus GLI sites, which are essential for *ptcprox* activity in the wing (Figure 2.1D-E), are in regions of mostly closed chromatin in the embryo, consistent with reporter data suggesting that Hh/GLI regulates *patched* via different enhancers in the embryo vs. wing. Every tested fragment in this 27 kb region drives expression in *ptc*-positive cells within at least one Hh-patterned tissue (e.g., Figure 2.3C-I; see also data from (Jenett et al., 2012; Kvon et al., 2014). These tissues represent cell types from nearly every stage of the *Drosophila* life cycle, from multiple embryonic tissues types to larval imaginal discs (Figure 2.3C,D,F) to adult tissues such as the gut and stem cell/niche systems in testes and ovaries (Figure 2.3E,G,H). Hh signaling participates in the development and maintenance of all of these tissues and cell types (Chen and Struhl, 1996; Michel et al.,

2012; Sahai-Hernandez and Nystul, 2013; Takashima and Murakami, 2001; Zhang and Kalderon, 2001). In many tissues, such as the embryonic ectoderm, multiple *ptc* modules are active. However, no single enhancer recapitulates the complete *ptc* expression pattern (Figure 2.3C-I; Figure 2.4). Together, these findings are consistent with DamID data indicating that Ci, the *Drosophila* GLI, occupies the entire *ptc* locus in embryos (Biehs et al., 2010). Distributed, multi-modular enhancer arrangements have been observed at other developmental gene loci (Barolo, 2011a), but *ptc* appears to be an unusually complex *cis*-regulatory locus—an idea supported by unusually dynamic chromatin structure (Figure 2.3B; Figure 2.4) (Thomas et al., 2011).

Bypassing enhancer-promoter specificity reveals diverse sub-patterns among segmentation stripe enhancers of *patched*

We find that, as in other tissues, the control of *patched* expression in the embryonic ectoderm is distributed among multiple stripe enhancers distributed across the locus. A small promoter element, fragment BA, which does not include the cluster of three optimal GLI sites (Figure 2.3A), drives broad, low-level, non-Hh-regulated expression (Figure 2.3J,J'), while several separable embryonic enhancers, when examined in vivo as GFP reporters containing the BA promoter element, drive *patched*-like segment-polarity stripe patterns (Figure 2.3K-N). In order to distinguish the information encoded in these embryo stripe enhancers from that provided by the endogenous *ptc* promoter, we also tested several *ptc* stripe enhancers with a minimal heterologous Hsp70 TATA/Inr-containing promoter ("hspmin") driving GFP. When paired with the hspmin promoter, three *ptc* enhancers drive *ptc*-like responses in the embryonic ectoderm (Figure 2.5; Figure 2.6). The intronic element 1EH is activated in

stripes to the anterior and posterior of each segmental stripe of Hh-expressing cells, much like the *ptc* gene's expression pattern in this tissue, though with broader stripes (Figure 2.5C-D). The distal 5' enhancer VT is active in a subset of *ptc*-positive cells on the anterior side of Hh-producing stripes in each segment (Figure 2.5H), while the more proximal 5' enhancer LK drives expression in *ptc*+ cells to the posterior of each Hh stripe (Figure 2.5K). Conditional overexpression of the Hh ligand augmented the posterior-stripe expression of enhancer LK, but did not affect the context specificity of its expression pattern (Figure 2.7).

2.3.2 GLI binding site affinity informs tissue specificity

We analyzed the sequence of the embryonic stripe enhancers described above to identify putative binding sites for the zinc finger transcription factor Cubitus interruptus (Ci), the *Drosophila* GLI, using an in vitro binding dataset (Hallikas et al., 2006) which corresponds well with GLI binding preferences in vivo (e.g., (Peterson et al., 2012) (see Methods). These Ci/GLI motifs were then altered by overlap extension PCR to abolish Ci binding (see Figure 2.8 for in vitro binding assays). Mutating Ci/GLI motifs in each embryonic stripe enhancer of *ptc* severely reduced its Hh responsiveness in vivo (Figure 2.5), demonstrating that low- to moderate-affinity GLI motifs—which account for most or all of the Ci/GLI sites in these modules (Figure2.8)—are required for strong activation in the embryonic ectoderm. Enhancer 1EH contains one optimal GBS (GACCACCCA), but targeted mutation of this motif did not have a strong effect on enhancer activity (Figure 2.5G, compare to 2.5D).

To test whether the low affinity of GLI sites in these enhancers contributed to transcriptional activation, we converted two of the low-affinity motifs in both the 1EH

module and the VT module to optimal GACCACCCA motifs. Because this two-site optimization increased Ci binding to those sites in vitro (Figure 2.8), and augmented the GLI1-responsiveness of the enhancer modules in mammalian cells (Figure 2.5M), we conclude that the sequence alteration improves Ci/GLI binding as predicted. In vivo, however, the GLI-optimized 1EH and VT enhancers are significantly less active in the embryonic ectoderm than their wild-type counterparts (Figure 2.5F,J; compare to 2.5D,H). Though perhaps counterintuitive, this finding is consistent with our group's previous reports that optimizing Ci/GLI motifs in Hh-regulated embryonic ectodermal enhancers of the wingless and stripe genes causes a decrease in enhancer activation in vivo (Ramos and Barolo, 2013; White et al., 2012). Interestingly, this effect appears to be tissue-/stage-dependent: optimizing two GLI motifs in the VT module decreases its activity in the embryonic ectoderm, as described above, but the same alteration conversely causes an increase of enhancer activity in the wing disc (Figure 2.5J'), where VT is normally weakly active (Figure 2.5H'). These results suggest that there is no level of GLI occupancy that allows optimal activity of this *ptc* enhancer in both the embryonic segments and the wing disc. Thus, nonconsensus GLI motifs can influence both the strength and the stage/tissue specificity of transcriptional responses to Hedgehog signaling.

2.3.3 Multiple wing disc enhancers of *patched* depend on GLI motifs of varying affinities to synergistically activate gene expression in response to Hh

The *ptc* response to Hh/GLI in the wing, as in the embryonic ectoderm, is distributed across many regions of the locus; in fact, most of the enhancers we tested in vivo had some activity in wing imaginal discs, though not all drove a strong or strictly

ptc-like expression pattern. We examined a few of these enhancers (Figure 2.9A)—all of which correspond to regions of accessible chromatin in wing discs (Figure 2.3B; (McKay and Lieb, 2013))—in greater detail with respect to their direct GLI inputs, using GFP reporters with a minimal TATA+Inr "hspmin" promoter. All of these wing stripe enhancers depend on GLI motifs for full activity in the wing imaginal disc (Figure 2.9B,D). However, the number and quality of GLI motifs in these modules bears no obvious relationship to the strength of enhancer activity in Hh-responding cells. For example, the previously characterized wing enhancer *ptcprox* (referred to here as module DB), with its cluster of three well-conserved optimal sites, is a weaker wing enhancer than several other regions of the *ptc* locus with fewer and/or lower-affinity GLI motifs (Figure 2.9B-D; Figure 2.10). Presumably this reflects a requirement for other transcription factor inputs into these modules, but it also demonstrates that binding site quality and number are not always accurate predictors of transcriptional responsiveness in vivo.

When these five wing enhancers (paired with the same hspmin promoter) were tested for GLI responsiveness in GLI1-expressing NIH/3T3 cells as luciferase reporters, we found that their levels of activation were not well correlated with either their expression levels in wing discs or the number or quality of GLI sites (Figure 2.9E, compare to 2.9C,D). However, the relative change in expression upon mutation of the GLI motifs, both in vitro and in vivo, was usually greater in modules with higher-quality GLI sites such as DB and ZY, compared to modules with weaker sites such as HF and YU (Figure 2.9D,E). Taking these results altogether, we conclude that, while GLI motif quality plays a role in Hh/GLI responsiveness (positively or negatively, depending on the

developmental context), it is not in itself a good indicator of which enhancers will be the most strongly active in a given tissue. Note that of the five wing enhancers examined here, the two with the highest expression levels in wings (HF and YU) have relatively few and/or low-quality GLI motifs, which nevertheless make significant contributions to their activity.

We were intrigued to learn from the above analysis that the canonical *patched* wing enhancer *ptcprox*/DB, with its cluster of optimal GLI motifs, drives relatively weak expression in vivo (though it is strongly responsive to Hh/GLI in vitro). This result led us to hypothesize that the high-affinity GLI sites in the DB module may require additional flanking sequence in order to exert a strong influence on gene expression in vivo. We first addressed this by testing a nested set of fragments from the -5.5 kb region upstream of *ptc*, all containing the 3 proximal optimal GBS (Figure 2.9F, top). All of the fragments (DB, GB, HB, and JB, in order of increasing size) drove *ptc*-like stripes in the wing, but while the anterior-posterior positions of these stripes were the same, the levels and dorsal-ventral extents of gene expression increased greatly with the addition of 5' flanking sequence (Figure 2.9G,H). 5' sequences containing relatively few and weak GLI motifs provided significant boosts to gene expression in the wing: for example, compare GB to DB, and HB to GB (Figure 2.9G,H,K). We next examined the direct contribution of the three optimal GLI sites (located within region D) to the in vivo activity of the nested fragments DB, GB, and JB. In all cases, mutating these three GLI sites (while leaving any other GLI motifs intact) significantly reduced peak gene expression in the wing stripe (Figure 2.9K) (Student's t-test; $p<0.0015$). However, the relative contribution of those three sites, measured as reduction of expression of a

3xGLI-mutant enhancer compared to wild-type, decreased as more *cis*-regulatory context was included—from a 99.9% reduction in module DB, to a 62% reduction in module GB, to a 26% reduction in module JB (Figure 2.9K). When taken out of their native sequence context and tested as a synthetic reporter, three optimal Ci/GLI sites are not capable of activating gene expression in the wing, though they can strongly synergize with other activator binding sites (Ramos and Barolo, 2013).

By breaking down module JB into smaller, non-overlapping sub-fragments (JI, H, GE, and DB; see Figure 2.9F), we were able to test the individual contributions of these regions, with very different predicted GLI binding patterns, to gene activation. Note that of these fragments, only module DB contains the cluster of three optimal GLI motifs. All of these modules were sufficient to drive expression in the wing disc, and as with the more distal *patched* wing enhancers examined above, little or no correlation between predicted GLI affinity and in vivo expression levels was observed (Figure 2.9G,I). In general, transcriptional synergy among sub-fragments was observed in vivo, such that the activity of a large fragment (JB, HB, GB) was greater than the sum of the individual activities of its constituent sub-fragments (Figure 2.9H-J).

The above findings suggest that, even in the wing, the promoter-proximal cluster of optimal GLI sites makes a relatively minor (though significant) contribution to the Hh/GLI response of *patched*, and that inputs from many additional *cis*-regulatory regions, some of which contain relatively few and/or weak GLI motifs, are integrated to produce the final pattern of transcriptional activation in this tissue.

2.3.4 An unusual tissue- or stage-specific Polycomb/Trithorax response element (PRE) alters the output of *patched* enhancers

In addition to the multi-modular complexity of the *ptc* locus, we found evidence of higher-level control of gene expression via specialized enhancer-promoter interactions. Some *ptc* enhancers drove qualitative and quantitatively different in vivo expression patterns when placed in *cis* to a homologous vs. a heterologous promoter (e.g., Figures 2.3M, 2.5K; see also below). In third-instar larvae, the *ptc* promoter and 5' sequences are bound by the Polycomb group (PcG) factors Pleiohomeotic (Pho), Polyhomeotic (Ph), Polycomb (Pc), and Posterior sex combs (Psc), as well as the Trithorax group (TrxG) factor Trl/GAF (Oh et al., 2013; Schaaf et al., 2013), and are marked with the repressive histone modification H3K27me3 (Figure 2.11A). These are all hallmarks of Polycomb/Trithorax response elements (PREs), *cis*-regulatory sequences which mediate epigenetic regulation by PcG/TxG factors (Bowman et al., 2014; Kassis and Brown, 2013; Schaaf et al., 2013). In embryos, however, the *ptc* promoter lacks PRE signatures: compared to the PRE-containing loci *en/inv*, AbdB, and Ubx, the *ptc* promoter region has low PcG occupancy (except for Pho) and low H3K27me3 (Schuettengruber et al., 2009) (Figure 2.12). A motif analysis of DNA sequences located 5' of the *ptc* promoter using the CIS-BP database (Weirauch et al., 2014) identified conserved binding motifs for the PcG protein Pho (ortholog of vertebrate YY1) and the TrxG transcription factor Trl/GAF (Figure 2.11B; Figure 2.13), consistent with our tissue-specific ChIP-seq binding data. These motifs are frequently enriched in PREs and often required for their function (Kassis and Brown, 2013). We also compared the core promoter motifs present in our two test promoters. The *ptc* core promoter region lacks a TATA box but contains a *Drosophila* Initiator element (Inr) and multiple GAF-binding GAGA motifs, as well as a pause button (PB) and a Downstream Promoter

Element (DPE) 3' of the transcription start site (Figure 2.11B; Figure 2.13). The GAGA + Inr + [PB and/or DPE] promoter configuration is associated with RNA polymerase II stalling at *Drosophila* promoters (Hendrix et al., 2008), and embryonic GRO-seq data identify *ptc* as one of the most heavily promoter-paused genes in the *Drosophila* genome (Saunders et al., 2013). The *ptc* promoter is highly dissimilar from the minimal Hsp70 "hspmin" promoter fragment used here as an alternative heterologous promoter, which has a TATA + Inr motif configuration (Emanuel and Gilmour, 1993).

When paired with the heterologous hspmin promoter, the *ptc* embryonic stripe enhancers PN and LJ activate expression in restricted subsets of the *ptc* segment-polarity stripe pattern (dorsal and posterior *ptc*-positive cells, respectively, within each embryonic segment) (Figure 2.11, left). However, when paired with the native *ptc* promoter fragment BA—which alone is broadly and weakly active (like the *ptc* gene itself) but unresponsive to Hh/GLI (Figures 2.3J,J' and 2.11C, top)—these enhancers drive stronger and more complete, *ptc*-like expression patterns in embryos (Figure 2.11C, right). Neither the core *ptc* promoter, fragment A, nor the upstream element, fragment B, were sufficient to enhance the Hh/GLI response of module PN (Figure 2.14A-H), suggesting that both the *ptc* core promoter, region A, and the promoter-proximal fragment B are required for this enhancement.

We next examined the promoter effects on *ptc* enhancer function in larval imaginal discs—where, in contrast to the embryo, we found strong PcG binding and the Polycomb-produced repressive mark H3K27me3 at and upstream of the promoter (Figure 2.11A; compare with Figure 2.12). Three of four tested wing-stripe enhancers (GC, YU, and ZY) showed reduced activity when paired with the *ptc*BA promoter,

compared to the heterologous hspmin promoter (Figure 2.11D,E). A fourth wing enhancer, HF, showed no difference in peak levels of activation, but had a higher baseline expression level across the wing when paired with *ptc*BA (Figure 2.11D,E).

Taken together, our genomics and functional results suggest that interactions among multiple regions of the *cis*-regulatory apparatus, including distal enhancers, 5' promoter-proximal sequences, and the core promoter are integrated to determine the final pattern and levels of gene expression in each tissue, in a manner that correlates with the presence or absence of PcG/TrxG regulation of a tissue- or stage-specific PRE.

Interestingly, the TrxG factor GAF—which has been associated with transcriptional activation in addition to its role in PRE-mediated repression (Adkins et al., 2006)—directly occupies the intronic embryonic stripe enhancer 1EH in the embryo, though to a lesser extent than the *ptc* promoter (Figure 2.11A; Figure 2.12), probably via conserved GAGA motifs in that element (Figure 2.15 and Figure 2.16). This may explain why the 1EH module, the most powerful embryonic stripe enhancer at the *ptc* locus, does not require the heavily GAF-occupied *ptc* promoter for strong activation in the embryo (Figure 2.14). GAF has recently been implicated in large-scale changes in gene regulation during the embryo-larva transition (Blanch et al., 2015).

2.3.5 *Drosophila* and vertebrate *patched* orthologs share a similar *cis*-regulatory strategy

Direct transcriptional activation of *ptc* by Hh/GLI, for the purpose of negative feedback regulation, is a regulatory circuit that predates the protostome-deuterostome split (Alexandre et al., 1996; Goodrich et al., 1996). In light of this unusually ancient evolutionary linkage, we asked whether the *cis*-regulation of the mammalian *Ptch1* gene resembles that of its insect ortholog. Two GLI-bound enhancers of mouse *Ptch1* have

been reported: a high-affinity promoter-proximal neural tube module v(Vokes et al., 2007) and a lower-affinity intronic limb bud mesenchyme module (Lopez-Rios et al., 2014). Our ChIP-seq analysis and previous studies (Lopez-Rios et al., 2014; Peterson et al., 2012; Vokes et al., 2007) identify many sites of significant GLI binding near m*Ptch1* and extending far upstream of the gene (Figure 2.17A,B). We tested seven of these GLI-bound regions as lacZ reporters in transgenic mice. As with enhancers of *Drosophila ptc*, we found that all tested *Ptch1* modules are active in different, overlapping subsets of the *Ptch1* expression pattern: no single module, no matter how strongly bound by GLI, recapitulates the complete Hh/GLI-responsive expression pattern of the parent gene (Figure 2.17A). Binding of GLI transcription factors to numerous sites in the *Ptch1* locus is dynamic across several different developmental contexts, perhaps suggesting that enhancer modules are differentially accessible in a developmentally regulated manner (Figure 2.17B, red arrows). Even taken together, the novel enhancers that we identify here and the previously identified modules do not account for the complete *Ptch1* pattern, suggesting that there are still more yet uncharacterized *Ptch1* enhancers to be found—presumably with relatively lower levels of GLI factor binding, since the highest-occupancy ChIP-seq peaks have now been tested. As in fly *ptc*, chromatin at the mouse *Ptch1* locus is H3K27-trimethylated at the promoter and in upstream noncoding regions (Figure 2.17C, compare with 2.11A), characteristic of PcG regulation (Bowman et al., 2014; Kassis and Brown, 2013).

Mouse *Ptch1* and fly *patched* share a general gene-expression profile (broad, weak basal expression plus a constitutive response to Hh/GLI), as well as a complex multi-enhancer organization in which the universal response to Hedgehog signaling is

not constituted in a master GLI response element, but distributed across many sub-tissue-specific GLI-regulated modules (Figure 2.17A,D,E). The GLI-regulated enhancers discovered to date at both the *Drosophila patched* and vertebrate *Ptch1* loci range from remote 5' elements, to promoter-proximal modules with optimal GLI motifs, to intronic enhancers. Hi-C topographical domain mapping and ChIP-seq occupancy data for the chromatin insulator/boundary factor CTCF indicate that both *Ptch1* and *patched* are flanked by a nearby 3' boundary (associated with a closely neighboring downstream gene) and a far more remote 5' boundary (Dixon et al., 2012; Wang et al., 2012) (Figure 2.17A,D,E; Figure 2.18). In both fly and mouse, all mapped enhancers fall within this topographical domain; they rely on GLI sites across a wide range of affinities; they appear to also require additional tissue-specific transcription factor inputs (Supplementary File 2); and they act through a PcG-bound, H3K27me3-marked promoter. This combination of regulatory mechanisms, ranging in scale from the single binding site to the locus as a whole, allows independent tuning of *ptc*/*Ptch1* expression—and therefore of Hh signaling levels—in individual tissues, without disrupting overall signaling required for normal development and adult tissue homeostasis (Figure 2.17A,D,E).

2.4 Discussion

This chapter, which examines the regulation of an essential developmental gene at the locus-wide level, demonstrates for the first time how a constitutive pathway-responsive gene is precisely regulated in diverse developmental contexts. This regulation consists of two major modes of control: First, *ptc* must be broadly active to

inhibit Smoothened activity, preventing the inappropriate activation of Hh target genes and subsequent faulty patterning, developmental defects, and disease (Barakat et al., 2010; Scales and de Sauvage, 2009; Teglund and Toftgård, 2010). Our reporter-based analysis of the locus has identified a promoter-proximal 5' sequence, region B, which in conjunction with the core promoter acts as the likely controller of the basal, signal-independent mode of *ptc* expression.

Second, *ptc* must be able to respond to Hh/GLI signaling in many developmental and stem-niche contexts. We show here that this is not accomplished by a master Hh/GLI response element, but rather by a collection of GLI-regulated enhancers spread across the *ptc* gene locus. These remote, tissue- and sub-tissue-specific enhancers work synergistically and are integrated with additional information from promoter-proximal sequences and the core promoter, responding to Hh signaling differently in each stage and tissue type. The basis for the tissue-specific activity of the numerous enhancers of *ptc* has not been directly examined yet, but in some cases we have identified likely candidate selector inputs by a combination of sequence motif searching and DNA conservation analysis (Figure 2.16; Figure 2.17D,E).

At least in some enhancer modules, we find that the quality/affinity of GLI binding motifs in these enhancers is extremely important for proper patterning, to the extent that optimizing GLI motifs in certain enhancers can cripple enhancer activity. Previous studies from our group and others have shown that low-to-moderate-affinity transcription factor binding sites can be important for limiting the response of a target gene to a pleiotropic, broadly active transcription factor or pathway (Crocker et al., 2015; Farley et al., 2015; Swanson et al., 2011); additional references within). Our

results with embryonic stripe enhancers of the *ptc* gene, however, point to a different type of regulatory control by weak binding sites, in which the occupancy of a transcription factor is modulated to an optimal level of activation within a tissue. The fact that optimal activation is achieved via suboptimal GLI binding sites in the *Drosophila* embryo may reflect the fact that Ci/GLI acts both as an activator and as a repressor through the same binding sites. Considering that the area and duration of Hh production in embryonic segments is vastly lower than that in larval/pupal wings, where half of the organ produces Hh protein for most of the developmental period of the animal, it is to be expected that the GLIA:GLIR ratio may be significantly lower in Hh-responsive cells in the embryonic segments than in late-larval wing discs. If so, the relatively high concentration of GLIR in Hh-responding embryonic cells may create selective pressures favoring lower GLI occupancy, as we have argued previously (Parker et al., 2011; White et al., 2012). Alternatively (or in addition), other mechanisms, such as overlapping binding specificities for non-GLI transcription factors, could play a role in the sub-optimization of GLI motifs observed here.

Our results from transgenic mouse reporter constructs and GLI ChIP-seq experiments suggest that the same large-scale *cis*-regulatory strategy at work in the fly locus is also present in vertebrates—at least to the extent that many enhancers, with widely varying levels of GLI input, act in combination but in a tissue-specific manner to independently adjust *Ptch1* expression levels in each cell type, giving the illusion of a universal response to Hh signaling but in fact representing a complex array of context-limited responses. However, we do not mean to propose that the specific *cis*-regulatory sequences that regulate *patched* and its orthologs are conserved across such large

evolutionary distances. Individual *patched* enhancers show no obvious sequence conservation from insects to mammals—unsurprisingly, given that most *ptc*-expressing, Hh/GLI-regulated organs present in modern animals did not exist at the time of the protostome-deuterostome divergence. Rather, we propose that *patched* and its orthologs share an ancient regulatory structure composed of relatively young enhancers. Based on our motif analyses and mutational studies, we propose that *patched* enhancer evolution likely involved both tuning of GLI occupancy and the co-option of tissue-specific inputs to produce new aspects of the *patched* expression pattern over time. This elaboration of new sub-patterns would not be necessary if *patched*/*Ptch1* were regulated by a unified "master" Hedgehog response element; under such a simple regulatory scheme, new domains of *patched* expression would appear automatically whenever Hh signaling were activated in new contexts. By contrast, our analysis suggests that robust response to Hh/GLI in both embryonic segment-polarity stripes and in larval wing stripes may not be achievable by the same enhancer module, perhaps due to conflicting requirements for GLI occupancy, or perhaps for other reasons whose basis in *cis*-regulatory logic is not yet clear.

The *cis*-regulatory architecture discovered here is also a remarkably ancient example of a direct transcriptional linkage and gene regulatory strategy in animals. Notch regulation of Hes/Hey genes is a circuit of comparable age, but its conservation is limited to the presence of CSL-family binding sites controlling orthologous target genes (Rebeiz et al., 2012), with no evidence of a larger shared gene-regulatory structure. Conversely, while the collinear structure of animal Hox gene clusters predates the Bilateria, the upstream regulators of Hox genes do not appear to be shared between

protostomes and deuterostomes (Duboule, 2007). Our results suggest that the locus-wide multi-modular structure by which *patched* responds to Hh/GLI to provide feedback inhibition is an unusually old *cis*-regulatory strategy. This strategy has the potential advantages of transcriptional precision and robustness ascribed to regulation by 'shadow' enhancers (Barolo, 2011b; Frankel et al., 2010; Perry et al., 2011; Wunderlich et al., 2015), but more importantly it allows for the independent modulation of signaling levels in a stage- and tissue-specific manner, without disturbing pathway activity in other contexts.

Our model has significant implications for the evolution of cell signaling pathways and their target genes and tissues. For example, it helps to explain two reported cases of morphological divergence caused by changes at the vertebrate *Ptch1* locus: modifications to an intronic enhancer of *Ptch1* produced major developmental adaptations in the bovine limb (Lopez-Rios et al., 2014), while sequence variation upstream of *Ptch1* is linked to cichlid craniofacial diversification (Roberts et al., 2011). In both cases, Hh signaling levels in other developmental contexts (for example, the spinal cord) were apparently unaffected by the adaptation, indicating that the affected enhancer module is stage/tissue-specific. Our conserved structure with flexible enhancers model of *patched* regulation is also consistent with the finding that, while zebrafish *Ptch1* resembles its mouse and fly orthologs in having multiple GLI-bound regions, few such sites are conserved among vertebrates (Peterson et al., 2012; Wang et al., 2013). Thus, while the *cis*-regulatory structure of *patched* appears to have ancient roots, it is also highly flexible, allowing a core component of the Hedgehog pathway to

function as an adaptable tissue-specific modifier of pathway activity and as a substrate for morphological evolution.

2.5 Materials and Methods

Transcription factor binding site prediction and ranking

Matrix similarity scores were calculated(Quandt et al., 1995) using in vitro Ci/GLI binding data (Hallikas et al., 2006). GBS were identified in silico by screening the *ptc* locus for defined motifs using GenePalette (Rebeiz and Posakony, 2004). PBEs (Polycomb-core-complex Binding Elements) were defined by Mohd-Sarip et al. (Mohd-Sarip et al., 2005).

DNA sequence alignments

Sequences and multi-species alignments were obtained from the UCSC Genome Browser (genome.ucsc.edu).

DNA cloning and mutagenesis

Wild-type *ptc* enhancers were amplified by PCR (Roche Expand High Fidelity PCR System) from BAC DNA (CH322-170A-12 or CH322-188E13). PCR primers are provided in Figure 2.19. Enhancer constructs were sub-cloned into the pENTR/D-TOPO plasmid (Life Technologies) by TOPO cloning. Enhancers tested with the hspmin promoter, taken from the *D. melanogaster* Hsp70 gene, were subsequently cloned into the pHPdesteGFP transgenesis vector via LR Cloning (Life Technologies). Enhancers tested with the endogenous *ptc* promoter were cloned by traditional methods into the

pStinger transgenesis vector (Barolo et al., 2000). Targeted GBS mutations were created by overlap-extension PCR (Swanson et al., 2010a). Promoter analysis described in Figure 2.11 and Figure 2.14 was done by replacing the minimal hsp70 promoter contained in pHPdeste with the designated *ptc* promoter region (see Figure 2.19 for sequences and restriction sites used). All enhancers and promoters were screened by restriction digest and sequencing.

Drosophila transgenesis

P-element transformation was performed as previously described (Swanson et al., 2010a) in the w1118 strain. Site-directed transformation by embryo injection was performed as described by (Bischof et al., 2007), with reporter transgenes integrated into a ΦC31 landing site at genomic position 86Fb.

Immunohistochemistry and confocal microscopy

Drosophila embryos and third-instar imaginal discs were fixed and stained using standard methods (Parker et al., 2011; Ramos and Barolo, 2013; White et al., 2012). Adult testes and ovaries were dissected between 0-2 days after adult hatching, fixed in 4% paraformaldehyde, washed in PBS with 0.1% TritonX, and subjected to antibody staining. Third-instar larval gut was dissected and fixed in the same matter as testes and ovaries. Primary antibodies used included rabbit anti-EGFP (Invitrogen 1:100), mouse anti-beta-galactosidase (Developmental Studies Hybridoma Bank 40-1a, 1:200), mouse anti-Engrailed (Developmental Studies Hybridoma Bank 4D9, 1:50), mouse anti-Wingless (Developmental Studies Hybridoma Bank 4D4, 1:50). In *Drosophila* embryos,

EGFP antibodies were used to visualize reporter expression; in imaginal discs, native GFP fluorescence was imaged directly. Antibodies obtained from the Developmental Studies Hybridoma Bank were developed under the auspices of the National Institute of Child Health and Human Development and maintained by the Department of Biological Sciences, The University of Iowa (Iowa City, IA). AlexaFluor488, AlexaFluor555, and AlexaFluo568 conjugates with secondary antibodies from Invitrogen were used at 1:2000 dilutions. DAPI was included in the Prolong Gold antifade mountant (Life technologies). Confocal images were captured on an Olympus FluoView 500 Laser Scanning Confocal Microscope mounted on an Olympus IX-71 inverted microscope, and on a Nikon A1 confocal microscope. Samples to be directly compared were fixed, prepared, and imaged under identical confocal microscopy conditions and settings. Quantitative GFP expression data from imaginal discs were collected from confocal images: all GFP-expressing fly stocks were crossed to the same reference line, dppD-Ci*ptc*, which drives DsRed expression in wing discs, as an internal normalization reference (Parker et al., 2011). DsRed and GFP were imaged for each disc, and GFP levels were normalized to peak DsRed fluorescence (Parker et al., 2011). Normalized GFP fluorescence data across wing discs (Figs. 4 and 5) were graphed in MATLAB (E. Ortiz-Soto, A.I.R. and S.B., manuscript in preparation). Two-tailed t tests for two samples with unequal variances were used to compare peak wing expression levels of GFP reporters.

Cell culture assay

NIH/3T3 cells were cultured at 37°C, 5% CO2, 95% humidity in Dulbecco's modified eagle medium (DMEM; Gibco, cat. #11965-092) containing 10% bovine calf serum (ATCC; cat. #30-2030) and penicillin/streptomycin/glutamine (Gibco, cat. #10378-016). Luciferase assays were performed by plating 2.5 x 104 cells/well in 24 well plates. The next day, cells were co-transfected using Lipofectamine 2000 with the DNA constructs indicated in each experiment in addition to ptcΔ136-GL3 (Chen et al., 1999; Nybakken et al., 2005) and pSV-Beta-galactosidase (Promega) constructs to report Hh pathway activation and normalize transfections, respectively. GLI1 cDNA was added where relevant to activate the Hh pathway. Cells were changed to low-serum media (DMEM supplemented with 0.5% bovine calf serum and penicillin/streptomycin/glutamine) 48 hours after transfection and cultured at 37°C in 5%CO2 for an additional 48 hours. Cells were harvested and luciferase and beta-galactosidase activities were measured using Luciferase Assay System (Promega, cat. # E1501) and BetaFluor -gal assay kit (Novagen, cat. #70979-3). Multiple assays were performed and each treatment group was assayed in triplicate. Two-tailed t tests for two samples with unequal variances were used to compare samples.

Quantitation of transgenic reporter expression data

Fluorescence data from wing confocal images were collected and quantified as previously described using the Matlab program Icarus (E. Ortiz-Soto, A.I.R., and S.B., manuscript in preparation) (Parker et al., 2011; Ramos and Barolo, 2013). Each experiment was performed at least two times, and fluorescence was measured from at least two wings per construct.

EMSA competition assays

Electromobility shift competition assays were performed as previously described (Parker et al., 2011). EMSA oligonucleotides sequences are provided in Figure 2.20.

hedgehog misexpression in *Drosophila*

Heat shock-inducible Hh (HS-hh) transgenic flies (Ingham, 1993) were crossed with flies harboring the *ptc*LK-GFP reporter transgene. Embryos were collected overnight at 25°C and heat shocked for 1 hour at 37°C, shifted to 25°C for 30 minutes, and fixed, stained and imaged as described above.

***Drosophila* larva ChIP-seq**

Protocol for carrying out ChIP in larval tissue has been described previously (Brown and Kassis, 2013), with minor changes: fixed brains and imaginal discs were dissected from 10 third instar larvae, and before incubating the sonicated chromatin with antibodies 3.3% of each sample was saved for input reactions. ChIP was performed with 1:100 antibody dilutions of anti-Pho, anti-Ph (a kind gift from Donna J Arndt-Jovin), anti-En (Santa Cruz Biotechnologies) and 1:200 dilutions of anti-H3K27me3 (Millipore, 17-622) antibodies. Following purification of immunoprecipitated DNA, Illumina libraries were prepared by using TruSeq DNA Sample Prep Kit V2 as described previously (ethanomics.wordpress.com/chip-seq-library-construction-using-the-illumina-truseq-adapters/). All ChIP-seq data sets were aligned using Bowtie (version 0.12.2) to the

Drosophila reference genome (releases 5.22 and 6.02). All ChIP-seq experiments were performed with 2 biological replicates.

ChIP-seq data can be accessed at this link http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=yrifkssspfuzbsp&acc=GSE76892.

Mouse ChIP-seq, transgenic reporters, and beta-galactosidase staining

E10.5 embryos were obtained from timed matings of Swiss Webster (Taconic) mice and micro-dissected in cold PBS to isolate the head, trunk (neural tube and somites) and limb buds. Each tissue was pooled separately and fixed for 30 minutes in 1% formaldehyde/PBS at room temp. Chromatin immunoprecipitation and embryoid body differentiation were performed as previously described (Peterson et al., 2012). Antibodies used for ChIP were goat anti-mouse GLI2 antibody (R&D Systems, AF3635) and H3K27me3 (Millipore, 07-449). High throughput single-end sequencing was done on the GAII or Hi-seq platform (Illumina) and reads were mapped using BWA (Li and Durbin, 2010). Sequence data for the 500 kb flanking *Ptch1* corresponds to (chr13:63,112,841-64,166,828 (mm9) have been deposited under accession number GSE71199. GLI3 ChIP-seq data correspond to GSE52939. Individual enhancer regions were defined by the conservation block encompassing the detected GLI binding region, amplified by PCR and cloned into a Gateway compatible version of the reporter construct previously described (Vokes et al., 2007). Mouse enhancer coordinates are provided in Figure 2.21. Transient transgenic embryos were analyzed at E10.5 and X-

gal stained for beta-galactosidase activity for 4h at 37C. ChIP-seq data are available (accession number GSE71199).

2.6 Acknowledgements

The nature of this work is highly collaborative, so there are many people to acknowledge in this chapter. Andrea Ramos, a former graduate student in the Barolo lab, generated the majority of wing data. Kevin Peterson, a researcher at Jackson Labs and former postdoctoral fellow in Andy McMahon's lab, generated all mouse data. Brandon Carpenter, a former Graduate Student in Ben Allen's lab generated the Luciferase data. Sandip De, a postdoctoral fellow in Judy Kassis' lab generated the polycomb ChIP-seq data in *Drosophila*. Additional authors who significantly contributed to this work are as follows: David S. Parker, Lauren E. Hillers, Victoria M. Blake, Yuichi Nishi, Matthew R. McFarlane, Ason C.Y. Chiang, and Scott Barolo. This research was supported by the Cellular and Molecular Biology Training Grant (NIH T32-GM007315) to D.S.L. and A.I.R., the University of Michigan Reproductive Sciences Training Program Fellowship to D.S.L., the Center for Organogenesis predoctoral fellowship (NIH T32-HD007505) to A.I.R. and D.S.L., a Center for Organogenesis Research Team Grant to B.L.A. and S.B., and by NIH grant GM076509 and NSF grant MCB-1157800 to S.B. A.P.M. is funded by National Institutes of Health (R01 NS033642). B.S.C. is supported by an EBS EDGE award. B.L.A. is funded by a Scientist Development Grant (11SDG638000) from the American Heart Association, and by a grant from the National Institutes of Health (R01 DC014428). We also thank Harold Smith (NIDDK) for Illumina NGS. S.D. and J.A.K. were supported by the Intramural Research Program of the NICHD/NIH. We thank Doug Epstein, Lindy Jensen, Charles Katzman, Annie Azrak,

Shelby Peterson, Autumn Holmes, Lisa Johnson, Jessica Frick, Eleanor Smith, Katherine Gurdziel and Elliott Ortiz-Soto for assistance. We thank the staff at the Microscopy and Image Analysis Laboratory (MIL) at the University of Michigan Medical School for assistance obtaining images, the members of the University of Michigan developmental genetics meeting, and the UM*fly* community for helpful discussions.

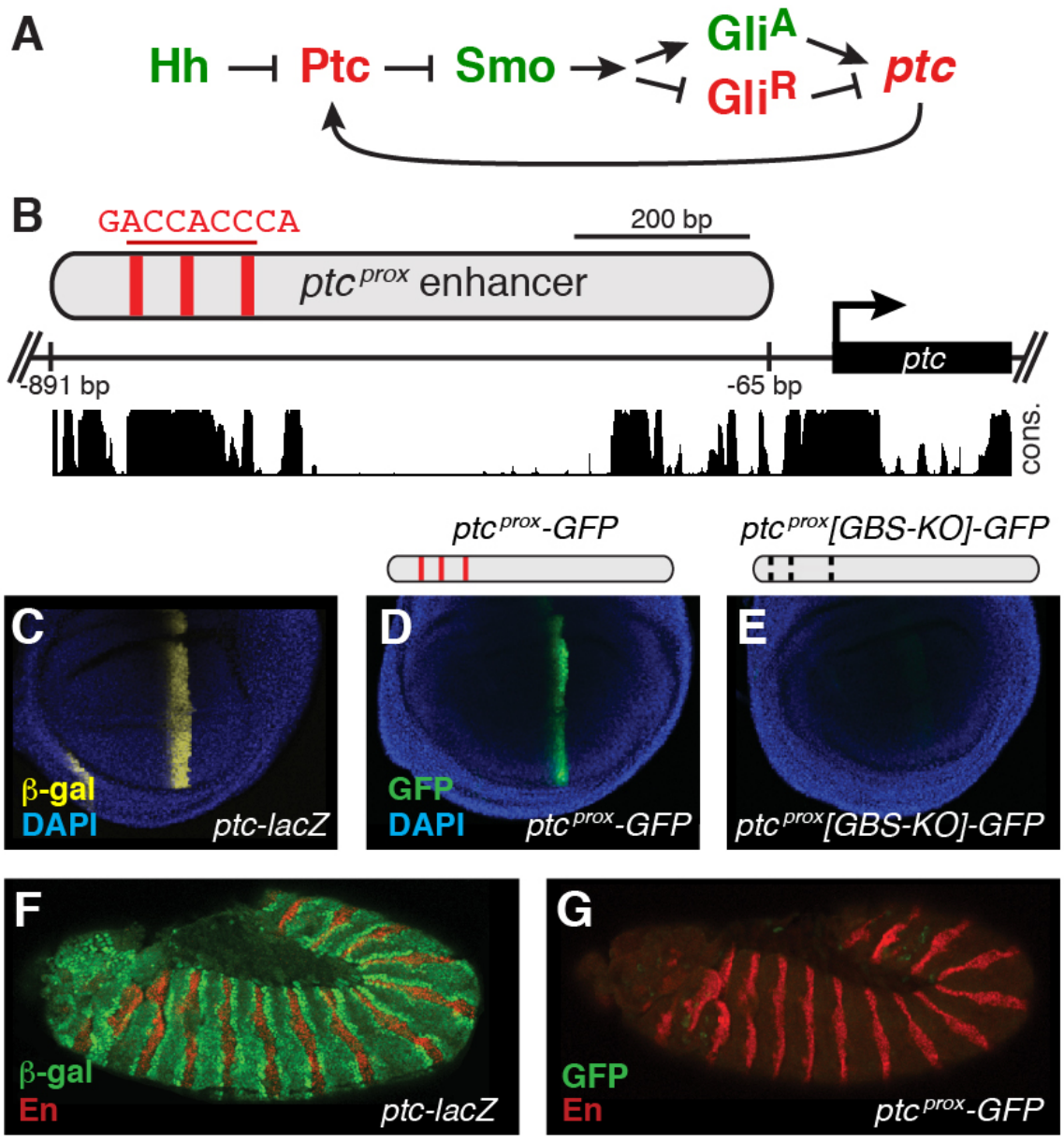


Figure 2.1 The promoter-proximal Hedgehog/GLI-responsive enhancer of *patched* is not sufficient to respond to Hh signaling in the embryo

(A) Summary of the conserved Hh-GLI-Ptc pathway feedback loop. (B) Promoter-proximal region of *D. melanogaster patched* (*ptcprox*). Optimal Ci/GLI binding sites are shown as red bars; sequence conservation among *Drosophila* species is indicated by the histogram at bottom. (C) Wing imaginal disc from a *ptc-lacZ* enhancer-trap larva, showing the *ptc* wing expression pattern. (D-E) Larval wing discs carrying *ptcprox*-GFP reporter transgenes with the minimal hsp70 promoter and either intact (D) or mutated (E) GLI sites. Sequences can be found in Supplementary File 1. (F) Stage 12 *ptc-lacZ* embryo. (G) The *ptcprox* enhancer is unresponsive to Hh in the embryonic ectoderm.

ptc^{prox} enhancer (DB) chr2R:4,536,264-4,537,090 827 bp

ATGCATGCGCAGCCTGCCACGCACGCGCTTCCCCCAAACAAATACACACACACACACTGAGACG
AAAGCTCCATTGGGCAGCGCTGCCGACGCTGAAGGCCGACATCGGCAGAGCTGAACGTTTGGGT
AGGGGACCACCCACATCGCTTGGCGGTTTCAGTTTAATGAAGGCAGAAACAAATTTATTTT
---A-AA-- Ci-1^{ko}

TGGGTGGTCCACACTGCAGCGAAAATAAACTACAGTGGCAACAACAAACCAGCAGCCAAGGCAC
--TT-T--- Ci-1^{ko}

TTTGGGTGGTCCATGCAAAAAAAAAACAAATTACGGCATGCGAATAACAATAGAAATTAGCGCT
--TT-T--- Ci-1^{ko}

CTCGTGGCGGAGCTATTTGGGTATATTAGAGCTACATATTTTATTTGTTTATAAAAAGTATAAA
TGTAAACAATGAGTTCCAAGCATTAAGTCCGTATGCTCAACAATTACATTATCATTATTATTAT
CACTTAAATATTTACAAAGGATATTTAAACAGTAATAGATATATATTTTATTTCTTAATTTCTG
TTAACATATGTATTTACATTGGTAGTTATTCTTTATTTTGCAACAAGCATTCATAAATTTTATA
TAACAAACTTGGTATTTTCTCGGAAAAACTCCTGAATCACCCCTCGGTATTTTGTGCGTTGAGC
TATCGTTAAAGCAGCCCTCGCAGAGAGCGTTCTCAAACCAAAATGGCCGCACACGAAACAAGAG
AGCGAGTGAGAGTAGGGAGAGCGTCTGTGTTGTGTGTTGAGTGTCGCCCACGCACACAGGCGCA
Ci-63
AAACAGTGCACACAGACGCCCGCTGGGCAAGAGAGAGTGAGAGAGAGAAACAGCGGCGCGCG

VT chr2R:4,522,304-4,523,363 1060 bp

CCGCATACCCTATGATATGACTCGACAGACCCCGAAGACTGAGATAATACACTTTCCGCATACC
CTATGATATGACTCTGCTTAAAAATGTAATGGTACTCAATCAAAATGTAACCATTGGTTTTTCT
TTTAAATATCAGCTATAGAAATGTCATATATCCTTAATTAAAACAATATTCCAATAATGTCTTG
CTGTATGATATATTTTCCTAATGTAATAAATATAGACTTATTATTATGCAAATTTTCGCTGCGT
GTTGCTGTGTATGTGTGATTGGTGAATGGCATGTTGCCACCCAAGTTGCCAGTTGGCTGGCCAA
Ci-93^{ko} ---A-AA-- Ci-141^{ko} --TT-T---
Ci-1 GA------- Ci-1 ---G---T-

GACTTAATAGCCGCTTTCATTGGCCAAAAGCCTGATAAAAGACCGCACACGAAGCTCTTCTTCG
Ci-59^{ko} ---A-A---

GAGCCAAGGCAGTCGTGGCTGTGTGGTGGTTTCGTGGGTTTGATGGTTGCTGGCTTGGTGGTTT
Ci-196^{ko} ---T-T--- Ci-203^{ko} --T--T---

GGCGGTTGGCCTTTCGACGTCTCATCGGCAACATTGCGTTCGCCTTGATGTTTATCGGCAATAT
TTAGCTGCAACATGTTTTGCCAAGGCATCAAACGGTAGCTCTTCTTAATTACAACAACCAAAAC
AACAGCCAGCCAGCTGGCCATTAAGATAAATCGGAAACGGAAATCGCTAATGAAAAATTTCCAT

ATTTATGGTAACGACAAAGAGCCGCTTCGCTGCAAAGGCAAACCGAAAACCAAAAGCAATAGCA
AAAACAATAACAAAAGCAACAAAAACAGCAACGGCAACCACTAAACAACTAAATTTCGTTGTCG
ACTGATTGGGGGGATGTGGATGT**GGCCACCAA**CGACGAACCAACCACT**TGGGGGGTC**CAGCCTC
Ci-47^{ko} ---A-AA-- **Ci-5^{ko} --TT-T---**

GCGTGATTTGTTCTCTTTTTTTTTTTTTTTCATTTTCATTTTATTTTTTTCTGGCCCGCCGTTT
TGCTTTTGGGCCCCGTCTCGTGTCTTTA**TTGGTGGTA**CGGGCTGCAATTAATTCCAACGAAATT
Ci-195^{ko} --TT-T---

TAATTAGAATTAAAATGTATATTTCATTGACTGGGCGCGATAAGGCAGCGCAAATGTCAATCAA
ACAAATGGGTCCCGAGAATATTAGGCGATTCGAATGCTGGCGCGAAAGATTAATCATAACTCAG
TCTTCGGGGTCTGTC

LK chr2R:4,530,049-4,531,613 1565 bp

CTACTTGGTTTGATAAATACGGTGTGTGTGAACCCAACTAATTGCTACTTGGTTTGATAAATGG
GCGCACGCCTAGAGCGGGCTAATCGGAAACTTCCGGCACATTGGAGCTCATAAATAACGACCAT
TGTGCACCTAGAAAAATTATTTAAATGTGATCTAAGTTCTAAGTTTCTAAGCAAAATGGGTTAT
ATGTAAATGATGATAAATATGCAACTGATCTGAGAAATTTGGATAAAGTTTGATGGACTTTATT
TGGAGAGGGAAGAGAAAAGTTCACATTTACTTTAAAAAATAATTGTATCACAAATAAAAGAGCG
AGAAAAGAGCTGATATTGAATATTATTGTTAAATTGAATATATACTTTTCCTTTAAAAAAGCAT
TGCATACTTTGCAACAGCATTAAAAGATACTTCAATCTCCTGTGTGCTGAAATTGAGAATGTAT
ACATATTGAACGTATTAATTAAAACAAGAAGCTGGAATTATGATTTAAATTTCTTATCGTAACT
TCCAGTTTATAATGGGTCATAGAAAGTGCCTATTTTTC**TGGGTGCAC**TGCTGGCCAAAACAAGT
--TT-T--- Ci-100^{ko}

GGACGA**GGCCACCAA**AT**GACGGCCCA**CGCCTGACCCCGACTGCGATGACCCAGGGACACTGTCG
---A-AA-- ---AA-A-- Ci-47^{ko},Ci-183^{ko}

GGCTTATCATGACCCGATATTGGACGACGTGTGCGGGCCGCCCCTGCCCCCGCCCCCCTCAGCT
GGATCAATTAAAACGCAGGCGAGGGCAGCGGGCGGCGGGCTAACAACTCAATACAATGTAAACA
AATGATATACGGTCGCAGGGCCACCACAGAGTCAGGGTCCATTTCCATTCCCATTTCCATGCCC
ATTTCCAGTTTGAGCTCGATTTTCATTTTCCGGAGCGATTCAGCGATTGAAACGTCTCTACCTT
TTCCACAAAAGCGACAGCCGCACAGCTGGACTCGGCCGACTGGCTCAGCAGTCGCTGCAAATCA
ATTCGAAAGAGATTTTCATTTGATCCGAGCTCGAGTTTGAGTCGTAAAAAGCGGGCCCGTGGTC
AGGCTAACATTCGACTGGTGATGGTGGTGGCCTGGTGGTCCCGGCCGATGTGCAGCGATCGTAT
CTGGCCAGATATGTCGGCCCAAACAAGTTGAGTTGTCAACGCAGCTCCGCAGGCCACGTGAAGT
GTCGTTGAGTTATGACTGTCCCGGGTTTTACCCTGAATCCAAATCCTAATCCGAAGCTGAAGCC
GCAGCTAGCTCAATGTTGGGTCTTTACGACTCTGCCGGCCATGTGGCTAATTGATCGTATTAAG
CTGTGGGAAAAGTAGTTTGGGTACTTGCTGATTATACATTGCTGATTGCACG**TGGGTGATC**CTC
Ci-16
GGCCAG**GAGCACCGA**CCGTCCAGGTGTCTTTTTCGATTGGATTTTTGAAAGTGCACAATCGAGT
---A--A---A-AA-- Ci-174^{ko}/184^{ko} (OVERLAPPING)

CTTGACCACAACATAGCTTTCTTCTCGTTGCCAGTGGGGTGTTGAGCTAAGCAAATGAAAAGAA
TGTTATAAACAAATGACTGTCTAAGCTTACACATTTGTAAGGCTTTATAATCGTGCTC
TGCCACCCACTCACACTTGGAGCCACACAATCTTCGGCCGCAAGTTCAACATTCAATTAGTTGG
---A-AA-- Ci-93^{ko}

GTTCACACACACCGT

1EH chr2R:4,542,479-4,545,408 2930 bp

GAAGTGCTTAACAAGTTAACTGTAATTTGACATGCTGGGCGTGTTGAAAGCAAAACACGATAAG
GTGTGCACGCCCAGAAAATGTCCACTTCTGGGAAATGTCATTTAAATACAATATGACCTGAAAA
ATTCGTAGCCATTCCATACAAAGAATTGTTAGTACTGAAGATTATGCCAATACCAATGGGTATA
TAGAGTCGTAGTTCCAAAAGACTTTACATGAACTTTAATCGGATTGGAGTTCATTCATAAAAAC
GCATCACTGCCGCTTTACATAATCCTTTGACCCATTATCAGGGGTTTGAACTTATTTTCCCCTT
CGAATGAACCCAATGGCCCACATAACTCCCACACAGTGATCCCACAATTGACAGCTTCATCTGC
TGCGGCCCCACACACTCAACGCACCAACGTGTCCAACTTGACTTGATCTGCCATCTGCAGCTGG
CGCGGGGAACTGGAACTCAAAAAAAAAAGAAAAGAAAAACACAAAATACAAAATACAAAATGTG
AGGGAAAAAATGAAAATAAACCGAGGAAAAAACACGAAATGCTGTTGTACAGGCGGCCACAGCT
TGGGCTTCATTAGGGACGGGACGGCAACGACTTTTGCCCGAGAGACGAGCCGCACATAAAACTG
TTATTGTCTTCATCAGTCTGAGGTCTCGGCGAAGCAAGACATAACAACGGCCCGACAGAGAGAG
AGAGAGAAGAAAAATCGGGAGAATTATGAAGACATTAACTC**GACCACACA**GCACGCTGCCGTAC
---A-A--- Ci-9^{ko}

CCGTACCTATATACCTATACCCAACCCATAACCATACACACACACACACACACACGCATCAA
CACACACACACACACGAACAGACAGGCCCAAAAACTCGAAACCCCAGCAGAGAGAAGCGGGGGC
TTACTTACGTTACGATCAAGACCTAGAGCCGAGCCAGAAAAAGGTATACTGCAGAGACAGAGAG
GAGGGCACAGTGACAGAGAGCGAATACCGGAAAGAAACATTCAAGCAATAATCACGGAATTCAC
TGAAAAAGCAGTTAACGGACAGAGGTCCACGGACGGACTGGCGGACGGACAAAACTCCGCTGAA
CGGACGCGAACTCGATACGATACGATACGATATGATGGCAAACCATCGGCAGTGGAGAATGAGT
GCGTGTGTACCTGTGGCGGGCCAGGTAGGTTGCGTTTTCGATTTTCAGACGAAATAATAAAGAA
ATTAAAATTGTTACGGCCTCTCCAGATCGCAGGGCAGGTGAACTTAGCTTCAGTTGAAATCGAA
GTGATTACAAAACTTTTTTGACAGACTGTCCAGAGAGAGAGGGAAGATAATCGGGTATCCATGC
CGGTGGTCCTCTTCTGTTTTGCACCTGGCGCTTCTGGAGTTAACTGTAAACCGTCTGCCTGGCC
GCTCACCTGAGCCGCACTGTGACCTTCAAACCTTTTTTTAACGATTTCGCTTTACGGGCATGCA
AATATGTTTTTTTTACTCCTTTTTTTCGACTCGCTCGATGCAAAAAGTAAAGAGCATAAAATGG
TAATAAATGCGGCAGACGGTTTGCTAGTCACGTTTTCGGGGCGCCATAAATTGGCGGCAGCAAG
GAATTTACATTCAACGGTTTCACATTCGTTTGGGCCTCATTTTTAATTGCCTCGAAATCAATTT
ACTTTTGCT**GTCCAACCA**GCCGAGTGCATG**TGTGTGGTC**CTTAATGGCGCCCAATAGACGGAAA
---A--A-- Ci-180^{ko} ---T-T--- Ci-9^{ko}

AATTAGTACGGAGCGCTAAACGTAATAAAGAATTCCATATCCCTGCTAAAAATAAACAATAAAT
AG**GCCCATCCA**TGACGCACTATATCCATTTCCACCAGCTATTTTGTTTGACAGCTGGGTGAATT
---A-AA-- Ci-140^{ko}

GGGATGTGAAACTAGAAAAAAAACACACTTCCATTCAAATGTCGGGCGGGCGAGATGGGAAAAA
AGGGGAAGGGTAACCACCATGAGATACACGCCAGATCGGATGTGGCAATGCAGAAGATTTCGTA
TTCAAAAACGTTATCAAGTGGCATATAACGTTTATGGTAGCTACAAGTTTTATGGCACACGCTT

ACGTGAGAAACATGGGAAAGCTGCGACACTCGCACACACATAACACACAAAGTCCACCGGATCA
GAGATTCCATATATAGAGCAGTCATCAGCAGTAATCCCTTGAGTCGGAATCGAAAGATATGAAA
AGATTCACGCGCTCGCACAATCCATAAATTTGGACCGCAATGGAATCGCCTGTCCATATACACA
ATGCCATAGATGCTCCCTATTCATATTCGATTCGTC**TGGGTGGTC**TGCCATATAGAGTTTCATT
--T-TT--- Ci-1^{ko}

TTATGATAAAAATGCTGTTTTTTGTGGTTTTGCATTCGATTTTTATTGCGGCATAAC
TGGGTGCGCTCATATTTAATGGCCCAACGCACCATATACGAGCCGAGAGACTCAAGTGCTTTTC
--TT-T--- Ci-83^{ko}
------GT- Ci-1

GGATTTCCAAATGGCATGGAAAAGTTTTCTATATACATATATATATATATATTTTTTTTTTCAA
ATTTAGAGTGGTTCCCGAAGCGGTCCGTGTAATTAAGTAGTCAGGCGCTG**TGGGGGGCC**GGAGA
Ci-23^{ko} --TT-T---
Ci-1 ----T--T-

AAATTCGATTTTTGCCTTTTGCCGCGCCAGTTGATGTGTCGGATCTTTTGATTGATTGATGCGT
GATGCGAAGATTTCGAGATTGTTTACAGAAGAGAACAGAAAACTTCAAACATACAGAATTTTAA
GTAAAGTAGCTCACAATTTTCTTTTCTCTTCATATTCTAGTAGAATCCTTTATGATTTCATAGT
TAATCGTATAGAAGTATGTTAATCTGTTAATATTTAAAATCGAGCTAGAGAGCTGAATTTTTCT
CCGTGTACTTGTGATACATTTCTACCCATCATCTCCGGTCTACCTGCACCTGCTGTTCCCAACT
GTTTTCGTTCTTTTCGGCCACCGGTCTGCCCCTAGCTCGATCTCATTGGTTGTCGTG

HF chr2R:4,533,138-4,535,559 2422 bp

TCCCACTTCATAACCCTCTGAAATCTGGACAAGACACTTCGATTCCAGCTAAAATAAACCTATG
GAGGAAATGTCGCCATGCGTGGGTTCGCTCAGTGATAGCCGAAAATCCATCCATCTGTACCGCA
AAAAATCTATTGAAAACAAAACCCAAATAACCTAGGATTGTACACTGTAAAAAAACGAACATTT
TCTATTGAAAAAGTTTTTGTTTCTATAAATACAACTGATGAAAGGTTTTAAAAATGAGCTAACC
CAATTTTTCTTATGCCAAAATAGTACAATACAAATCGTCTCCGATTGATTTTTTGAAATCGAAG
TATCCCGAATTTGTTCAGCTTTTTGACCATAATTTTTGCCAAGTGCACTGCGAGAAGCAGAGCT
GAAAGCATGGCAG**GACCACACA**AGAAGCTGCAAAAGCCACACACACCGTCGAGGGGACGGGGGC
---A-A--- Ci-9^{ko}

AGACTGCTCAGTTCCCCAGTTCCTCAGTTGGTGATGTCGGTTGCCGGTCGCCGGTCGCCGGTTG
CCGTTGGACGTGGCGGTTGATTACATGTACAAACGAACCGAGCCGATCCGATCCGATCCGAGCC
CAGTTCCAGGCCCAGTCCGATCTGGTCCCCCAGATCGCAGAGCCAACCGCAAAAGCCGACGCCA
AGATAACGAGTTAACCCAAACCCAGACCGCAGGCACCGGGGAATCGGGATGACGGGCCATCGGG
ACAAGGGGAGTCGCCCG**AACCACACA**CGCCATGTCGCTGGGGATTCGGAATCGCGGCCTACGTA
---A-A--- Ci-196^{ko}

TTACCGGTGGCAAATAGTAATTATTACATTAAATGCACCGCAATAATGCAGCCGCGACGTCAAC
ATTTAGCCATAAAAATTGCTTGCACTCCGCTGTTGGAAATGCAGTTAAGCTTAAAGTGGCGGTG
CAAGAGGTACACCGAAAATAATTACAAATACCAAATGTTATGATATTATGATAGATTAAACCTT
TAAATTGATATATTTATGGCATTACTATACATCAAGTAGCGAATCCGAATAGAAATAGAATTGA

CTTCACTGAATCACGTGAAAGATCTCTTTAAATTTTTTTTTTTATGTAAACCTTGTGAATGGGT
TGTGGCGGTGGAAAAGGGGGATCTAGAGAAAACAACCAGCCGGGTAATGGTAATGGCGCTATGG
GCCAATGACAAATGCAATAGCGTTGGTTTGGTAATCGGGGAGCGGGACGAGAGGATTGTGTATC
AGTTGGGGTCTAGGGAACGCTGCACTATAAATTAAGACATTATACGATTAAAAACAAACAGCCT
GTCGTTAAAGTCGCCAAAGGAGGAAAAGAGAGCGGAGAGCGGAGCGAGCAGATCAGGGTTAGAA
TCCAGCTTAAGCAACGGGTTGATGGGCGAGACCAACCGATTGGAATCATTCACATTTTATAGTA
ATGGTAAGGGTATGTTGCTTAATGTTGAGAATTTAGTAAAAATTTTAAATTTATTTAAATGACT
TTTGCCACAACTCTACTCCCTTTTCCACTAAGTCTGCCTCCCGAAAAAGCTAAAAAAAAATAGT
TTTTGAATGCGGGGTTCTTAGGAGCTTTGTAGAGCTCTTAACGCTTATAACATGAAAAATATTT
ATATTTGGTTTATCGCTTTAAACGATCCTGAATTGAGAAATATAGATTGAAACAGAATTCATTA
CCATTTAAGAATATCATTATTTATGGGGGAGTAATGCGCCTCCGAGTAGGCAATGCTTTTCTTG
ACATTGTTACTAAGAATTGTGAATGATATTTGGGCGTGGATCAACGCCGATTAAAAGCTGCTTT
TGCTTCCAGGCGGCCAGAGAAGAGATCCAAACTTCAACTCCAGCCATAAAAGCAACAACATTTC
CGTCTCCCCCTTGTAGCTCCCCTTCCTCCGGCTCTTCCACTCTCCACGAAACGGCAAATGAAGC
TCTCAAAGCGAACTGTGCTTCGCTGGTGGTCCATTGGCAGCTGCCGCCACACAGGCGCTGCTTT
TGTGTGTGTGTGAATATCAATCTTGCTCTCCCTCTCTTTTTATCTCTCTTGGGGAATTGGAGCT
GCATGCGAATTGAGCGACAGCAAAACGAACTGCAAGTCATTGAGAGGAGAGCAAAAACTCGAGA
GCAAGCCAAAGATGGCGCAATCTGGGGAGAGCGAAATAAAGCTAAAATATGCATGTTGGAGAAA
AAATGCCGCCCATGTCGCCAAAATGCGCCACACGCAGAGTGAGCGGGCGGAGGTGGGAGTAATG
GAAAGGGCGATGAGGGAACGATTAGCTTGAAGAGAGAGAACAACAAATGAATGTGCTGCAACGT
TAGTTCAGGTGAGCGAGTTAGAGAGAGAGTTGTTGTTTTTTGATTGTAATAGCTCGC
TTGGTGGTGGGTCCACATTCACATCTCCCTCTCCCACTCTTTCTCCCCGAAAGAGAGAGCG
---T--T------ Ci-202^{ko}/157^{ko} (OVERLAPPING)

YU chr2R:4,520,384-4,522,692 2309 bp

GCCCTGTCGTCTTTGTCTTCTTTATTCCCAGCGCTCC**TGGGCGATC**GA
Ci-106^{ko} --TT-T---

CG**TGTGTGGCC**CTCCACATCCCATTTGCGATCGGAGCGGATTGTGGGGCATTATTGGACCCTGT
---T-T--- Ci-44^{ko}

CTTTGTGCCGCGATTTAAACATACATATTCGGTTTTCTAGGAATCGGGCAATCCTAGTGCGCCA
AACGTTCGGAAAGCTTGGCTTTATTATTATCGTTCCGAGGAAAGCGAAATTAGGTTTTTATTTC
GGGTAGGGCCTGGGAATGCGGATGGAGGTTAAGGTTTGGAGACCAAGCACACACCCAGAAGAAA
AAAACTTTTTTAACATTTTCAATTTTATTTTGCTTATTTTATCTTTGCTCGCAGCCATAACGTC
GAGATCTTCCAGGGTGCTGTATCCGTGGTGAAGGGTTCGGGGGGCA**GAGCTCCCA**CTATTAATA
---A-AA-- Ci-88^{ko}

CTACAATTTTGGAAAAGGTTTATTTTATTATTCTTGATTATTTCCATTGGTATTTAAAAACAAT
ATTCTTTCTGGTACTCACACTAGCTAGATAGTCGAGAAACTTATAGCTCTTACTGTCGTTTTTT
ATTGATAAAAAAAATTCTTGAACTTTTTTTCGCCGACAGCAATGAGACTGCTCCAAAAATGGAT
ATCCCTAATTCTTTTCGAAAAGTATGTATATATTGACTTTTTTCAGTTTCATTGATCGAATTCA
GCCAAAGACTGACGTTTTACCCATTCACGATCTCATCTTCGGTCTTAGACAGCAAGGAATGGGC

TGCGGTTAGTTAACTTAGGGGTAATCATCAGCCAAATGATAAAGGCTAGATCCTCTACTATCAG
ATACCTATTACTCAAATAATTTGTTTTCATACAATAAAGATAGTCTTTTATCCTGTATGTAAAT
CATTTATGATTTATATGATTAAAACCAATTTTCGTCCCGTTAATAAAAAAAAAAATTATTTATA
AAATCTTTAAAAATATAACAGTTTATAACTATTACAGCACCAGGTCAACTTCTTCGATCCCATC
TCGGATTTTTTCAAGCCGTTCTTCTCTGTAATTGGCCCAAAGAACGTAGAATTGCCCAGCGGCA
GCAATAAAATTATGACAAGCAACGGCTGTAACATCAAATGTTGGCAACAATTATTGCGTTGTTG
ACAGCAAGATAGATAGCAATCATTTGCATCGCCATCGCTGGAAAAACCG**AACCACCAA**CAGCAA
Ci-199^{ko} ---A-AA--

TTGTTTGCGGCTGTTGGGCGTGTAATTTACAGCGCCAAGATCAAAAAACAAGATGCATGTTGCA
AGTTGCAGATTGCAGGCGCAACAAAAGGGGAGCAGCACACGAGCAACATGTTACTTTTGACAAT
ATAACGCAACGGCGTGGCGACAACCAACTGGGAGCATCTCGGATCGGCCGTATATATATGTACA
TATCTGTATATATGCACATATGTATGTACATACATATGTATCCCCCAGACATCGGAGGCATTCG
GATTTCACGGTGGTCCCGGCATCTTCATCCCAGGCTTAGTTTCTGCTGTCAATCTTTGGCAAAA
GGCAAATGCGGCTACTGTTTGCCGCACACACAGGCACAGATACAGATACAGATACAGATACAAA
TACAGATACAGATGCAGTCCCATCGTATAATTGCATATAGCGCCAGTGGAAGTTCCGGTCGGTT
GGCTTTTTTTTGCCCGTTGCACGCGCTTTCTTTTTTGTGCGGCGCCGCCGACAGACGGGGCTAC
ACTGAGAGAAATAATTGCATTTTGTTAGTTTTAGGAATCTTACGAATTCGAAATTGTAATTCTA
ATTATCACTTAGATACTACAAGATCATCTATATATCTTTTGAGAATATCATATTTGCTTTCTGA
AAAAAATGATTTTTAAATTTCTTCAAAATTTGTTTTACTTATTTAACTATTCTATATATATGAT
TTGCACAAACTTAAATATAATACACTTTCCGCATACCCTATGATATGACTCTGCTTAAAAATGT
AATGGTACTCAATCAAAATGTAACCATTGGTTTTTCTTTTAAATATCAGCTATAGAAATGTCAT
ATATCCTTAATTAAAACAATATTCCAATAATGTCTTGCTGTATGATATATTTTCCTAATGTAAT
AAATATAGACTTATTATTATGCAAATTTTCGCTGCGTGTTGCTGTGTATGTGTGATTGGTGAAT
GGCATGT**TGCCACCCA**AGTTGCCAGT**TGGCTGGCC**AAGACTTAATAGCCGCTTTCATTGGCCAA
--TT-T--- Ci-93^{ko} --TT-T--- Ci-141^{ko}

AAGCCTGATAAAA**GACCGCACA**CGAAGCTCTTCTTCGGAGCCAAGGCAGTCGTGG
---A-AA-- Ci-59^{ko}

C**TGTGTGGTG**GTTTCGTGGGTTTGATGGTT
---T-T--- Ci-196^{ko}

ZY chr2R:4,519,864-4,520,686 823 bp

CCGGATCGACCTAGGTAAGGGTAAGGATCCTGGTCCAGAATATATATACTGTTCCTGTGACTTA
GCATTGAAAAAATCTTGTATATATAGTATTCCCATTAAAATACCTTAAACCCTCTTCTAATATA
AATCCCATCTTCCATTGCTCAAAATTGATACTTTTTTTTCTCACTGTACCATCTCGCTCACCAA
CACACATGCAGCAACGAGAAAGAGAGCAACGACGTGTCTCTTCG**TGGGTGGCC**CTTAATTCGAC
Ci-3^{ko} --TT-T---

GTCATCTTCT**TGGGTGGTC**CATATTAGCCGGCCTGCTCTCTCCCCCCTTTTGATCTCAAGTGCG
--TT-T--- Ci-1^{ko}

AGCGAGTGCGATGCATGTGGATAGGTTCCCCATTGATCATTTTCTCGTTTGATAGTGCGTGACG
TCTCTGGGCAGCCGTCTCGCTCGCACTAATTTATGCCGCCCGAATGGTGGGGCGCCCATTTTTT

GATCCTTCTTCTGCTGCTCACTCGCGCTCATTTATTCGAGCATAAAATTAGCAATTATTTGTTT
TTCTTGCCGCCCTGTCGTCTTTGTCTTCTTTATTCCCAGCGCTCC**TGGGCGATC**GACG
Ci-106^{ko} --TT-T---

TGTGTGGCCCTCCACATCCCATTTGCGATCGGAGCGGATTGTGGGGCATTATTGGACCCTGTCT
---T-T--- Ci-44^{ko}

TTGTGCCGCGATTTAAACATACATATTCGGTTTTCTAGGAATCGGGCAATCCTAGTGCGCCAAA
CGTTCGGAAAGCTTGGCTTTATTATTATCGTTCCGAGGAAAGCGAAATTAGGTTTTTATTTCGG
GTAGGGCCTGGGAATGCGGATGGAGGTTAAGGTTTGGAGACCAAGCACACACCCAGAAGAA

GB chr2R;4,534:284-4,537,090 2807 bp

CTATGGGCCAATGACAAATGCAATAGCGTTGGTTTGGTAATCGGGGAGCGGGACGAGAGGATTG
TGTATCAGTTGGGGTCTAGGGAACGCTGCACTATAAATTAAGACATTATACGATTAAAAACAAA
CAGCCTGTCGTTAAAGTCGCCAAAGGAGGAAAAGAGAGCGGAGAGCGGAGCGAGCAGATCAGGG
TTAGAATCCAGCTTAAGCAACGGGTTGATGGGCGAGACCAACCGATTGGAATCATTCACATTTT
ATAGTAATGGTAAGGGTATGTTGCTTAATGTTGAGAATTTAGTAAAAATTTTAAATTTATTTAA
ATGACTTTTGCCACAACTCTACTCCCTTTTCCACTAAGTCTGCCTCCCGAAAAAGCTAAAAAAA
AATAGTTTTTGAATGCGGGGTTCTTAGGAGCTTTGTAGAGCTCTTAACGCTTATAACATGAAAA
ATATTTATATTTGGTTTATCGCTTTAAACGATCCTGAATTGAGAAATATAGATTGAAACAGAAT
TCATTACCATTTAAGAATATCATTATTTATGGGGGAGTAATGCGCCTCCGAGTAGGCAATGCTT
TTCTTGACATTGTTACTAAGAATTGTGAATGATATTTGGGCGTGGATCAACGCCGATTAAAAGC
TGCTTTTGCTTCCAGGCGGCCAGAGAAGAGATCCAAACTTCAACTCCAGCCATAAAAGCAACAA
CATTTCCGTCTCCCCCTTGTAGCTCCCCTTCCTCCGGCTCTTCCACTCTCCACGAAACGGCAAA
TGAAGCTCTCAAAGCGAACTGTGCTTCGCTGGTGGTCCATTGGCAGCTGCCGCCACACAGGCGC
TGCTTTTGTGTGTGTGTGAATATCAATCTTGCTCTCCCTCTCTTTTTATCTCTCTTGGGGAATT
GGAGCTGCATGCGAATTGAGCGACAGCAAAACGAACTGCAAGTCATTGAGAGGAGAGCAAAAAC
TCGAGAGCAAGCCAAAGATGGCGCAATCTGGGGAGAGCGAAATAAAGCTAAAATATGCATGTTG
GAGAAAAAATGCCGCCCATGTCGCCAAAATGCGCCACACGCAGAGTGAGCGGGCGGAGGTGGGA
GTAATGGAAAGGGCGATGAGGGAACGATTAGCTTGAAGAGAGAGAACAACAAATGAATGTGCTG
CAACGTTAGTTCAGGTGAGCGAGTTAGAGAGAGAGTTGTTGTTTTTTGATTGTAATAGCTCGC
TTGGTGGTGGGTCCACATTCACATCTCCCTCTCCCACTCTTTCTCCCCGAAAGAGAGAGCGGGA
Ci-200 Ci-160

GCGAAGGGGCACGAGGGGAGCACGATGACTATGCAGTTGCATTCAATTTGAATTTCCATGGTGC
TGATGATTCGAGCGCCAATTTTTTCGAAGAGTTCTTATTTGTTTACTTCGTTGTTGTTGCCTCA
ATTGGAAAGGGAAAATGTGGAATGCGGAGAAACACCAGAAGCAAATGCATTTCCATTCATAAAT
CCAAAGAAGTTTTAAAGATAACATGTCATTTGGCTTAAGTTCGTGGTGCACAAAAAAGATCGGT
TTGCGGTTGTCGCATGAAAATGAGTTTATTCCATTGGTATATTATTATTCAGAAATTAAAAAAA
AACTTGTTTAGTCTATTTTTTTTTTTTAAATAAAAAAAAAAAAATTCTTTTATAAGTCGATTTTA
GAGTAAATATTTAAAGACTACGTCTAATAAACATATAATTTGTTCTGTGTTTTAATTTGCCGGC
AAAAACAAACCTACT**TGTGTGGTC**CTCGCACACTCATAACCCCTCGCATATTTGAGATTCATGG
Ci-9
GGCAAAGAGGCTGCAAAAACAATGGAAAGGGAAAAGCAGAAACATCCTGCCGCTCATAATTTAG
CATCGGAACATGCAAAAACAGACATCATCGCATGGGGCAGCAGCAACAGCCATAAAACCAACAA

CACGAGCAATGTAAAGCTAACAAATTTGCCAACAGTTCGCGGCACGGCTACACACACACACATG
CATGCGCAGCCTGCCACGCACGCGCTTCCCCCAAACAAATACACACACACACACTGAGACGAAA
GCTCCATTGGGCAGCGCTGCCGACGCTGAAGGCCGACATCGGCAGAGCTGAACGTTTGGGTAGG
G**GACCACCCA**CATCGCTTGGCGGTTTCAGTTTAATGAAGGCAGAAACAAATTTATTTT
---A-AA-- Ci-1^{ko}

TGGGTGGTCCACACTGCAGCGAAAATAAACTACAGTGGCAACAACAAACCAGCAGCCAAGGCAC
---T-TT-- Ci-1^{ko}

TT**TGGGTGGTC**CATGCAAAAAAAAAACAAATTACGGCATGCGAATAACAATAGAAATTAGCGCT
---T-TT-- Ci-1^{ko}

CTCGTGGCGGAGCTATTTGGGTATATTAGAGCTACATATTTTATTTGTTTATAAAAAGTATAAA
TGTAAACAATGAGTTCCAAGCATTAAGTCCGTATGCTCAACAATTACATTATCATTATTATTAT
CACTTAAATATTTACAAAGGATATTTAAACAGTAATAGATATATATTTTATTTCTTAATTTCTG
TTAACATATGTATTTACATTGGTAGTTATTCTTTATTTTGCAACAAGCATTCATAAATTTTATA
TAACAAACTTGGTATTTTCTCGGAAAAACTCCTGAATCACCCCTCGGTATTTTGTGCGTTGAGC
TATCGTTAAAGCAGCCCTCGCAGAGAGCGTTCTCAAACCAAAATGGCCGCACACGAAACAAGAG
AGCGAGTGAGAGTAGGGAGAGCGTCTGTGTTGTGTGTTGAGTGTC**GCCCACGCA**CACAGGCGCA
Ci-63
AAACAGTGCACACAGACGCCCGCTGGGCAAGAGAGAGTGAGAGAGAGAAACAGCGGCGCGCG

JB chr2R:4,531,613-4,537,090 5478 bp

TACGTACTCTTATTACTCCACTCCACTTCACTTCGCGGCAGTGTAAATAAAGACGGGTCATGGG
ATGGGGCACGAAGTGGGGTTAAATGGCCAGGGGCATTGGGGAGTAACGAGGCATACCTGCGGGC
CGTAAATAAGCAACTCAATGAGCAGCGAACGCCAGCGCACGTGCCTCATAAACCACCGGCCGAG
TCACCAGTATCCATACACTCACAGAAAAATGGTGGAACAAATTCGATTTAAAATTTATATTGAC
ATTTATTTAGCACTTCAAGCTACAATAGAAAGACGTGACACATTAAAATATTTCATGTATCTTT
CTTCAAATAAATAATAGCTACAAATTTTTGCCTGTGTACTTAAGTCGTATTCGAGCCGAGTTGG
AGTCAAGTCGAATCGAATCGAGTTGGA**GACCACCCA**GGTAACGATTTGTGCAGTCATTCCACGA
Ci-1
ATTCCCAGCTGAGGGGGCGTCGCGTCGGTTCTTCGTGGCTTTTGCGTCCAATTCTGCAGACGGC
GGGCGCTGGGGGGCATTTCGCATTTCACCAGGCTAAGTATGCTAAGGATGGCAAACGGTATCCG
TCCGGGACCATTGTCAAAGTCCAAATTATGCATGCGTTTTCAGAGCTGGGTATACATACATATA
TAGCCGGACAGCTGCGGGAGTGAAAATAAAACCACAGGAATTAATTTCGTTCAAAATAAATAGA
TGTACTTGCTGTTTTGAACTTTCTAATCATCTACCTACAGACTTTAAGATTTGGAATTTCCATA
ACAATGTATTCCCTGTTACGCATTGTAAGTTTCTCGCTGGACACAGTGCAATATTCTTATGACC
AGCACTGTACTTATGTAAATTCCGCCGGTGGTTGTGGTTGTGGTGTGCAGGGCAATAAGGTTCG
GCGCATAAC**GCCCACGCA**GTCGCAGTATCGTCATGGTCGCCGTAATCGCAGTTGGGGCAGCGGG
Ci-63
CGAAATTAATTTGACAGCCAACGCAAAACGCATGCGTCGTCGACATTCCTCGCCTTCA
CCCCACCCACCCACTGATCGCCGTCTCTTGCACCACCCTGCTCCCCGGATATATCCACGATCTC
Ci-104 Ci-102
CGGGGATTCGACGACGACGACAGGCAGACCCAGTGACGTTCCGGGCTTATCGCCTTACTCTTAT
ATCTATCTTAAGCTTGTTGTTGGCCCGCCTCCTTCCCGCTCATTGTCGTTATCGCCTTGCAGTC

ACATTAACACATCCGACTTCAAAAAGGGTTATGCGATGAGCTTAACTGATGTCGACGGAGCGGG
AACTACGGGCAAATTAATGGGTTGGGGATTTGCAGCCAGCGTTGCAAAGGGAACGGAATCGAAG
AGAACGTAGCAATGTCATCTTAATTTTGGTTCTCAATATGATCTCTTTAGTGTAGAGCAGAAAC
CCAGATTACATTGCCCATCAATCAATCTTCAAAATGCCTTCAAAAAAGTGAACTGCAATAGAAT
ATATCTGGAATAAACACCTTGGACTTACTGAATCGCTTCTCATTCTACTTCCAATTAATTCCCA
CTTCATAACCCTCTGAAATCTGGACAAGACACTTCGATTCCAGCTAAAATAAACCTATGGAGGA
AATGTCGCCATGCGTGGGTTCGCTCAGTGATAGCCGAAAATCCATCCATCTGTACCGCAAAAAA
TCTATTGAAAACAAAACCCAAATAACCTAGGATTGTACACTGTAAAAAAACGAACATTTTCTAT
TGAAAAAGTTTTTGTTTCTATAAATACAACTGATGAAAGGTTTTAAAAATGAGCTAACCCAATT
TTTCTTATGCCAAAATAGTACAATACAAATCGTCTCCGATTGATTTTTTGAAATCGAAGTATCC
CGAATTTGTTCAGCTTTTTGACCATAATTTTTGCCAAGTGCACTGCGAGAAGCAGAGCTGAAAG
CATGGCAG**GACCACACA**AGAAGCTGCAAAAGCCACACACACCGTCGAGGGGACGGGGGCAGACT
Ci-9
GCTCAGTTCCCCAGTTCCTCAGTTGGTGATGTCGGTTGCCGGTCGCCGGTCGCCGGTTGCCGTT
GGACGTGGCGGTTGATTACATGTACAAACGAACCGAGCCGATCCGATCCGATCCGAGCCCAGTT
CCAGGCCCAGTCCGATCTGGTCCCCCAGATCGCAGAGCCAACCGCAAAAGCCGACGCCAAGATA
ACGAGTTAACCCAAACCCAGACCGCAGGCACCGGGGAATCGGGATGACGGGCCATCGGGACAAG
GGGAGTCGCCCG**AACCACACA**CGCCATGTCGCTGGGGATTCGGAATCGCGGCCTACGTATTACC
Ci-196
GGTGGCAAATAGTAATTATTACATTAAATGCACCGCAATAATGCAGCCGCGACGTCAACATTTA
GCCATAAAAATTGCTTGCACTCCGCTGTTGGAAATGCAGTTAAGCTTAAAGTGGCGGTGCAAGA
GGTACACCGAAAATAATTACAAATACCAAATGTTATGATATTATGATAGATTAAACCTTTAAAT
TGATATATTTATGGCATTACTATACATCAAGTAGCGAATCCGAATAGAAATAGAATTGACTTCA
CTGAATCACGTGAAAGATCTCTTTAAATTTTTTTTTTTTATGTAAACCTTGTGAATGGGTTGTGG
CGGTGGAAAAGGGGGATCTAGAGAAAACAACCAGCCGGGTAATGGTAATGGCGCTATGGGCCAA
TGACAAATGCAATAGCGTTGGTTTGGTAATCGGGGAGCGGGACGAGAGGATTGTGTATCAGTTG
GGGTCTAGGGAACGCTGCACTATAAATTAAGACATTATACGATTAAAAACAAACAGCCTGTCGT
TAAAGTCGCCAAAGGAGGAAAAGAGAGCGGAGAGCGGAGCGAGCAGATCAGGGTTAGAATCCAG
CTTAAGCAACGGGTTGATGGGCGAGACCAACCGATTGGAATCATTCACATTTTATAGTAATGGT
AAGGGTATGTTGCTTAATGTTGAGAATTTAGTAAAAATTTTAAATTTATTTAAATGACTTTTGC
CACAACTCTACTCCCTTTTCCACTAAGTCTGCCTCCCGAAAAAGCTAAAAAAAAAATAGTTTTTG
AATGCGGGGTTCTTAGGAGCTTTGTAGAGCTCTTAACGCTTATAACATGAAAAATATTTATATT
TGGTTTATCGCTTTAAACGATCCTGAATTGAGAAATATAGATTGAAACAGAATTCATTACCATT
TAAGAATATCATTATTTATGGGGGAGTAATGCGCCTCCGAGTAGGCAATGCTTTTCTTGACATT
GTTACTAAGAATTGTGAATGATATTTGGGCGTGGATCAACGCCGATTAAAAGCTGCTTTTGCTT
CCAGGCGGCCAGAGAAGAGATCCAAACTTCAACTCCAGCCATAAAAGCAACAACATTTCCGTCT
CCCCCTTGTAGCTCCCCTTCCTCCGGCTCTTCCACTCTCCACGAAACGGCAAATGAAGCTCTCA
AAGCGAACTGTGCTTCGCTGGTGGTCCATTGGCAGCTGCCGCCACACAGGCGCTGCTTTTGTGT
GTGTGTGAATATCAATCTTGCTCTCCCTCTCTTTTTATCTCTCTTGGGGAATTGGAGCTGCATG
CGAATTGAGCGACAGCAAAACGAACTGCAAGTCATTGAGAGGAGAGCAAAAACTCGAGAGCAAG
CCAAAGATGGCGCAATCTGGGGAGAGCGAAATAAAGCTAAAATATGCATGTTGGAGAAAAAATG
CCGCCCATGTCGCCAAAATGCGCCACACGCAGAGTGAGCGGGCGGAGGTGGGAGTAATGGAAAG
GGCGATGAGGGAACGATTAGCTTGAAGAGAGAGAACAACAAATGAATGTGCTGCAACGTTAGTT
CAGGTGAGCGAGTTAGAGAGAGAGTTGTTGTTTTTTGATTGTAATAGCTCGC
TTGGTGGTGGGTCCACATTCACATCTCCCTCTCCCACTCTTTCTCCCCGAAAGAGAGAGCGGGA
Ci-200 Ci-161
GCGAAGGGGCACGAGGGGAGCACGATGACTATGCAGTTGCATTCAATTTGAATTTCCATGGTGC
TGATGATTCGAGCGCCAATTTTTTCGAAGAGTTCTTATTTGTTTACTTCGTTGTTGTTGCCTCA

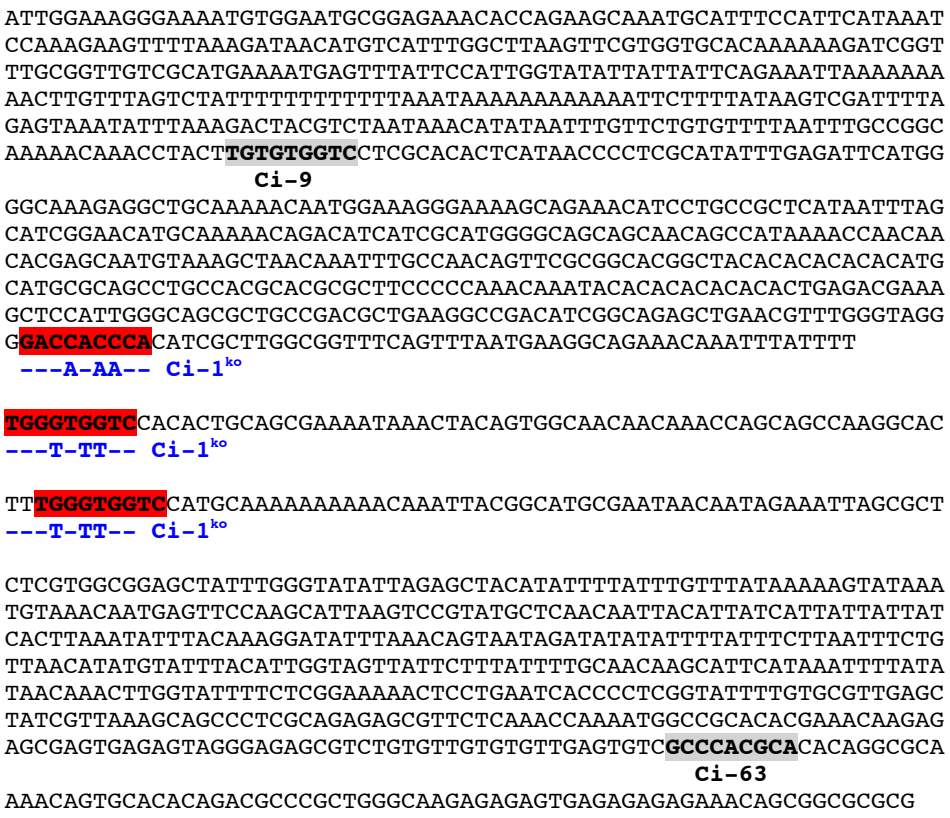


Figure 2.2 *patched* enhancer sequences with targeted GLI/Ci binding site mutations

Sequences were taken from the dm3 build of the UCSC genome browser. Optimal Ci binding sites are highlighted in red, low affinity Ci binding sites are highlighted in gray and further annotated with a numerical ranking corresponding to the affinity predictions previously defined(Hallikas et al., 2006). Blue text indicates how each Ci binding site was mutated, for example in *ptc*-prox enhancer DB, the optimal Ci-1 binding site of GACCACCCA was mutated to GACaCaaCA

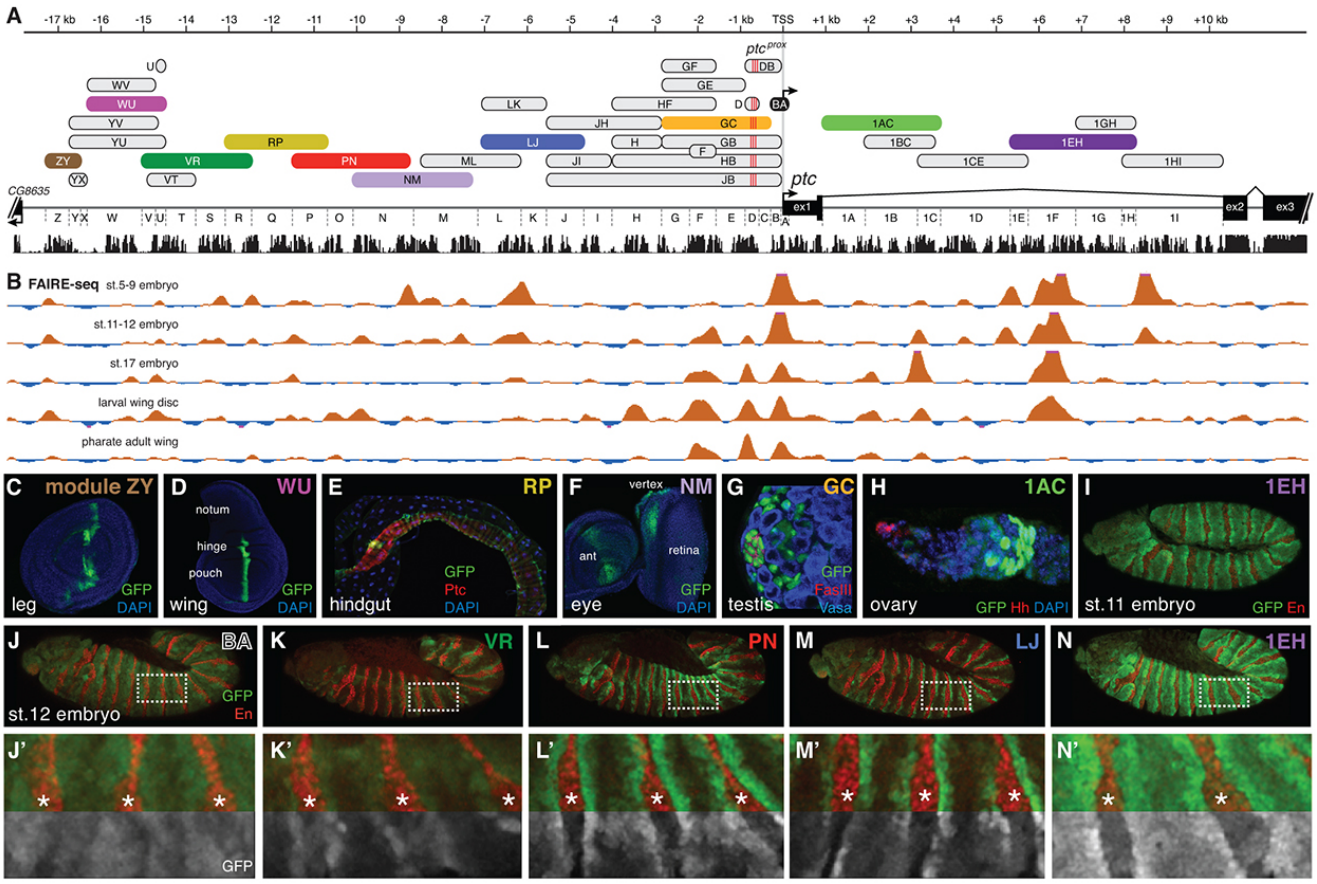


Figure 2.3 The universal Hedgehog response of *patched* is mediated by a large array of tissue- and stage-specific enhancers

(A) The *ptc* genomic locus. DNA regions tested for in vivo enhancer activity are shown as oblong shapes; a sequence conservation histogram is shown below the exon map. (B) In vivo DNA accessibility at the *ptc* locus, as determined by FAIRE-seq across various developmental stages and tissues (McKay & Lieb, 2013); DNAse-seq data are shown in Figure 2—figure supplement 1. (C-I) Selected larval, adult, and embryonic tissues from GFP transgenic animals. Enhancer names are as in (A). (J) Activity of the *ptc*BA promoter region in a stage 12 embryo, driving broad, low expression in the anterior (En-negative) compartment of each segment. (K-N) *ptc*-like ectodermal stripes driven by four separate *ptc* enhancers in *cis* to the *ptc*BA promoter. (K'-N') Enlarged views of dashed boxes in (J-N); asterisks show stripes of En-positive, Hh-producing cells in the posterior compartment of each segment. GFP signals are isolated (greyscale) in the lower half of K'-N' to show the lack of stripe enhancer activity in the posterior (En-positive) compartment.

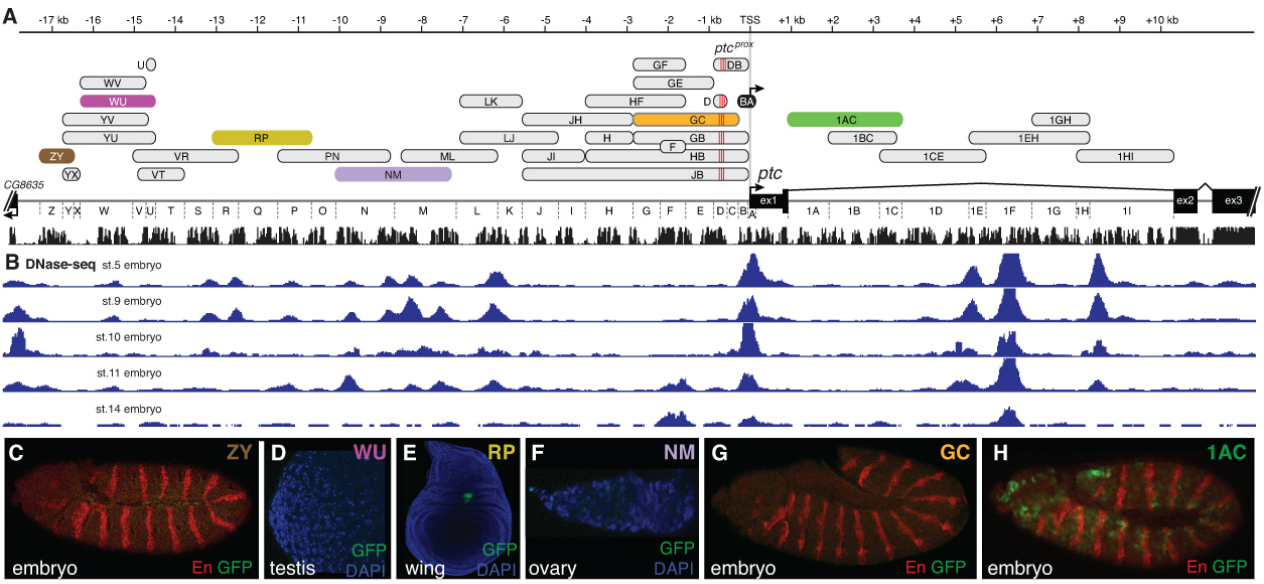


Figure 2.4 Individual *patched* enhancers exhibit tissue-restricted responses to Hedgehog signaling

(A) Map of the *ptc* locus. Enhancers described in this figure are colored; all other tested enhancers are grey. Sequence conservation among 12 *drosophilids* is shown in black. (B) DNAse-seq different embryonic stages in the *ptc* locus (Thomas et al., 2011) (B to H) enhancer reporter constructs, all of which are Hh/GLI-responsive in other contexts (see Figure 2.3), shown here in a developmental context in which they fail to drive a *ptc*-like response to Hh/GLI.

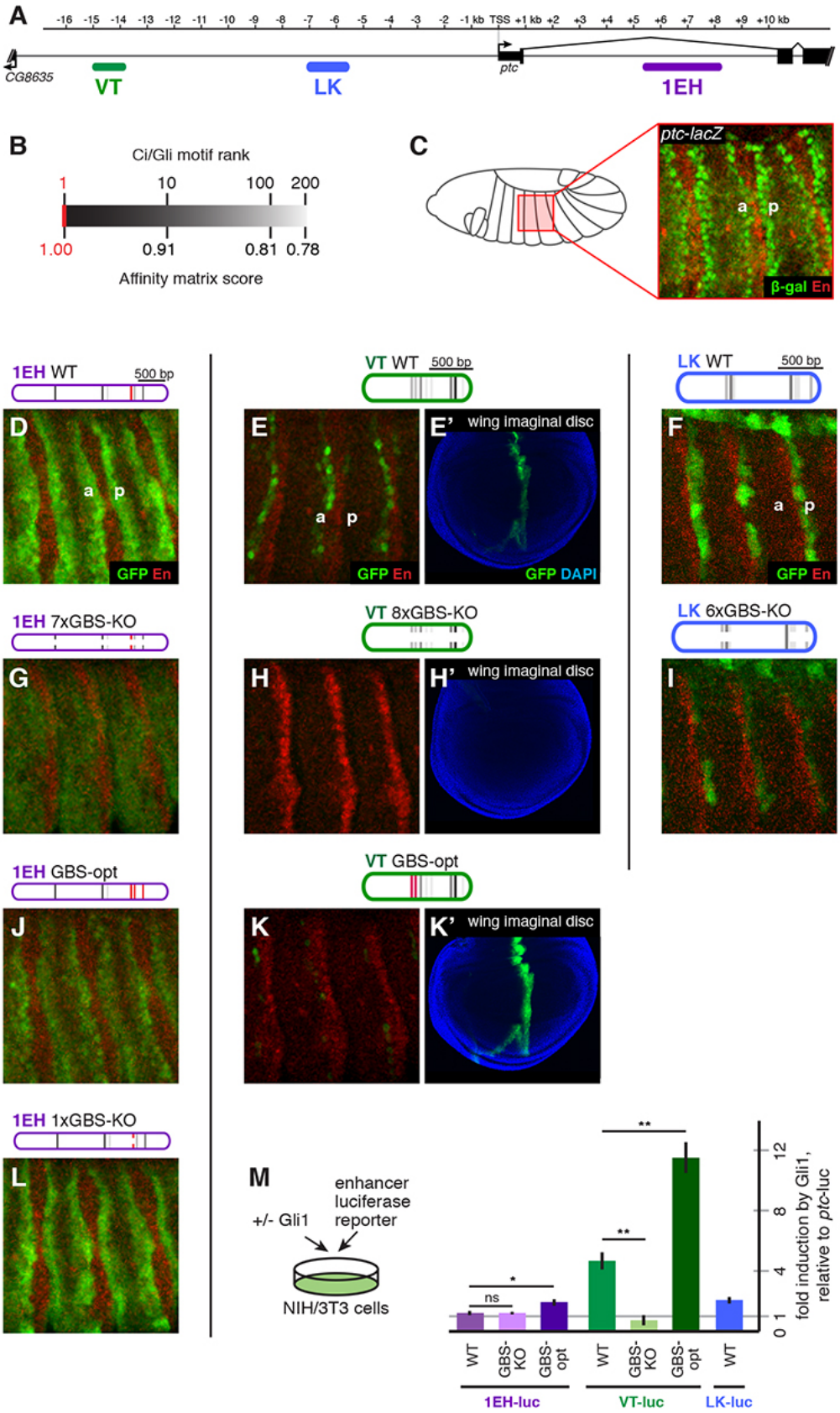

A
-16 -15 -14 -13 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 kb TSS +1 kb +2 +3 +4 +5 +6 +7 +8 +9 +10 kb
CG8635
VT
LK
ptc
1EH
B
Ci/Gli motif rank
1
10
100
200
1.00
0.91
0.81
0.78
Affinity matrix score
C
ptc-lacZ
a
p
β-gal En
1EH WT
500 bp
D
a
p
GFP En
VT WT
500 bp
E
a
p
GFP En
E'
wing imaginal disc
GFP DAPI
LK WT
500 bp
F
a
p
GFP En
1EH 7xGBS-KO
G
VT 8xGBS-KO
H
H'
wing imaginal disc
LK 6xGBS-KO
I
1EH GBS-opt
J
VT GBS-opt
K
K'
wing imaginal disc
1EH 1xGBS-KO
L
M
enhancer
luciferase
reporter
+/- Gli1
NIH/3T3 cells
**
**
*
ns
0
1
4
8
12
fold induction by Gli1,
relative to ptc-luc
WT
GBS-KO
GBS-opt
WT
GBS-KO
GBS-opt
WT
1EH-luc
VT-luc
LK-luc

Figure 2.5 Embryonic ectoderm enhancers of *patched* require suboptimal GLI sites to respond to Hh signaling

(A) Map of the *ptc* locus showing enhancers tested in this figure. (B) Scale of predicted relative binding affinity for Ci, used to label predicted Ci/GLI motifs in enhancer diagrams; the optimal motif (in red) is GACCACCCA. (C) Cartoon depiction of the embryo with the representative region in this figure highlighted in red, left. Stage 13 embryo carrying the *ptc-lacZ* enhancer-trap, right. (D,H,K) Expression of wild-type (WT) 1EH, VT and LK enhancers in embryos (wing imaginal disc displayed in H') Full embryos for all panels can be seen in Figure 2.6. (E,I,L) GLI binding site knockout (GBS-KO) versions of those enhancers, (wing imaginal disc displayed in I'). In addition, enhancer LK also responds to Hh signaling via overexpression (Figure 2.7) (F,J) Optimization of two non-consensus GLI mofits (GBS-opt) in enhancers 1EH and VT (wing imaginal disc displayed in J') (G) The single optimal consensus motif is mutated in the 1EH module (1xGBS-KO). The relative affinity of GBS were assessed by competitive EMSA analysis (Figure 2.8). All enhancer and GLI sites can be found in Figure 2.2. (M) GLI1-responsiveness of embryonic enhancers in NIH/3T3 cells, error bars indicate s.d. Student's t-test; *, $p<0.05$; **, $p<0.005$.

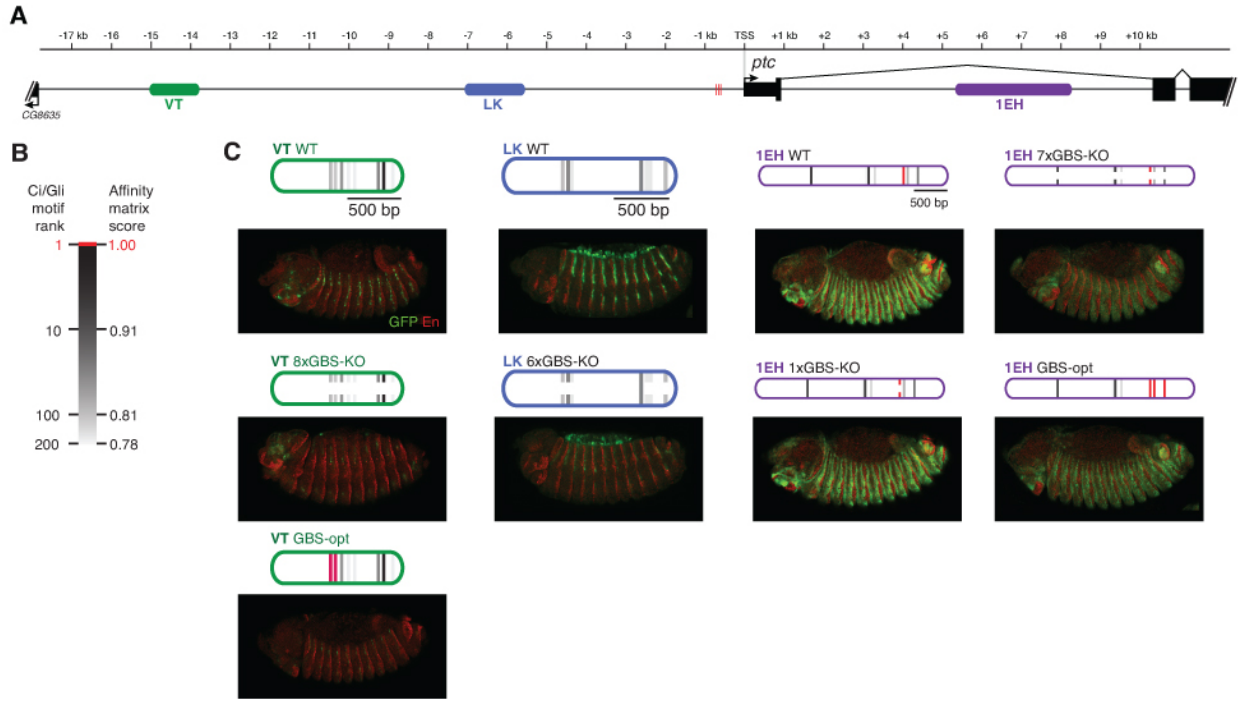


Figure 2.6 Embryonic ectoderm enhancers of *patched* require non-consensus, suboptimal GLI sites to respond to Hh signaling

These are whole-embryo views of the enhancer expression patterns presented in Figure 2.5.

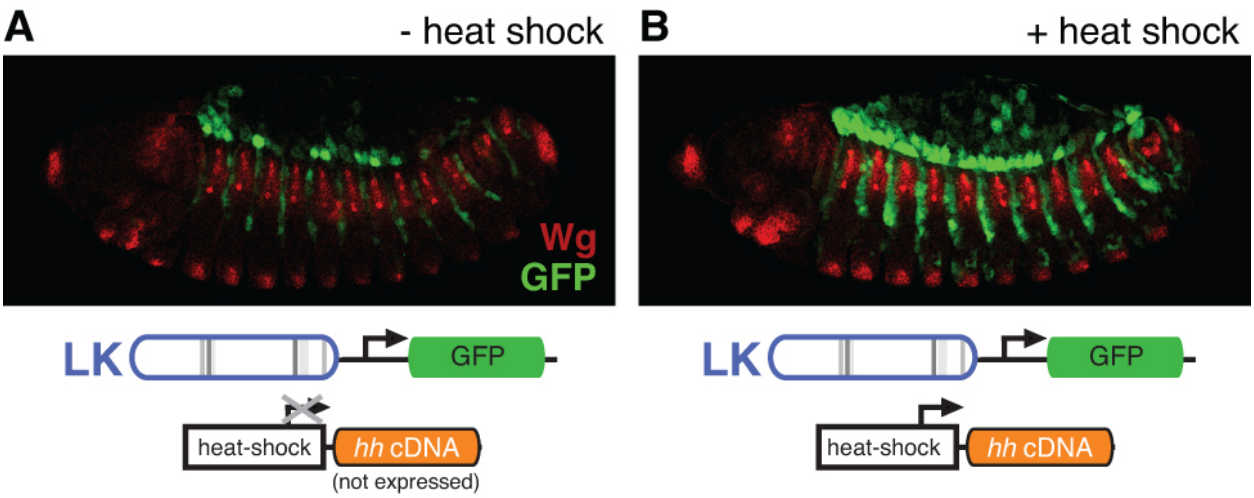


Figure 2.7 *patched* enhancer LK responds to Hh signaling in the embryonic ectoderm

(A) GFP expression (green) shows expression of a wild-type LK enhancer reporter in doubly transgenic embryos containing a hedgehog cDNA under the control of the heat-inducible Hsp70 promoter, in the absence of heat shock. (B) Embryo of the same genotype as (A), but with heat shock. Hh pathway activation causes an intensification and expansion of enhancer activity in segment-polarity stripes to the posterior of stripes of normal Hh signaling, but does not produce ectopic expression in the anterior stripes of Hh-responsive, *ptc*-positive cells (labeled in red with the Hh/GLI target Wingless, which also maintains its restricted expression pattern).

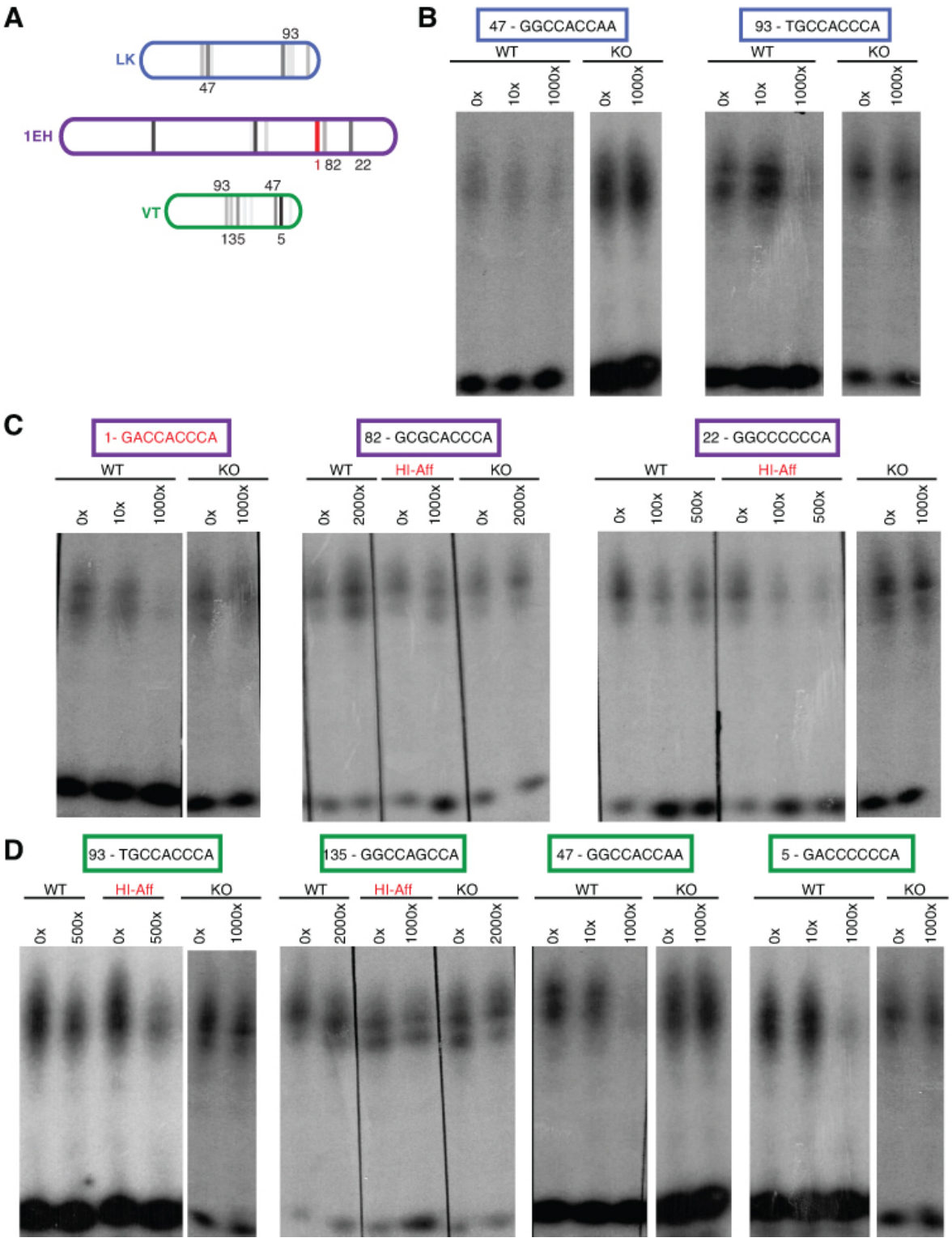


Figure 2.8 Ci binds to non-consensus GLI motifs in *patched* embryonic ectoderm enhancers in vitro

(A) Diagram showing three embryonic-ectoderm enhancers of *ptc* and their predicted GBS (using the same affinity rank scale as in Figure 2.5). GBS noted with an affinity score were tested in these experiments. (B-D) EMSA competition assays in which cold (unlabeled) probes bearing GBS of noted rank compete with a radiolabeled probe containing the consensus GBS (GACCACCCA, rank 1) for binding to the Ci DNA-binding domain. GBS sequences and affinity ranks are in colored boxes and the concentrations of competitor oligonucleotides, relative to the concentration of radiolabeled probe, are noted.

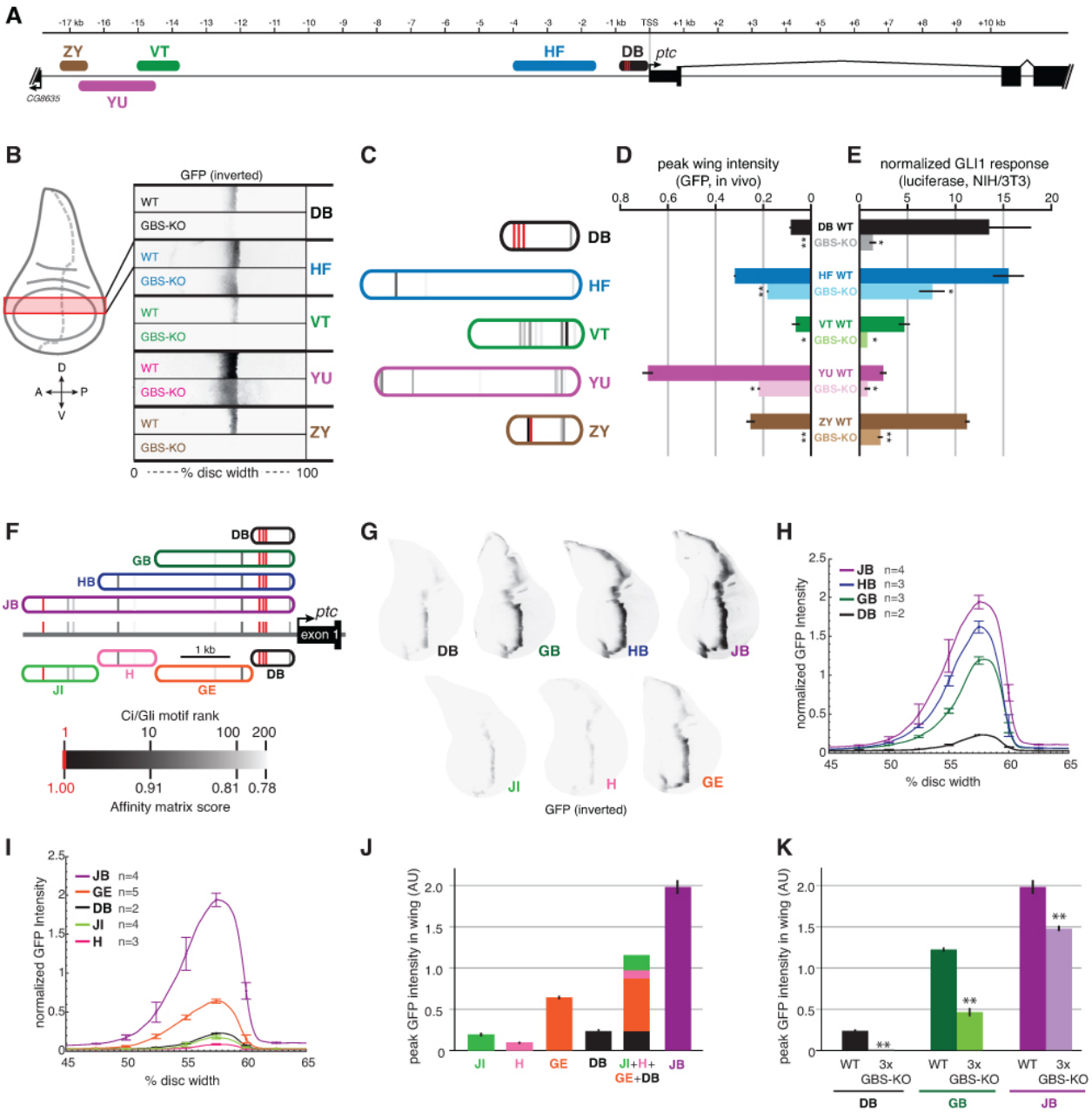


Figure 2.9 Wing enhancers of *patched* synergize to respond to Hh/GLI, largely via low-affinity GBS

(A) Map of the *ptc* locus showing enhancers tested in this figure. (B) Activity of WT and GBS-KO versions of five *ptc* enhancers across the wing imaginal disc. Color-inverted grayscale images of disc segments spanning the anterior-posterior axis; GFP fluorescence appears black. The relative affinity of GBS were assessed by competitive EMSA analysis (Figure 2.10). (C) Diagrams of five wing enhancers and their GBS; all enhancers are color-coded as in (A). (D) Peak wing disc GFP intensity driven by WT and GBS-KO wing enhancers. (E) GLI1-responsiveness of the same enhancers in

NIH/3T3 cells, error bars indicate s.d. Student's t-test; *, $p<0.05$; **, $p<0.005$. (F) Tested enhancers ≤6 kb upstream of the *patched* promoter are mapped to the *ptc* locus with best predicted high- and low- affinity GBS in red and gray, see GBS affinity color scale, bottom. (G) promoter proximal enhancer activity in wing imaginal discs as revealed by transgenic larvae (inverted grayscale represents GFP expression). (H,I) Quantitation of GFP enhancer activity across the wing disc. (J-K) Quantitation of peak GFP fluorescence in transgenic wing discs carrying enhancer-GFP reporters. All enhancer and GLI binding sites can be found in Figure 2.2.

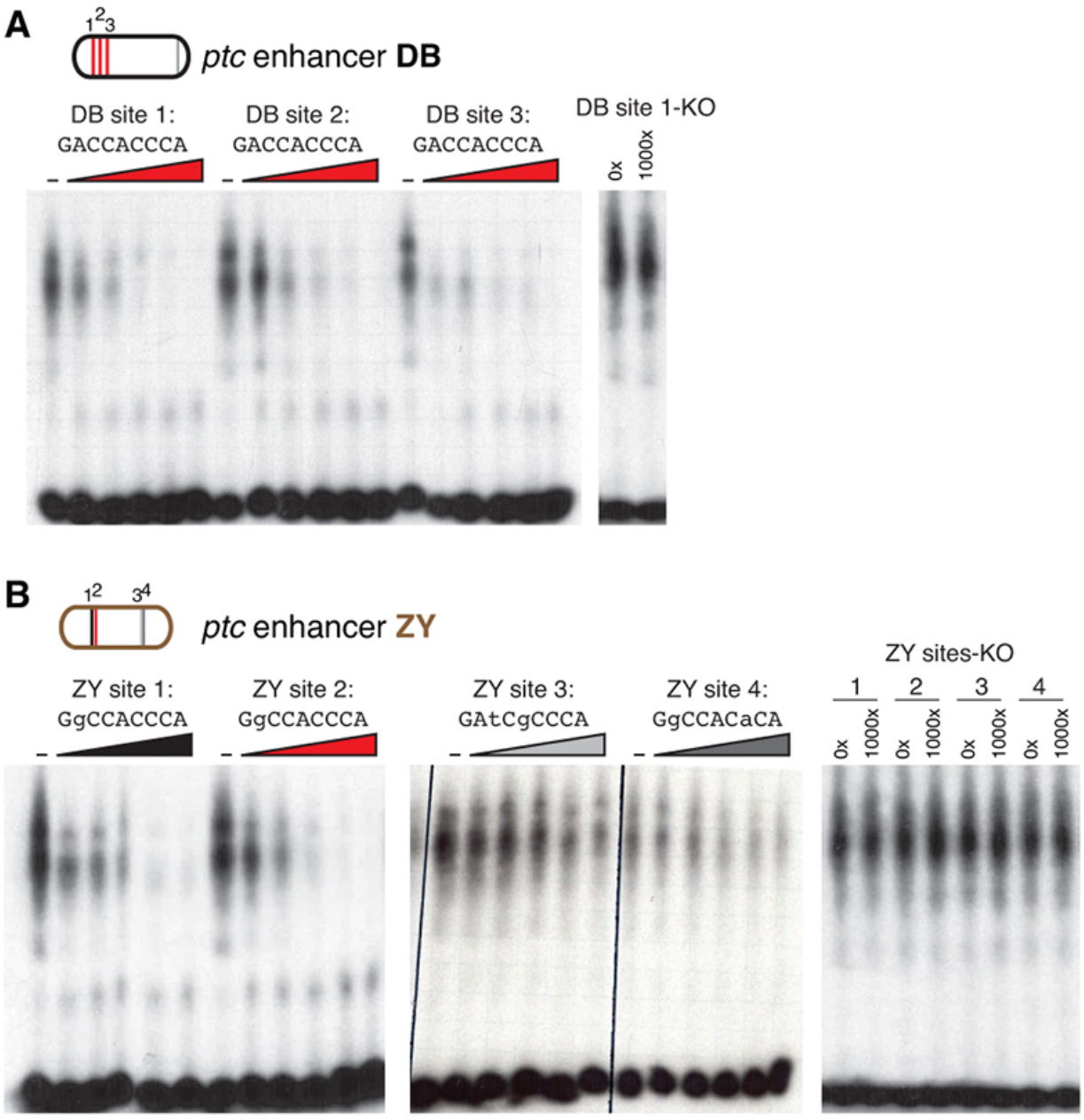


Figure 2.10 GLI binds to low affinity motifs in wing enhancers

(A) EMSA competition assays in which increasing concentrations of unlabeled oligonucleotides containing the optimal GBS from the *ptc*DB enhancer compete with a radiolabeled probe containing the consensus GBS (specifically, DB site 1) for binding to the Ci DNA-binding domain. Unlabeled competitor concentrations, relative to labeled probe, are 0x, 50x, 100x, 250x, 500x, and 1000x, from left to right. Far right, an unlabeled probe containing the mutated motif used in GBS-KO enhancers does not compete with the labeled probe for Ci binding. (B) Four GBS from the *ptc*ZY enhancer (colored according to the affinity scale in Fig. 2), tested as unlabeled competitors against a labeled probe containing an optimal GBS (DB site 1, tested in the far left of

panel [A]). Oligonucleotide concentrations are the same as in (A). Far right, GBS-KO mutated versions of the four ZY GBS examined to exhibit Ci binding in vitro.

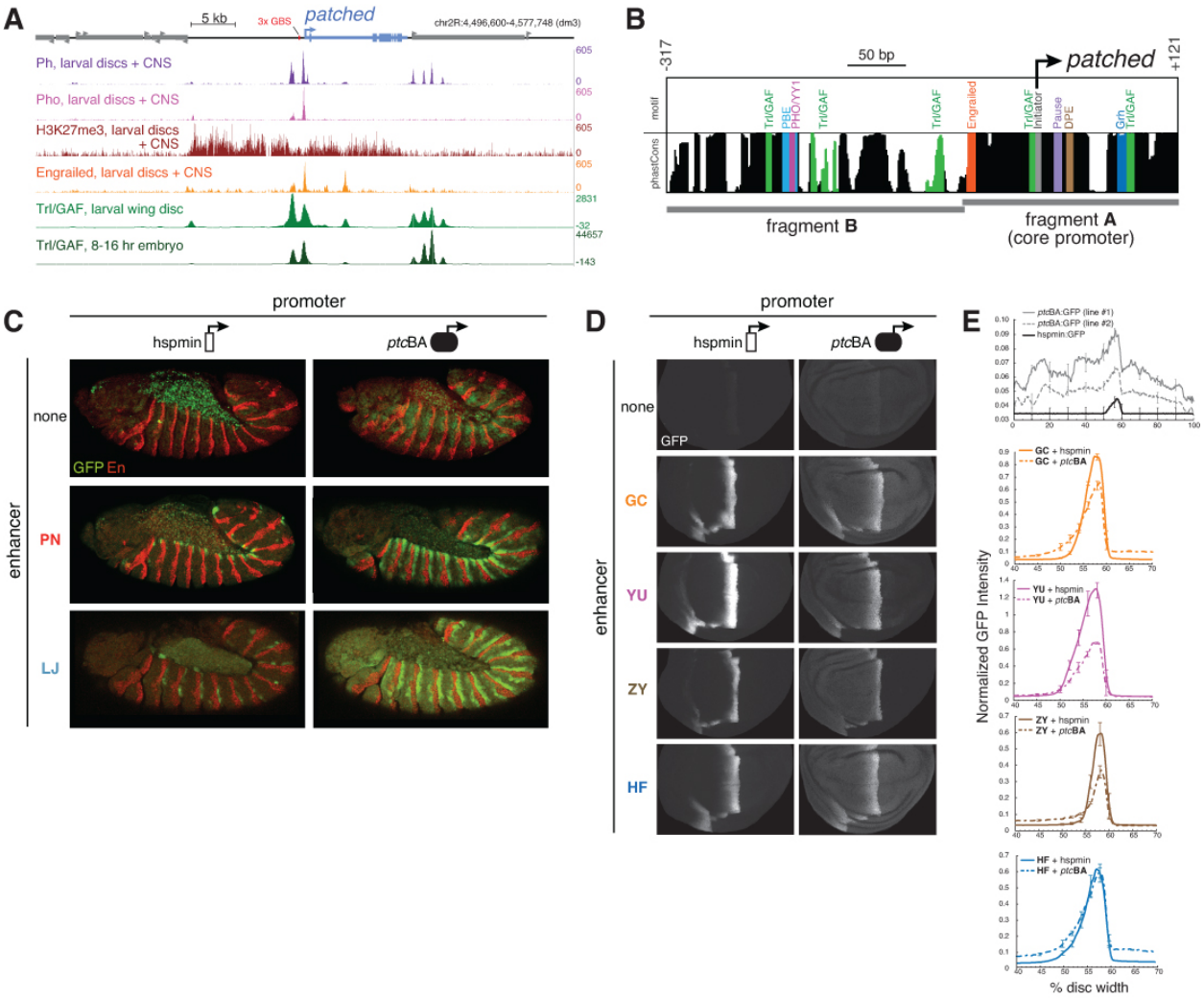


Figure 2.11 Developmentally dynamic Polycomb/Trithorax Response Elements (PREs) near the *patched* promoter modulate enhancer activity differently in the embryo and developing wing

(A) ChIP-seq data from larval tissues showing binding of PcG proteins, the H3K27me3 histone mark, Engrailed, and Trithorax-like(Trl)/GAF in the neighborhood of *ptc*. Embryonic Trl/GAF binding data (bottom) are taken from (Oh et al., 2013). For comparison with larval PcG/TrxG binding, see embryonic ChIP-seq data in Figure 2.12. (B) DNA motif analysis of the *ptc*BA promoter. Additional 5' flanking sequence is analyzed in Figure 2.13. (C) Two embryonic ectodermal enhancers of *ptc* (rows) tested on different promoters (columns) in stage 12 transgenic embryos. Related experiments are presented in Figure 2.14. (D) Four wing enhancers of *ptc* tested on a minimal TATA+Inr promoter (hspmin) or the native *ptc* promoter (fragment BA; -317 to +121). (E) Quantitation of wing GFP expression across wing imaginal discs carrying the transgenes shown in panel D.

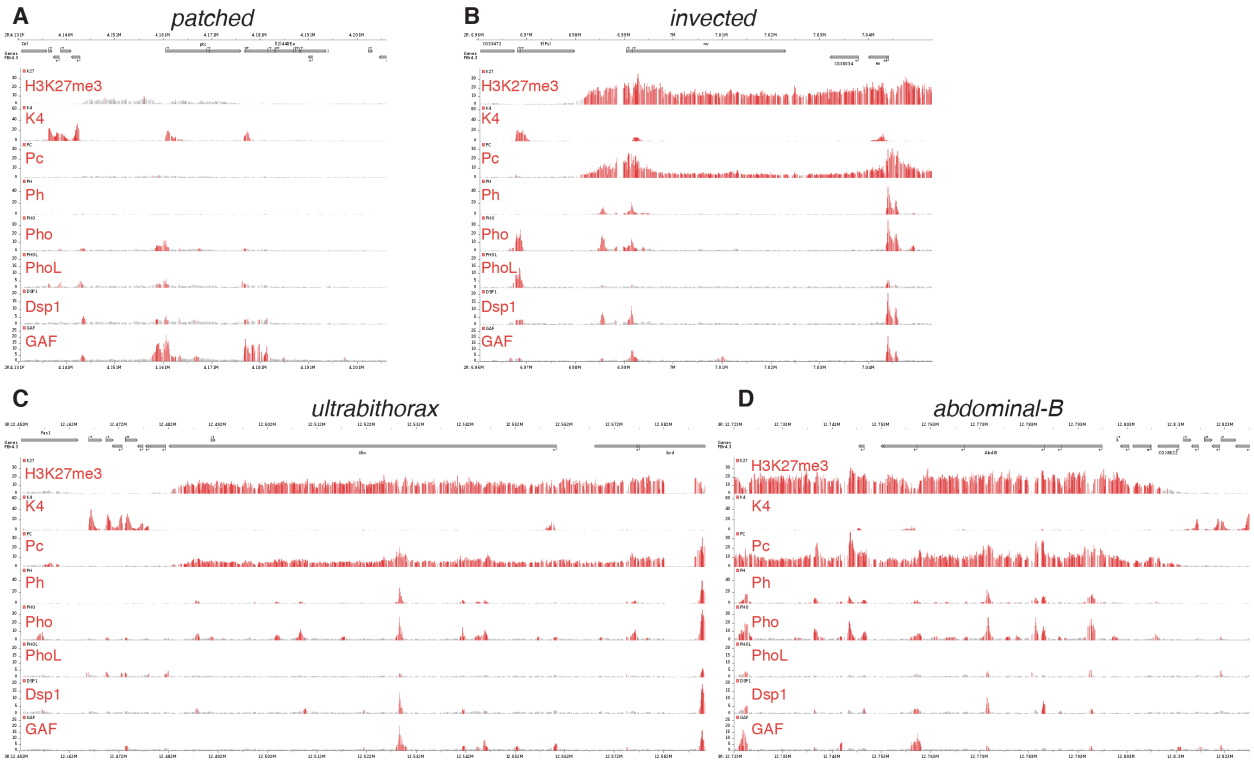


Figure 2.12 The *patched* promoter region, which is heavily bound by PcG proteins and H3K27me3-marked in larval discs, does not show PRE signatures in the embryo

(A) Embryo ChIP-seq binding data are taken the Cavalli group's Polycomb and Trithorax Genome Browser (Schuettengruber et al., 2009) and show the H3K4me3 mark (which is associated with active or potentially active promoters), the TrxG factor Trl/GAF (Oh et al., 2013), and the PcG protein Pho are present at the *ptc* promoter in the embryo. (B-D) Classic PRE-containing loci such as *en/inv*, *ubx* and *abd-B*. Embryonic *patched* (in A) has low levels of H3K27 trimethylation and poor occupancy of the PRE-associated proteins Polycomb (Pc) and Polyhomeotic (Ph) compared to these examples of PRE associated genes.

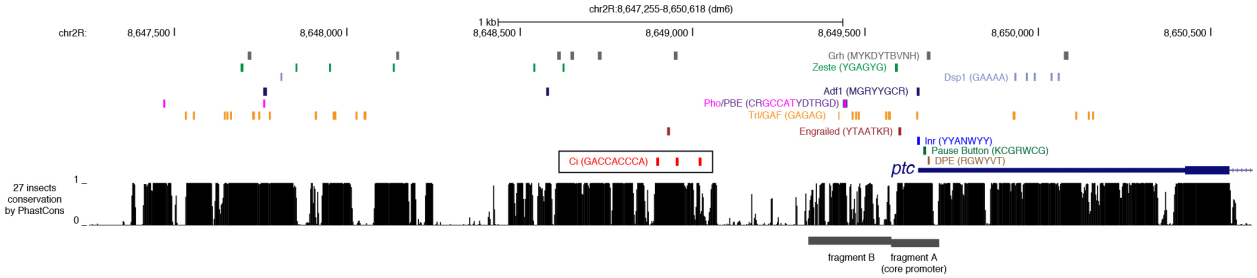


Figure 2.13 Conserved PRE-associated transcription factor binding motifs in promoter-proximal sequences of the *patched* gene

Gene map and sequence conservation histogram are taken from the UCSC Genome Browser (genome.ucsc.edu). The cluster of optimal GBS is outlined in a black box; various PRE-associated transcription factor binding motifs and core promoter motifs are shown as colored boxes. The *patched* promoter lacks a TATA motif in all species.

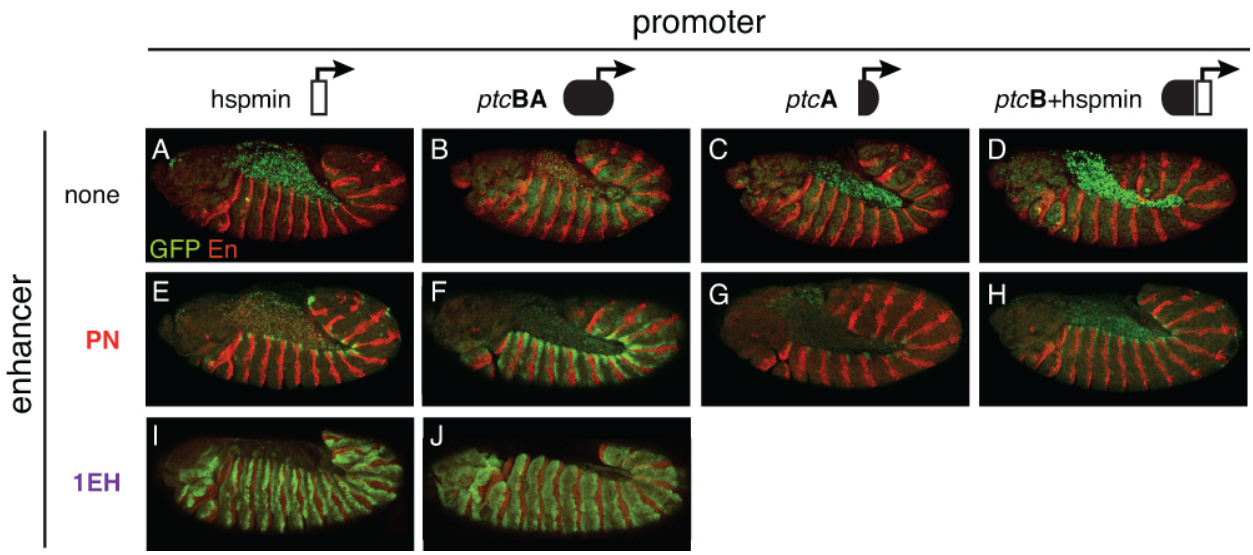


Figure 2.14 Polycomb/Trithorax Response Elements (PRE) near the *patched* promoter modulates enhancer activity

(A-J) Two embryonic ectodermal enhancers of *ptc* (PN, 1EH in rows) tested on different promoters (columns) in stage 12 transgenic embryos.

Enhancer	Transcription Factor	TF Family/Homologs	Consensus Binding Motif
ZY	Distaless (Dll)	DLX	TRATWDB
ZY	Tango/Spineless (Tgo/Ss)	bHLH-PAS/Arnt	RTCASRCR
YU	Scalloped (Sd)	TEAD	DGHATNT
YU	Engrailed (En)	En	YTAATKR
VT	Odd-Skipped (Odd)	Osr	SRGTAGC
VT	Engrailed (En)	En	YTAATKR
RP	Forkhead (Fkh)	Forkhead/FOXA	TRDDYAAACA
RP	Engrailed (En)	En	YTAATKR
PN	Pannier (Pnr)/dGATAa	GATA	WGATAA
LK	Sloppy Paired (Slp1)	Forkhead/FOXG	RWAAAYA
JB	GAGA Factor/Trl (GAF)	Trithorax	GAGAG
JB	Scalloped (Sd)	TEAD	DGHATNT
JB	Engrailed (En)	En	YTAATKR
1EH	GAGA Factor/Trl (GAF)	Trithorax	GAGAG
1EH	Pleiohomeotic (Pho)	YY1	GCCAT
1EH	Engrailed (En)	En	YTAATKR
B	GAGA Factor/Trl (GAF)	Trithorax	GAGAG
B	Pleiohomeotic (Pho)	YY1	GCCAT

Figure 2.15 Conserved TF motifs identified in *ptc* enhancer sequences

Consensus motifs were obtained from the CIS-BP database (Weirauch et al., 2014). Enhancer nomenclature is given in Figure 2A. IUPAC naming is used in the “Consensus Binding Motif” colu mn (R = A/G; Y = C/T; S = G/C; W = A/T; K = G/T; M = A/C; B = C/G/T; D = A/G/T; H = A/C/T; V = A/C/G; N = A/C/G/T).

Sequence conservation of selected TF binding motifs in enhancers of *patched.* Selected conserved Tango/Spineless (Tgo/Ss) motifs are in gray; Distaless (Dll) in light blue; Engrailed (En) in green; Pannier (Pnr/dGATAe) in purple; Scalloped (Sd) in blue; Cubitus interruptus (Ci) in red; Odd-Skipped (Odd) in dark red; Forkhead (Fkh) in dark yellow; Sloppy-paired (Slp) in blue-green; Pleiohomeotic (Pho) in pink, GAGA Factor (GAF/Trl) in yellow. Blocks of sequence separated by line breaks are not necessarily contiguous.

Enhancer ZY

```
 D. melanogaster  tct----------------tcgtgggtggcccttaat------------------tcgacgtca-
     D. simulans  tct----------------tcgtgggtggcccttaat------------------tcgacgtca-
    D. sechellia  tct----------------tcgtgggtggcccttaat------------------tcgacgtca-
       D. yakuba  tct----------------tcgtgggtggcccttaat------------------tcgacgtca-
       D. erecta  tct----------------tcgtgggtggccccttaat------------------tcgacgtca-
    D. biarmipes  tct----------------tcgtgggtggcccttaat------------------tcgtcgtca-
  D. bipectinata  tct----------------tcgtgggtggcccttaat------------------tcgtcgtcag
   D. eugracilis  tct----------------tcgtgggtggcccttaat------------------tcgtcgtca-
      D. elegans  tct----------------tcgtgggtggcccttaat------------------tcgtcgtca-
   D. takahashii  tct----------------tcgtgggtggcccttaat------------------tcgtcgtca-
     D. rhopaloa  tct----------------tcgtgggtggcccttaat------------------tcgtcgtca-
   D. ficusphila  tct----------------tcgtgggtggcccttaat------------------tcgtcgtca-
D. pseudoobscura  tcttgctctccgctctgcgtcgtgggtggcatttgat------tcgtcgtcgtcgtcgtcgtca-
   D. persimilis  tcttgctctccgctctgcgtcgtgggtggcatttgattcgtcgtcgtcgtcgtcgtcgtcgtca-
      D. miranda  tcttgctctccgctctgcgtcgtgggtggcatttgat------tcgtcgtcgtcgtcgtcgtca-
   D. willistoni  tct----------------tcgtgggtggcctt----------------------tcgttgtca-
      D. virilis  tct----------------tcgtgggtggcctttat------------------tcgttgtt--
   D. mojavensis  ----------------tcgtgggtggtctttat------------------tcgttgct--

 D. melanogaster  tcttcttgggtggtcca----------------------tattagccggcctgct--------------
     D. simulans  tcttcttgggtggtcca----------------------tattagccggcctgtt--------------
    D. sechellia  tcttcttgggtggtcca----------------------tattagccggcctgct--------------
       D. yakuba  tcttcttgggtggtcca----------------------tatttgccggcctgct--------------
       D. erecta  tcttcttgggtggtcca----------------------tattagccgtcctgct--------------
    D. biarmipes  tcctcttgggtggtcca----------------------tattagccggccagtt--------------
  D. bipectinata  tctgcttgggtggcctc----------------------tattagccgtcccgct--------------
   D. eugracilis  tcttcttgggtggtcca----------------------tattagctggccttct--------------
      D. elegans  tcttcttgggtggtctc----------------------tattagccggcctgct--------------
     D. kikkawai  tctgcttgggtggtc-----------------------tattagccggcctgct--------------
   D. takahashii  tcctcttgggtggtcca----------------------tattagccggcctgtt--------------
     D. rhopaloa  tcttcttgggtggtctc----------------------tattagccggcctgct--------------
   D. ficusphila  tcttcttgggtggtctc----------------------tattagccggcctgct--------------
D. pseudoobscura  tcatcatgggtggtc-----------------------tattagcccg--------------------
   D. persimilis  tcgtcatgggtggtc-----------------------tattagcccg--------------------
      D. miranda  tcgtcatgggtggtc-----------------------tattagcccg--------------------
   D. willistoni  tcgtcatgggtggtc-----------------------tattagctgcctcgcttctctatctttctc
      D. virilis  tcgtcttgggtggtctatctcttgccagtgtgtgtctctctgt---ctcgctcact--------------
   D. mojavensis  tcgcctgggtggtccaactcttggcagtgtgtggctctatat-agctcgctctcg--------------

 D. melanogaster  attgatcatt-ttctcgttgatagtgcgtgacg--t---ctctgggcag------------------cc
     D. simulans  attgatcatt-ttctcgttgatagtgcgtgacg--t---ctctgggcag------------------cc
    D. sechellia  attgatcatt-ttctcgttgatagtgcgtgacg--t---ctctgggcag------------------cc
       D. yakuba  attgatcatt-ttctcgttgatagtgcgtgacg--t---ctctgggcag------------------cc
       D. erecta  attgatcatt-ttctcgttgatagtgcgtgacg--t---ctctgggcag------------------cc
    D. biarmipes  attgatcatt-ttctcgttgatagtgcgtgacg--t---ctctgttcgg------------------ct
  D. bipectinata  attgatcattcccctcgttgatagtgcgtgacgt-t---ctccagtact------------------cc
   D. eugracilis  attgatcatt-ttctcgttgatagtgcgtgacg--tcgcctctggcga------------------ct
      D. elegans  attgatcatt-ttctcgttgatagtgcgtgacg--t---ctctgagtaga-----------------cc
     D. kikkawai  attgatcatt-ctctcgttgatagtgcgtgacg--a---cgcttgagagggctgccagaggagccgtcc
   D. takahashii  attgatcatt-ttctcgttgatagtgcgtgacg--t---ctcgggtcgg------------------ct
     D. rhopaloa  attgatcatt-ttctcgttgatagtgcgtgacg--t---ctctgggtagg-----------------cc
   D. ficusphila  attgatcatt-ttctcgttgatagtgcgtgacgtct---ctctggacag------------------cc
D. pseudoobscura  attgatcatt-ctctcgttgatagtgcgtgacg--t---tttagtgttg------------------ct
```

```
   D. persimilis  attgatcatt-ctctcgtttgatagtgcgtgacg--t---tttagtgttg------------------ct
      D. miranda  attgatcatt-ctctcgtttgatagtgcgtgacg--t---tttagtgttg------------------ct
   D. willistoni  attgatcatt-ttcttgtttgatagtgcgtgacg--t---tttcaactctcact--------------ct
      D. virilis  attgatcatt-ttcttgtttgatagtgcgtgacg--t---tttcgtaatg------------------ct
   D. mojavensis  attgatcatt-ttctcgtttaatagtgcgtgacg--t---tttcataatg------------------ct
```

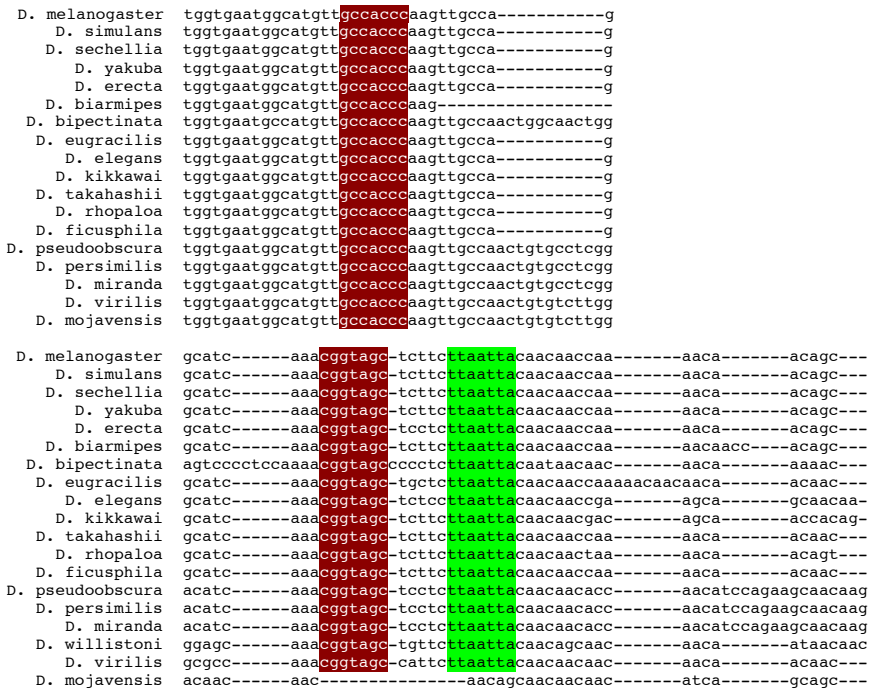
```
 D. melanogaster  cttctt-----ctgc------tgctcactcgc-----gctcatttattcgagcataaaattagcaattat
     D. simulans  cttctt-----ctgc------tgctcactcgc-----gctcatttattcgagcataaaattagcaattat
    D. sechellia  cttctt-----ctgc------tgctcactcgc-----gctcatttattcgagcataaaattagcaattat
       D. yakuba  cttctt-----ct--------cactccctcgc-----gctcatttgttcgagcataaaattagcaattac
       D. erecta  cttctg-----ct--------ccctcgctcgc-----gctcatttattcgagcataaaattagcaattat
    D. biarmipes  cttcgt-----ctgc------tgctcactcgc-----gctcatttattcgagcataaaattagcaattat
  D. bipectinata  cttcgc-----ctgc------tcttcattcgc-----gctcatttattcgagcataaaattagcaattat
    D. eugracilis  cttcgt-----ctccgctgctggctcactcgc-----gctcatttattcgagcataaaattagcaattat
       D. elegans  cttcgt-----ctgc------gtctcactcac-----gctcatttattcgagcataaaattagcaattat
      D. kikkawai  ctttgc-----ctgcccttc-gtctcgctagc-----tctcatttattcgagcataaaattagcaattat
    D. takahashii  cttcgt-----ctcc------tgctcactcgc-----gctcatttattcgagcataaaattagcaattat
      D. rhopaloa  cttcgt-----ctgc------ggctcactcgc-----gctcatttattcgagcataaaattagcaattat
    D. ficusphila  cttcgt-----ctgc------tgctcactcgc-----gctcatttattcgagcataaaattagcaattat
 D. pseudoobscura  cttctttactcttgg------ggcacactcacgctcagctcatttattcgagcataaaattagcaattat
    D. persimilis  cttcttaatccttgg------ggcacactcacgctcagctcatttattcgagcataaaattagcaattat
       D. miranda  cttctttactcttgg------ggcacactcacgctcagctcatttattcgagcataaaattagcaattat
    D. willistoni  ctactt---------------ggggcactcag-----gctcatttattcgagcataaaattagcaattat
       D. virilis  cttcaa---------------------------------cgtttttattcgagcataaaattagcaattat
    D. mojavensis  cttttg---------------------------------agtttttattcgagcataaaattagcaattat
```

Enhancer YU

```
 D. melanogaster  tgttggca-acaatta-ttgcgttgttgacagcaagatagatagcaatcatttgc
     D. simulans  tgttggca-acaatta-ttgcgttgttgacagcaagatagatagcaatcatttgc
    D. sechellia  tgttggca-acaatta-ttgcgttgttgacagcaagatagatagcaatcatttgc
       D. yakuba  tgttggca-acaatta-ctgtgttgttgacagcaagatagatagcaatcatttgc
       D. erecta  tgtt-gca-acaatta-ttgtgttgttgacagcaagatagatagcaatcatttgc
    D. biarmipes  tgtttgca-acaatta-ttgtgttgttgacagcaagatagataggaataatttgc
  D. bipectinata  ------ct-ac-atta-ttttgatgttgacagcaagatagatagcaataatttgc
    D. eugracilis  tgttggca-acaatta-ttgtgttgttgacagcaagatagatagcaataatttgc
       D. elegans  -----gca-acaatta-ttgtgatgttgacagccagatagatagcaataatttgc
      D. kikkawai  tgtcggca-acaatta-ttgtgttgttgacggccagatagatagcaataatttgc
    D. takahashii  tgttggca-acaatta-ttgtgttgttgacagcaagatagatagcaataatttgc
      D. rhopaloa  tgttggca-acaatta-ttgtgttgttgacagcaagatagatagcaataatttgc
    D. ficusphila  tgttggca-acaatta-ttgtgttgttgacagcaagatagatagcaacaatttgc
 D. pseudoobscura  tgttacca-acaatta-ttccgatgttgacagcaagatagatagcaataatttgc
    D. persimilis  tgttacca-acaatta-ttccgatgttgacagcaagatagatagcaataatttgc
       D. miranda  tgttacca-acaatta-ttccgatgttgacagcaagatagatagcaataatttgc
       D. virilis  tgttgcca-acaattattgtgttgttgacagcaagacagatagcaacagcagca
    D. mojavensis  tgttgcca-acaattattgtgttgttgacagcaagacagatagcaacagcaaca

 D. melanogaster  gtcccatcgtataattgcatatagcgccagtggaagttccggtcggttggc--tttttt
     D. simulans  gtccgatcgtataattgcatatagcgccagtggaagttccggtcggttggc--tttttt
    D. sechellia  gtccgatcgtataattgcatatagcgccagtggaagttccggtcggttggc--tttttt
       D. yakuba  -tgcgatcgtataattgcatatagcgccagtggaagttccggtcggttggc--tttttg
       D. erecta  ---cgatcggataattgcatatagcgccagtggaagttccggtcggttggc--tttttt
    D. biarmipes  ------tcgtataattgcatatagcgccagttgaagttccggtcggttggg--tttttt
  D. bipectinata  ------tcttataattgcatatagcgccagttgaagttccggtcggttgag--tttttg
    D. eugracilis  ------tcgtataattgcatatagcgccagttgaagttccggtcggttggg--tttttt
       D. elegans  ------tcgtataattgcatatagcgccagttgaagttccggtcggttggc--tttttt
      D. kikkawai  ------tcttataattgcatatagcgccagttgaagttccggtcggttggg--tttctt
    D. takahashii  ------tcgtataattgcatatagcgccagttgaagttccggtcggttggg--tttttt
      D. rhopaloa  ------tcgtataattgcatatagcgccagttgaagttccggtcggttggg--tttttt
    D. ficusphila  ------tcgtataattgcatatagcgccagttgaagttccggtccgttggg--tttttt
 D. pseudoobscura  ------tcttataattgcatatagcgccagttgcagttccggtcggttggg--tttttt
    D. persimilis  ------tcttataattgcatatagcgccagttgcagttccggtcggttggg--tttttt
       D. miranda  ------tcttataattgcatatagcgccagttgcagttccggtcggttgggttttttttt
    D. willistoni  ------ttttataattgcatatagcgccagttgcagttccggtccgatgag--tttttg
       D. virilis  ------gtttataattgcatatagcgccagttgcagttccggtccgttgag--tttttt
    D. mojavensis  ------gcttataattgcatatagcgccagttgcagttccggtcggttgag--tttttt
```


D. melanogaster tgttgctgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggct
D. simulans tgctgctgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggct
D. sechellia tgctgctgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggct
D. yakuba tgttgctgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggct
D. erecta tgtgcctgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggct
D. biarmipes tgtccctgtgtatgtgtgattggtgaatggcatgttgccacccaag----------
D. bipectinata tgtatacgagtatgtgtgattggtgaatgccatgttgccacccaagttgccaactggcaactggctggct
D. eugracilis tgctcctgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggcc
D. elegans tgttcttgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggct
D. kikkawai tgcccttgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggct
D. takahashii tgtcccctgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggct
D. rhopaloa tgtgcttgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttagct
D. ficusphila tgtgcttgtgtatgtgtgattggtgaatggcatgttgccacccaagttgcca----------gttggct
D. pseudoobscura agtactggtatatgtgtgattggtgaatggcatgttgccacccaagttgccaactgtgcctcggttggct
D. persimilis agaactggtatatgtgtgattggtgaatggcatgttgccacccaagttgccaactgtgcctcggttggct
D. miranda agtactggtatatgtgtgattggtgaatggcatgttgccacccaagttgccaactgtgcctcggttggct
D. virilis tgtttgtgtgtggtgtgggattggtgaatggcatgttgccacccaagttgccaactgtgtcttggttggct
D. mojavensis tgtgtgtgtatgagtgggattggtgaatggcatgttgccacccaagttgccaactgtgtcttggttggct

Enhancer VT


D. melanogaster tggtgaatggcatgttgccacccaagttgcca----------g
D. simulans tggtgaatggcatgttgccacccaagttgcca----------g
D. sechellia tggtgaatggcatgttgccacccaagttgcca----------g
D. yakuba tggtgaatggcatgttgccacccaagttgcca----------g
D. erecta tggtgaatggcatgttgccacccaagttgcca----------g
D. biarmipes tggtgaatggcatgttgccacccaag----------
D. bipectinata tggtgaatgccatgttgccacccaagttgccaactggcaactgg
D. eugracilis tggtgaatggcatgttgccacccaagttgcca----------g
D. elegans tggtgaatggcatgttgccacccaagttgcca----------g
D. kikkawai tggtgaatggcatgttgccacccaagttgcca----------g
D. takahashii tggtgaatggcatgttgccacccaagttgcca----------g
D. rhopaloa tggtgaatggcatgttgccacccaagttgcca----------g
D. ficusphila tggtgaatggcatgttgccacccaagttgcca----------g
D. pseudoobscura tggtgaatggcatgttgccacccaagttgccaactgtgcctcgg
D. persimilis tggtgaatggcatgttgccacccaagttgccaactgtgcctcgg
D. miranda tggtgaatggcatgttgccacccaagttgccaactgtgcctcgg
D. virilis tggtgaatggcatgttgccacccaagttgccaactgtgtcttgg
D. mojavensis tggtgaatggcatgttgccacccaagttgccaactgtgtcttgg
D. melanogaster gcatc------aaacggtagc-tcttcttaattacaacaaccaa-------aaca-------acagc---
D. simulans gcatc------aaacggtagc-tcttcttaattacaacaaccaa-------aaca-------acagc---
D. sechellia gcatc------aaacggtagc-tcttcttaattacaacaaccaa-------aaca-------acagc---
D. yakuba gcatc------aaacggtagc-tcttcttaattacaacaaccaa-------aaca-------acagc---
D. erecta gcatc------aaacggtagc-tcctcttaattacaacaaccaa-------aaca-------acagc---
D. biarmipes gcatc------aaacggtagc-tcttcttaattacaacaaccaa-------aacaacc----acagc---
D. bipectinata agtcccctccaaaacggtagccccctcttaattacaataacaac-------aaca-------aaaac---
D. eugracilis gcatc------aaacggtagc-tgctcttaattacaacaaccaaaaacaacaaca-------acaac---
D. elegans gcatc------aaacggtagc-tctcctaattacaacaaccga-------agca-------gcaacaa-
D. kikkawai gcatc------aaacggtagc-tcttcttaattacaacaacgac-------agca-------accacag-
D. takahashii gcatc------aaacggtagc-tcttcttaattacaacaaccaa-------aaca-------acaac---
D. rhopaloa gcatc------aaacggtagc-tcttcttaattacaacaactaa-------aaca-------acagt---
D. ficusphila gcatc------aaacggtagc-tcttcttaattacaacaaccaa-------aaca-------acaac---
D. pseudoobscura acatc------aaacggtagc-tcctcttaattacaacaacacc-------aacatccagaagcaacaag
D. persimilis acatc------aaacggtagc-tcctcttaattacaacaacacc-------aacatccagaagcaacaag
D. miranda acatc------aaacggtagc-tcctcttaattacaacaacacc-------aacatccagaagcaacaag
D. willistoni ggagc------aaacggtagc-tgttcttaattacaacagcaac-------aaca-------ataacaac
D. virilis gcgcc------aaacggtagc-cattcttaattacaacaacaac-------aaca-------acaac---
D. mojavensis acaac------aac----------aacagcaacaacaac-------atca-------gcagc---

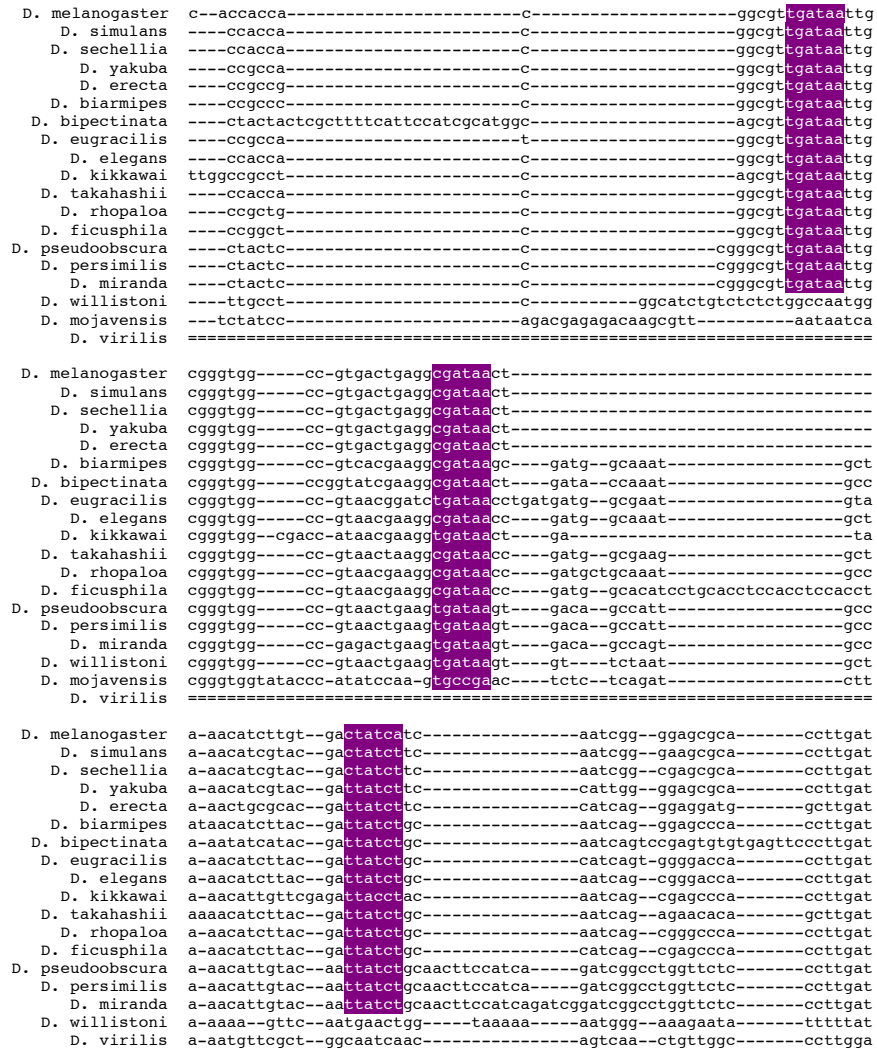
```
 D. melanogaster  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggtaacgacaaag--agccgc-ttc
     D. simulans  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggtaacgacaaag--agccgc-ttc
    D. sechellia  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggtaacgacaaag--agccgc-ttc
       D. yakuba  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggta-------------------tc
       D. erecta  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggtaacgacaaag--agccgc-gtc
    D. biarmipes  aatcggaaacggaaatcgctaatg-aaa--aatctccatatttatggtggcgataaag--agccac----
  D. bipectinata  aatcggaaacggaaatcactaatg-aaa--aatttccatatttatggtggcaataaag--agtc-----t
   D. eugracilis  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggtggcgataaag--agccac----
      D. elegans  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggtggcgataaag--agccca----
     D. kikkawai  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggtggcgataaag--cgatac-aac
   D. takahashii  aatcggaaacggaaatcgctaatg-aaa--aatctccatatttatggcggcgataaag--agccac----
     D. rhopaloa  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggcggcgataaag--agccc-----
   D. ficusphila  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatggcggcgataaag--agccgcagcc
D. pseudoobscura  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatgcagccgataaag--cgctac----
   D. persimilis  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatgcagccgataaag--cgctac----
      D. miranda  aatcggaaacggaaatcgctaatg-aaa--aatttccatatttatgcagccgataaag--cgctac----
   D. willistoni  aatcggaaacggaaatcgctaatgaaaa--aatttccatatttatgc--------atg--tgcc------
      D. virilis  aatcggaaacggatatcgctaatg-aaa--aatttccatatttatg-tggcgataaaagcagc-ac----
   D. mojavensis  aatcggaaacggatatcgctaatg-aaacgaatttccatatttatg-tgttaataaaaccagcaac----

 D. melanogaster  tacgggctgcaattaattccaacgaa-attta--attag-aat
     D. simulans  tacgggctgcaattaattccaacgaa-attta--attag-aat
    D. sechellia  tacgggctgcaattaattccaacgaa-attta--attag-aat
       D. yakuba  tacggactgcaattaattccaacgaa-attta--attag-aat
       D. erecta  tacgggctgcaattaattccaaccaa-attta--attag-aat
    D. biarmipes  cacgggctgcaattaattccaccgaa-attta--attag-aat
  D. bipectinata  tatgtgttgcaattaattccaatgaa-attta--attag-aat
   D. eugracilis  tacgggctgcaattaattccaacgaa-attta--attag-aat
      D. elegans  tatgggctgcaattaattccaacaaa-attta--attag-aat
     D. kikkawai  tacggactgcaattaattccaacgaa-attta--attag-aat
   D. takahashii  tacgggctgcaattaattccaacgaa-attta--attag-aat
     D. rhopaloa  tatgggctgcaattaattccaacgaa-attta--attag-aat
   D. ficusphila  tatgggctgcaattaatcccaacgaa-attta--attag-aat
D. pseudoobscura  tatgtgtcgcaattaatttcaacgaa-attta--attag-aat
   D. persimilis  tatgtgtcgcaattaatttcaacgaa-attta--attag-aat
      D. miranda  tatgtgtcgcaattaattttaacgaa-attta--attag-aat
   D. willistoni  catag--cgcaattaatttcatcaaa-tttca--ttcagaaat
      D. virilis  tgtgtgatgcaattaatttcatcgaatatttgttgttgg-aat
```

Enhancer RP

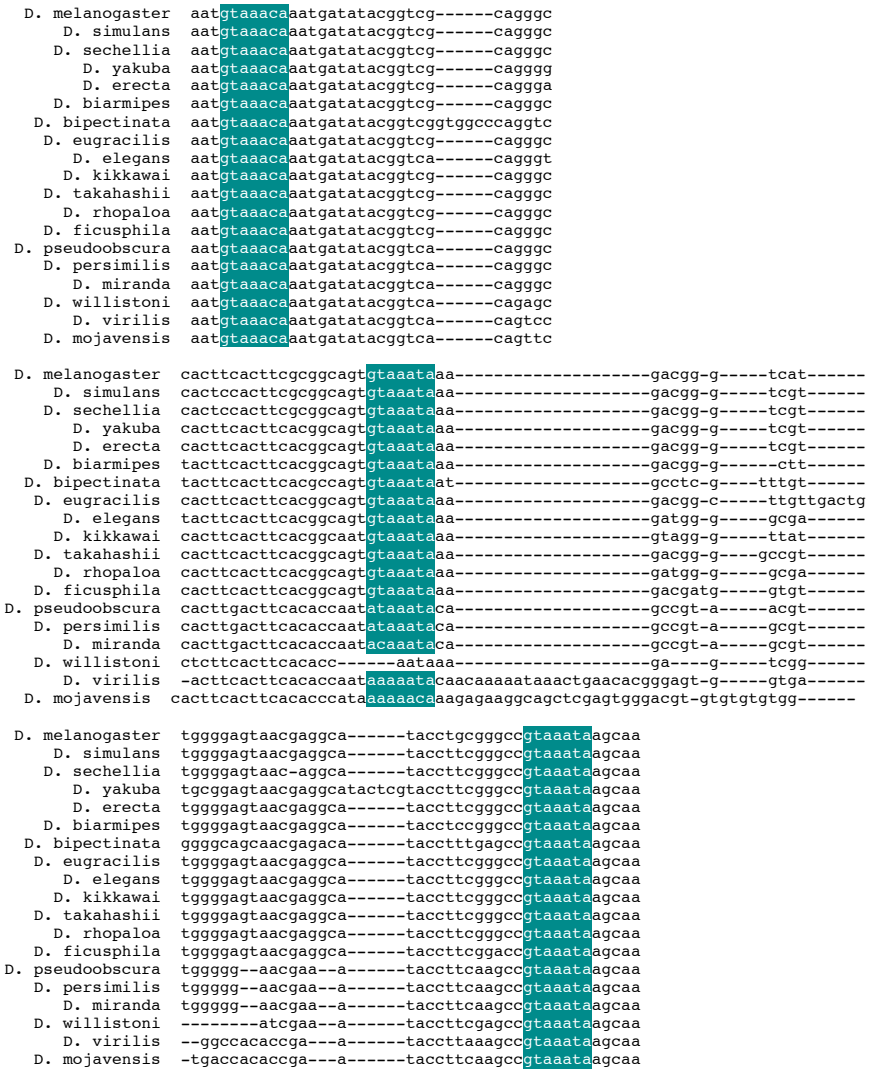
```
 D. melanogaster  gcccgtgcccaatggccataattaattgct
     D. simulans  gcccgtgcccaatggccataattaattgct
    D. sechellia  gcccgtgcccaatggccataattaattgct
       D. yakuba  gcccgtgcccaatggccataattaattgct
       D. erecta  gcccgtgcccaatggccataattaattgct
    D. biarmipes  gcccgtgcccaatggccataattaattgct
  D. bipectinata  gtccaaacccgatggctataattaattgct
   D. eugracilis  gcccgtgcccaatggccataattaattgct
      D. elegans  gcccgtgcccaatggccataattaattgct
     D. kikkawai  gtccgtgcccaatggctataattaattgct
   D. takahashii  gcccgtgcccaatggccataattaattgct
     D. rhopaloa  gcccgtgcccaatggccataattaattgct
   D. ficusphila  gcccgtgcccaatggccataattaattgct
D. pseudoobscura  gtccctgcccaatggccataattaatcgct
   D. persimilis  gtccctgcccaatggccataattaatcgct
      D. miranda  gtccctgcccaatggccataattaatcgct
   D. willistoni  gttcatgcccattggctataattaattggt
      D. virilis  tccaatactcaatgaccataattaattggt
```

```
 D. melanogaster  ag--tcgtgctaaggtcaaacaaacagcc
       D. yakuba  ag--tcgtgctaaggtcaaacacacagcc
       D. erecta  ag--tcgtgctaaggtcaaacacacagcc
    D. biarmipes  aa--ccgtgctaaggtcaaacacacagcc
  D. bipectinata  aa--ccatgctaaggtcaaacacagagcc
   D. eugracilis  aa--ccgtgctaaggtcaaacacacagcc
      D. elegans  aa--tcatgctaaggtcaaacacacagcc
     D. kikkawai  aa--gcatgctaaggtcaaacacagagcc
   D. takahashii  aa--tcgtgctaaggtcaaacacacagcc
     D. rhopaloa  aa--tcgtgctaaggtcaaacac--agcc
   D. ficusphila  aa--tcgtgctaaggtcaaacacacagcc
D. pseudoobscura  aa--ccatgctaaggtcaaacacacagcc
   D. persimilis  aa--ccatgctaaggtcaaacacacagcc
      D. miranda  aa--ccatgctaaggtcaaacacacagcc
   D. willistoni  gatttcaggcaaaggtcaaacacgtgcta
      D. virilis  gc--acatgctgaggtcaaacgcacagag
   D. mojavensis  gc--acatgctgaggtcaaagcacacac

 D. melanogaster  agccgtgacag-acttgtgggcaaacagaaacca
    D. sechellia  agccgtgacag-acttgtgggcaaacagacacaa
       D. yakuba  agccgtgacag-acttgtgggcaaacagacacca
       D. erecta  ggccgtaacag-acttgtgggcaaacagacacca
    D. biarmipes  agccgtggcag-acttgtgggtaaacagactcca
  D. bipectinata  ggccgtgacag-acttgtgggtaaacagacacca
   D. eugracilis  agccgtgacag-acttgtgggtaaacagacacca
      D. elegans  agccgtgacag-acttgtgggtaaacagaaacca
     D. kikkawai  ggccgtgacag-acttgtgggtaaacagacacca
   D. takahashii  agccgtgacag-acttgtgggtaaacagaaacca
     D. rhopaloa  agccgtgacag-acttgtgggtaaacagacacca
   D. ficusphila  agccgtgactg-acttgtgggtaaacagaaacca
D. pseudoobscura  ggccgtgacag-acttgtgggtaaacagaaacca
   D. persimilis  ggccgtgacag-acttgtgggtaaacagaaacca
      D. miranda  ggccgtgacag-acttgtgggtaaacagaaacca
   D. willistoni  aaccgtgacaacaattgtgggtaaacagaaacca
      D. virilis  -accatgacgt-tgttgcgggtaaataaaacca
   D. mojavensis  -accatgacca-tgttgtgggtaaatagaactca

 D. melanogaster  ttc--acacaaccgcaaagcagttgaacaaacaatgat------cc------g
     D. simulans  ttc--acacaaccgcaaagcagttgaacaaacaatgat------cc------a
    D. sechellia  ttc--acacaaccgcaaagcagttgaacaaacaatgat------cc------g
       D. yakuba  ttt--acacaaccgcaaggcagttgaacaaacaatgat------cc------g
       D. erecta  ttt--acacaaccgcaaagcagttgaacaaaca----------------------
    D. biarmipes  ttc--acacaaccgcaaagcagttgaacaaacaatgat------ct------g
  D. bipectinata  tcc--acacaaccgcaaaacagttgaacaaacaatgat------cc------g
   D. eugracilis  ttc--acacaaccgcaaagcagttgaacaaacaatgat------ct------g
      D. elegans  ttc--acacaaccgcaaagcagttgaacaaacaatgat------ct------g
     D. kikkawai  ttt--acacaaccgcagaacagttgagcaaacaatgat------ccacaacag
   D. takahashii  ttc--acacaaccgcaaagcagttgaacaaacaatgat------cc------g
     D. rhopaloa  tcc--acacaaccgcaaagcagttgaacaaacaatgct------ct------g
   D. ficusphila  ttc--acacaaccgcaaagcagttgaacaaacaatgct------cc------g
D. pseudoobscura  ttt--acacaaccgcaaaacagttggacaaacaatgat------cc------g
   D. persimilis  ttt--acacaaccgcaaaacagttggacaaacaatgat------cc------g
      D. miranda  ttt--acacaaccgcaaaacagttggacaaacaatgat------cc------g
   D. willistoni  ttt--acacaaccgcaaaacatttgtacaaacaatgat------cg------g
      D. virilis  ttt--acacagccgcag------------aacaatgat----ctct------t
   D. mojavensis  tttacacacaaccacaa-------gaac-aacaatgatttccctct------
```

Enhancer PN

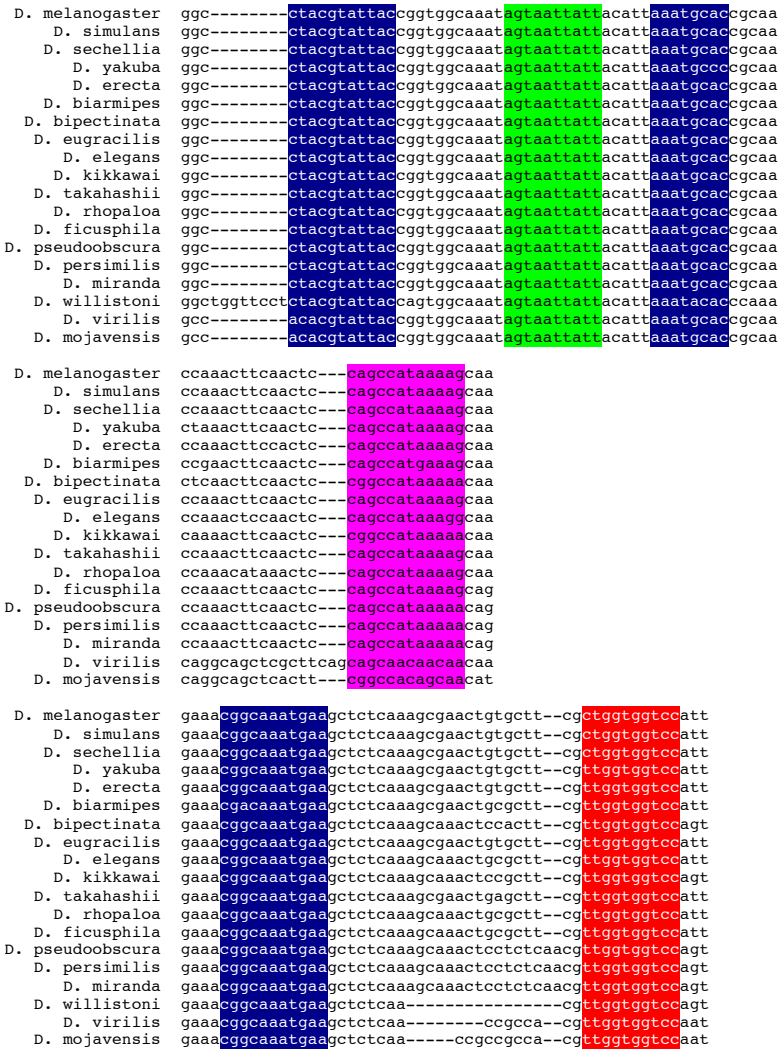

D. melanogaster
D. simulans
D. sechellia
D. yakuba
D. erecta
D. biarmipes
D. bipectinata
D. eugracilis
D. elegans
D. kikkawai
D. takahashii
D. rhopaloa
D. ficusphila
D. pseudoobscura
D. persimilis
D. miranda
D. willistoni
D. mojavensis
D. virilis

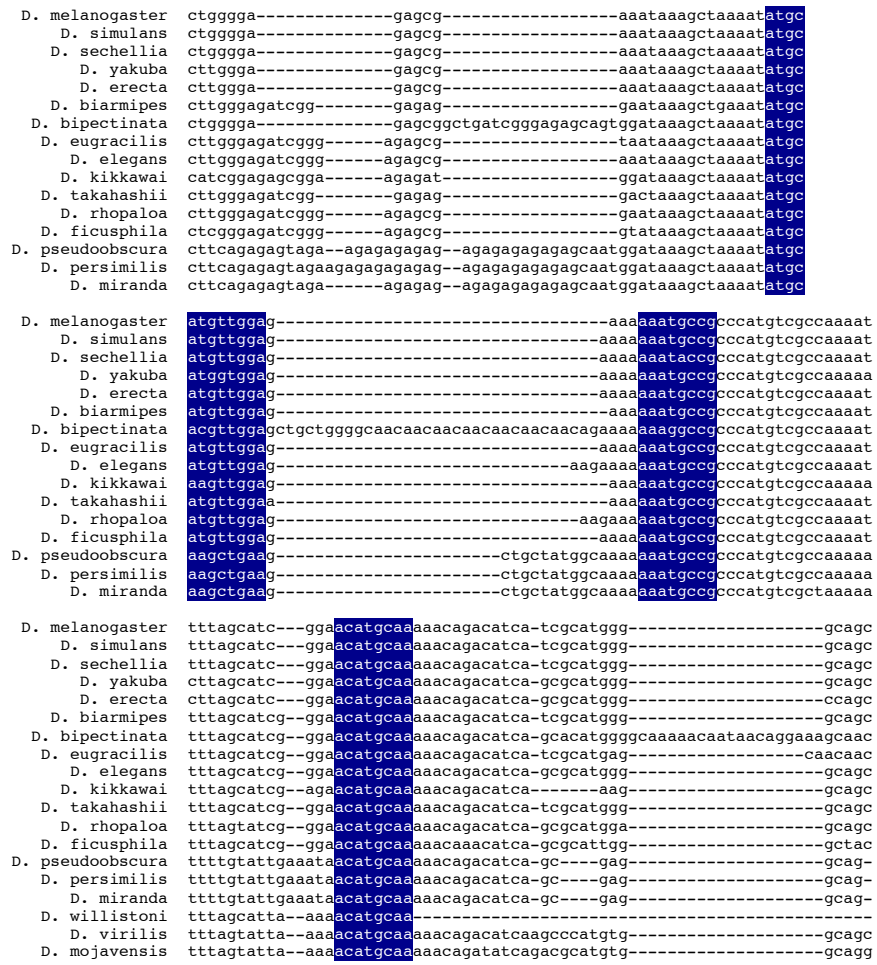
Enhancer LJ

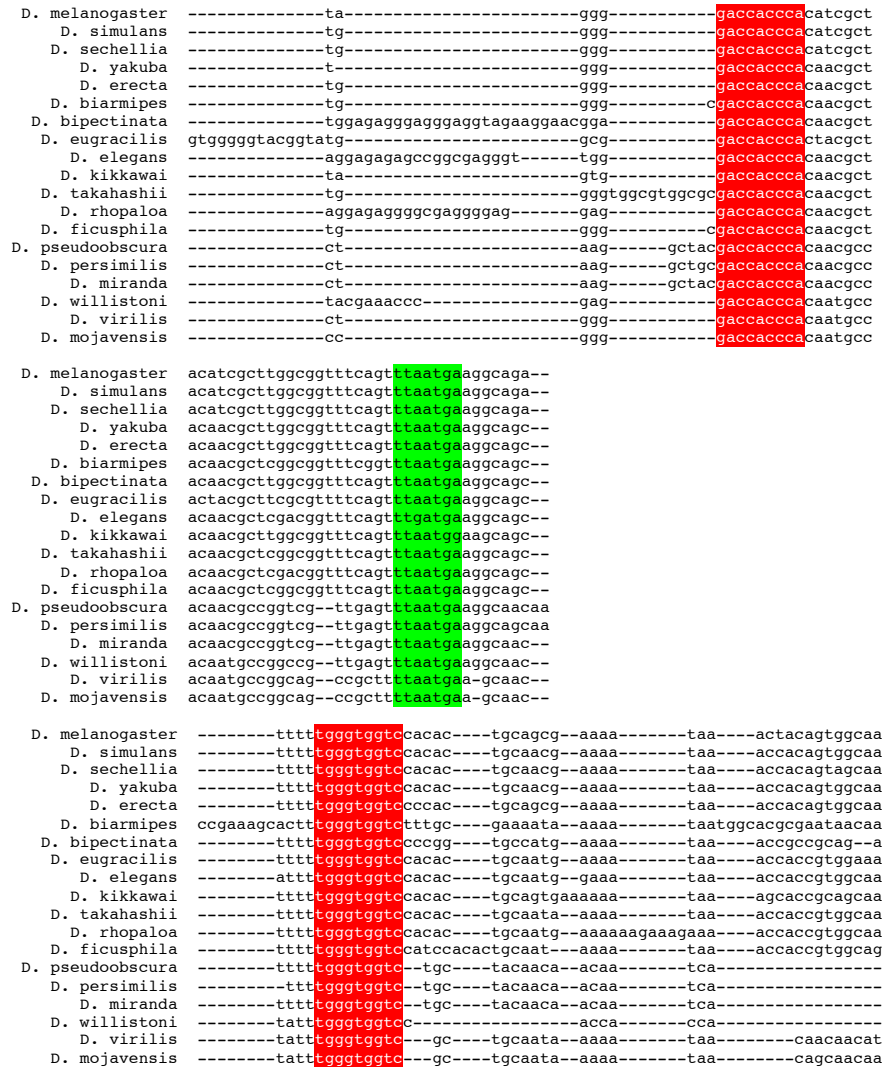

D. melanogaster aatgtaaacaaatgatatacggtcg------cagggc
D. simulans aatgtaaacaaatgatatacggtcg------cagggc
D. sechellia aatgtaaacaaatgatatacggtcg------cagggc
D. yakuba aatgtaaacaaatgatatacggtcg------cagggg
D. erecta aatgtaaacaaatgatatacggtcg------caggga
D. biarmipes aatgtaaacaaatgatatacggtcg------cagggc
D. bipectinata aatgtaaacaaatgatatacggtcggtggcccaggtc
D. eugracilis aatgtaaacaaatgatatacggtcg------cagggc
D. elegans aatgtaaacaaatgatatacggtca------cagggt
D. kikkawai aatgtaaacaaatgatatacggtcg------cagggc
D. takahashii aatgtaaacaaatgatatacggtcg------cagggc
D. rhopaloa aatgtaaacaaatgatatacggtcg------cagggc
D. ficusphila aatgtaaacaaatgatatacggtcg------cagggc
D. pseudoobscura aatgtaaacaaatgatatacggtca------cagggc
D. persimilis aatgtaaacaaatgatatacggtca------cagggc
D. miranda aatgtaaacaaatgatatacggtca------cagggc
D. willistoni aatgtaaacaaatgatatacggtca------cagagc
D. virilis aatgtaaacaaatgatatacggtca------cagtcc
D. mojavensis aatgtaaacaaatgatatacggtca------cagttc
D. melanogaster cacttcacttcgcggcagtgtaaataa----------gacgg-g-----tcat------
D. simulans cactccacttcgcggcagtgtaaataa----------gacgg-g-----tcgt------
D. sechellia cactccacttcgcggcagtgtaaataa----------gacgg-g-----tcgt------
D. yakuba cacttcacttcacggcagtgtaaataa----------gacgg-g-----tcgt------
D. erecta cacttcacttcacggcagtgtaaataa----------gacgg-g-----tcgt------
D. biarmipes tacttcacttcacggcagtgtaaataa----------gacgg-g------ctt------
D. bipectinata tacttcacttcacgccagtgtaaataat----------gcctc-g----tttgt------
D. eugracilis cacttcacttcacggcagtgtaaataa----------gacgg-c-----ttgttgactg
D. elegans tacttcacttcacggcagtgtaaataa----------gatgg-g-----gcga------
D. kikkawai cacttcacttcacggcaatgtaaataa----------gtagg-g-----ttat------
D. takahashii cacttcacttcacggcagtgtaaataa----------gacgg-g----gccgt------
D. rhopaloa cacttcacttcacggcagtgtaaataa----------gatgg-g-----gcga------
D. ficusphila cacttcacttcacggcagtgtaaataa----------gacgatg-----gtgt------
D. pseudoobscura cacttgacttcacaccaatataaataca----------gccgt-a-----acgt------
D. persimilis cacttgacttcacaccaatataaataca----------gccgt-a-----gcgt------
D. miranda cacttgacttcacaccaatacaaataca----------gccgt-a-----gcgt------
D. willistoni ctcttcacttcacacc------aataaa----------ga----g-----tcgg------
D. virilis -acttcacttcacaccaataaaaataacaacaaaaataaactgaacacgggagt-g-----gtga------
D. mojavensis cacttcacttcacacccataaaaacaagagaaggcagctcgagtgggacgt-gtgtgtgtgg------
D. melanogaster tggggagtaacgaggca------tacctgcgggccgtaaataagcaa
D. simulans tggggagtaacgaggca------taccttcgggccgtaaataagcaa
D. sechellia tggggagtaac-aggca------taccttcgggccgtaaataagcaa
D. yakuba tgcggagtaacgaggcatactcgtaccttcgggccgtaaataagcaa
D. erecta tggggagtaacgaggca------taccttcgggccgtaaataagcaa
D. biarmipes tggggagtaacgaggca------tacctccgggccgtaaataagcaa
D. bipectinata ggggcagcaacgagaca------tacctttgagccgtaaataagcaa
D. eugracilis tggggagtaacgaggca------taccttcgggccgtaaataagcaa
D. elegans tggggagtaacgaggca------taccttcgggccgtaaataagcaa
D. kikkawai tggggagtaacgaggca------taccttcgggccgtaaataagcaa
D. takahashii tggggagtaacgaggca------taccttcgggccgtaaataagcaa
D. rhopaloa tggggagtaacgaggca------taccttcgggccgtaaataagcaa
D. ficusphila tggggagtaacgaggca------taccttcggaccgtaaataagcaa
D. pseudoobscura tggggg--aacgaa--a------taccttcaagccgtaaataagcaa
D. persimilis tggggg--aacgaa--a------taccttcaagccgtaaataagcaa
D. miranda tggggg--aacgaa--a------taccttcaagccgtaaataagcaa
D. willistoni --------atcgaa--a------taccttcgagccgtaaataagcaa
D. virilis --ggccacaccga---a------taccttaaagccgtaaataagcaa
D. mojavensis -tgaccacaccga---a------taccttcaagccgtaaataagcaa

Enhancer JB

```
 D. melanogaster  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccagc
     D. simulans  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccagc
    D. sechellia  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccagc
       D. yakuba  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccagc
       D. erecta  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccagc
    D. biarmipes  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccagc
  D. bipectinata  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccaac
   D. eugracilis  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccagc
      D. elegans  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccaga
     D. kikkawai  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccaac
   D. takahashii  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccagc
     D. rhopaloa  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccaga
   D. ficusphila  gagaccaccca ggtaacgatttgtgcagtcattccacgaattccagc
D. pseudoobscura  ggaccaccca ggtaacgatttgtgcagtcattccacgaattctcgga
   D. persimilis  ggaccaccca ggtaacgatttgtgcagtcattccacgaattctcgga
      D. miranda  ggaccaccca ggtaacgatttgtgcagtcattccacgaattctcgga
   D. willistoni  gag--cacaggtataacgataagagaagt-------------------
      D. virilis  gagaccaccca ggcgtcgatttgtgcagtcattccacgaattcttggc
   D. mojavensis  gagaccaccca ggcgacgatttgtgcagtcattccacgaattcttggc

 D. melanogaster  cgaaattaatttgaca gccaacgcaaaacgcatgcgtcgtcgacattcctcgcc----------------
     D. simulans  cgaaattaatttgaca gccaacgcaaaacgcatgcgtcgtcgacattcctcgcc----------------
    D. sechellia  cgaaattaatttgaca gccatcgcaaaacgcatgcgtcgtcgacattcctcgcc----------------
       D. yakuba  tgaaattaatttgaca gccaacgcaaaacgcatgcgtcgtcggcattcctcgccagcactctacccccctt
       D. erecta  tgaaattaatttgaca gccaacgcaaaacgcatg---cgtcgacattcttcgcc----------------
    D. biarmipes  cgaaattaatttgaca gccaacgcaaaacgcatgcttcgtcaacattcctcgcc----------------
  D. bipectinata  cgaaattaatttgaca gccaacgcaaaacgcatgcgtcttcgacattcttcact----------------
   D. eugracilis  cgaaattaatttgaca gccaacgcaaaacgcatgcgtcgtcaacattcttcgcc----------------
      D. elegans  cgaaattaatttgaca gccaacgcaaaacgcatgcgtcgtcaacattcctcgcc----------------
     D. kikkawai  cgaaattaatttgaca gccaacgcaaaacgcatgcgtcgtcaacattcctggcc----------------
   D. takahashii  cgaaattaatttgaca gccaacgcaaaacgcatgcttcgtcaacattcctcgcc----------------
     D. rhopaloa  cgaaattaatttgaca gccaacgcaaaacgcatgcgtcgtcaacattcctcgcc----------------
   D. ficusphila  cgaaattaatttgaca gccaacgcaaaacgcatgcgtcgtcaacattcctcgcc----------------
D. pseudoobscura  cgaaattaatttgaca gccaacgcaaaacgcatgcatcgtcaacattcttctgg----------------
   D. persimilis  cgaaattaatttgaca gccaacgcaaaacgcatgcatcgtcaacattcttctgg----------------
      D. miranda  cgaaattaatttgaca gccaacgcaaaacgcatgcatcgtcaacattcttctgg----------------
   D. willistoni  cgaaattaatttgaca accaacgcaaaacgcatgcgtcgtcaacatttttcagc----------------
      D. virilis  cgaaattaatttgaca accaacgcaaaacgcatgcgccttcaacattccacaac----------------
   D. mojavensis  cgaaattaatttgaca accaacgcaaaacgcatgcgccgtcaacattccatgac----------------

 D. melanogaster  cgtggcggttgattacatgtacaaacgaaccgag
     D. simulans  cgtggcggttgattacatgtacaaacgaaccgag
    D. sechellia  cgtggcggttgattacatgtacaaacgaaccgag
       D. yakuba  cgtggcggttgattacacgtataaacgaagcgag
       D. erecta  cgtggcggttgattacacgtataaacgaacccag
    D. biarmipes  cggggcggttgattacatgaacaaacgaaccgaa
  D. bipectinata  cggggcggttgattacatgaacaaacgaaccgat
   D. eugracilis  cggggcggttgattacatgaacaaacgaaccgaa
      D. elegans  cggggcggttgattacatgaacaaacgaaccgaa
     D. kikkawai  cggggcggttgattacatgaacaaacgaaccgat
   D. takahashii  cggggcggttgattacatgaacaaacgaaccgaa
     D. rhopaloa  cggggcggttgattacatgaacaaacgaaccgaa
   D. ficusphila  cg-ggcggttgattacatgaacaaacgaaccgag
D. pseudoobscura  cggggcggttgattacatgaacaaacgaaccgaa
   D. persimilis  cggggcggttgattacatgaacaaacgaaccgaa
      D. miranda  cggggcggttgattacatgaacaaacgaaccgaa
   D. willistoni  ctgggcggttgattaca-----------------
      D. virilis  ctcgacggttgattacatgaacaaacgaagcgaa
   D. mojavensis  ctggacggttgattacatgaacaaacgaagcgaa
```


D. melanogaster ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. simulans ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. sechellia ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. yakuba ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcccgcaa
D. erecta ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. biarmipes ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. bipectinata ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. eugracilis ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. elegans ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. kikkawai ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. takahashii ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. rhopaloa ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. ficusphila ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. pseudoobscura ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. persimilis ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. miranda ggc--------ctacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. willistoni ggctggttcctctacgtattaccagtggcaaatagtaattattacattaaatacaccaaa
D. virilis gcc--------acacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. mojavensis gcc--------acacgtattaccggtggcaaatagtaattattacattaaatgcaccgcaa
D. melanogaster ccaaacttcaactc---cagccataaaagcaa
D. simulans ccaaacttcaactc---cagccataaaagcaa
D. sechellia ccaaacttcaactc---cagccataaaagcaa
D. yakuba ctaaacttcaactc---cagccataaaagcaa
D. erecta ccaaacttccactc---cagccataaaagcaa
D. biarmipes ccgaacttcaactc---cagccatgaaagcaa
D. bipectinata ctcaacttcaactc---cggccataaaaacaa
D. eugracilis ccaaacttcaactc---cagccataaaagcaa
D. elegans ccaaactccaactc---cagccataaaggcaa
D. kikkawai caaaacttcaactc---cggccataaaaacaa
D. takahashii ccaaacttcaactc---cagccataaaagcaa
D. rhopaloa ccaaacataaactc---cagccataaaagcaa
D. ficusphila ccaaacttcaactc---cagccataaaagcag
D. pseudoobscura ccaaacttcaactc---cagccataaaaacag
D. persimilis ccaaacttcaactc---cagccataaaaacag
D. miranda ccaaacttcaactc---cagccataaaaacag
D. virilis caggcagctcgcttcagcagcaacaacaacaa
D. mojavensis caggcagctcactt---cggccacagcaacat
D. melanogaster gaaacggcaaatgaagctctcaaagcgaactgtgctt--cgctggtggtccatt
D. simulans gaaacggcaaatgaagctctcaaagcgaactgtgctt--cgctggtggtccatt
D. sechellia gaaacggcaaatgaagctctcaaagcgaactgtgctt--cgctggtggtccatt
D. yakuba gaaacggcaaatgaagctctcaaagcgaactgtgctt--cgttggtggtccatt
D. erecta gaaacggcaaatgaagctctcaaagcgaactgtgctt--cgttggtggtccatt
D. biarmipes gaaacgacaaatgaagctctcaaagcgaactgcgctt--cgttggtggtccatt
D. bipectinata gaaacggcaaatgaagctctcaaagcaaactccactt--cgttggtggtccagt
D. eugracilis gaaacggcaaatgaagctctcaaagcgaactgtgctt--cgttggtggtccatt
D. elegans gaaacggcaaatgaagctctcaaagcaaactgcgctt--cgttggtggtccatt
D. kikkawai gaaacggcaaatgaagctctcaaagcaaactccgctt--cgttggtggtccagt
D. takahashii gaaacggcaaatgaagctctcaaagcgaactgagctt--cgttggtggtccatt
D. rhopaloa gaaacggcaaatgaagctctcaaagcaaactgcgctt--cgttggtggtccatt
D. ficusphila gaaacggcaaatgaagctctcaaagcaaactgcgctt--cgttggtggtccatt
D. pseudoobscura gaaacggcaaatgaagctctcaaagcaaactcctctcaacgttggtggtccagt
D. persimilis gaaacggcaaatgaagctctcaaagcaaactcctctcaacgttggtggtccagt
D. miranda gaaacggcaaatgaagctctcaaagcaaactcctctcaacgttggtggtccagt
D. willistoni gaaacggcaaatgaagctctcaa----------cgttggtggtccagt
D. virilis gaaacggcaaatgaagctctcaa--------ccgcca--cgttggtggtccaat
D. mojavensis gaaacggcaaatgaagctctcaa-----ccgccgcca--cgttggtggtccaat

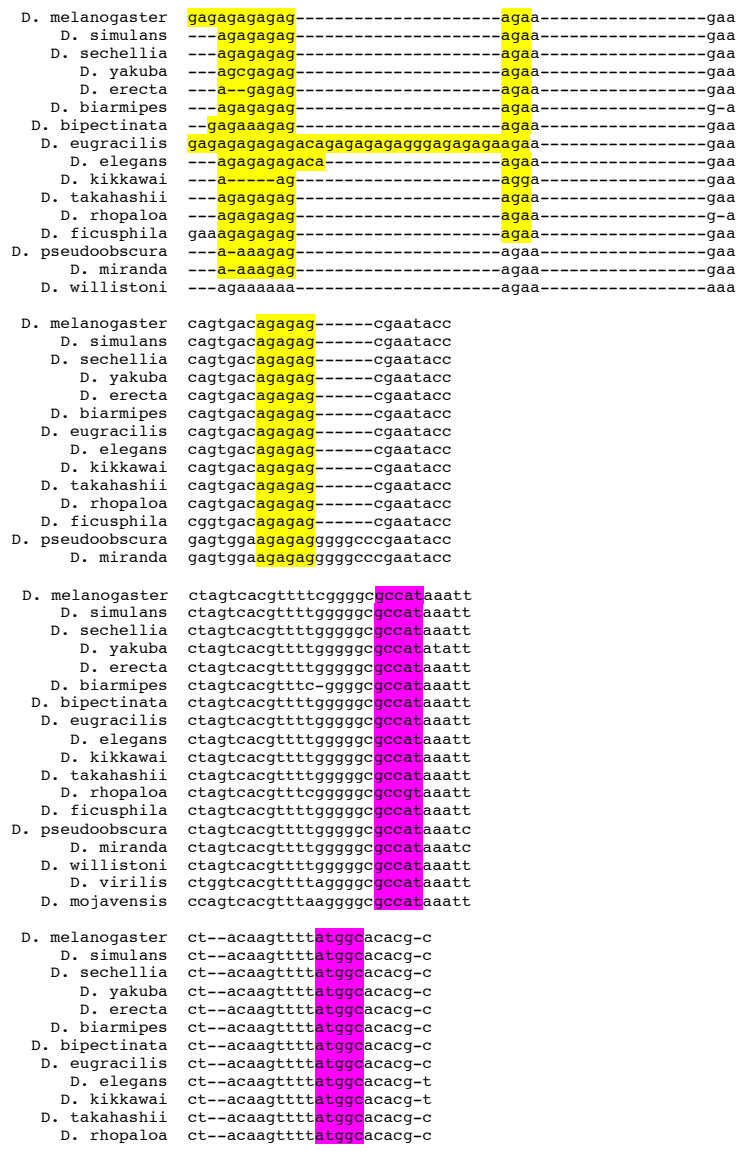

D. melanogaster ctgggga----------gagcg----------aaataaagctaaaatatgc
D. simulans ctgggga----------gagcg----------aaataaagctaaaatatgc
D. sechellia ctgggga----------gagcg----------aaataaagctaaaatatgc
D. yakuba cttggga----------gagcg----------aaataaagctaaaatatgc
D. erecta cttggga----------gagcg----------aaataaagctaaaatatgc
D. biarmipes cttgggagatcgg--------gagag----------gaataaagctgaaatatgc
D. bipectinata ctgggga----------gagcggctgatcgggagagcagtggataaagctaaaatatgc
D. eugracilis cttgggagatcggg------agagcg----------taataaagctaaaatatgc
D. elegans cttgggagatcggg------agagcg----------aaataaagctaaaatatgc
D. kikkawai catcggagagcgga------agagat----------ggataaagctaaaatatgc
D. takahashii cttgggagatcgg--------gagag----------gactaaagctaaaatatgc
D. rhopaloa cttgggagatcggg------agagcg----------gaataaagctaaaatatgc
D. ficusphila ctcgggagatcggg------agagcg----------gtataaagctaaaatatgc
D. pseudoobscura cttcagagagtaga--agagagagag--agagagagagagcaatggataaagctaaaatatgc
D. persimilis cttcagagagtagaagagagagagag--agagagagagagcaatggataaagctaaaatatgc
D. miranda cttcagagagtaga------agagag--agagagagagagcaatggataaagctaaaatatgc
D. melanogaster atgttggag----------aaaaaaatgccgcccatgtcgccaaaat
D. simulans atgttggag----------aaaaaaatgccgcccatgtcgccaaaat
D. sechellia atgttggag----------aaaaaaatgccgcccatgtcgccaaaat
D. yakuba atggtggag----------aaaaaaatgccgcccatgtcgccaaaaa
D. erecta atgttggag----------aaaaaaatgccgcccatgtcgccaaaat
D. biarmipes atgttggag----------aaaaaaatgccgcccatgtcgccaaaat
D. bipectinata acgttggagctgctggggcaacaacaacaacaacaacaacagaaaaaaaggccgcccatgtcgccaaaat
D. eugracilis atgttggag----------aaaaaaatgccgcccatgtcgccaaaat
D. elegans atgttggag----------aagaaaaaaatgccgcccatgtcgccaaaat
D. kikkawai aagttggag----------aaaaaaatgccgcccatgtcgccaaaaa
D. takahashii atgttggaa----------aaaaaaatgccgcccatgtcgccaaaat
D. rhopaloa atgttggag----------aagaaaaaatgccgcccatgtcgccaaaat
D. ficusphila atgttggag----------aaaaaaatgccgcccatgtcgccaaaat
D. pseudoobscura aagctgaag----------ctgctatggcaaaaaaatgccgcccatgtcgccaaaaa
D. persimilis aagctgaag----------ctgctatggcaaaaaaatgccgcccatgtcgccaaaaa
D. miranda aagctgaag----------ctgctatggcaaaaaaatgccgcccatgtcgctaaaaa
D. melanogaster tttagcatc---ggaacatgcaaaaacagacatca-tcgcatggg----------gcagc
D. simulans tttagcatc---ggaacatgcaaaaacagacatca-tcgcatggg----------gcagc
D. sechellia tttagcatc---ggaacatgcaaaaacagacatca-tcgcatggg----------gcagc
D. yakuba cttagcatc---ggaacatgcaaaaacagacatca-gcgcatggg----------gcagc
D. erecta cttagcatc---ggaacatgcaaaaacagacatca-gcgcatggg----------ccagc
D. biarmipes tttagcatcg--ggaacatgcaaaaacagacatca-tcgcatggg----------gcagc
D. bipectinata tttagcatcg--ggaacatgcaaaaacagacatca-gcacatggggcaaaaacaataacaggaaagcaac
D. eugracilis tttagcatcg--ggaacatgcaaaaacagacatca-tcgcatgag----------caacaac
D. elegans tttagcatcg--ggaacatgcaaaaacagacatca-gcgcatggg----------gcagc
D. kikkawai tttagcatcg--agaacatgcaaaaacagacatca-------aag----------gcagc
D. takahashii tttagcatcg--ggaacatgcaaaaacagacatca-tcgcatggg----------gcagc
D. rhopaloa tttagtatcg--ggaacatgcaaaaacagacatca-gcgcatgga----------gcagc
D. ficusphila tttagcatcg--ggaacatgcaaaaacaaacatca-gcgcattgg----------gctac
D. pseudoobscura ttttgtattgaaaaacatgcaaaaacagacatca-gc----gag----------gcag-
D. persimilis ttttgtattgaaaaacatgcaaaaacagacatca-gc----gag----------gcag-
D. miranda ttttgtattgaaaaacatgcaaaaacagacatca-gc----gag----------gcag-
D. willistoni tttagcatta--aaaacatgcaa----------
D. virilis tttagtatta--aaaacatgcaaaaacagacatcaagcccatgtg----------gcagc
D. mojavensis tttagtatta--aaaacatgcaaaaacagatatcagacgcatgtg----------gcagg


D. melanogaster
D. simulans
D. sechellia
D. yakuba
D. erecta
D. biarmipes
D. bipectinata
D. eugracilis
D. elegans
D. kikkawai
D. takahashii
D. rhopaloa
D. ficusphila
D. pseudoobscura
D. persimilis
D. miranda
D. willistoni
D. virilis
D. mojavensis

```
  D. melanogaster  caac---------aaaccagc-agc----ca-aggcacttt---gggtggtccatgcaaaa---------
      D. simulans  caac---------aaaccagc-agc----cg-aagcacttt---gggtggtccgtgcaaaa---------
     D. sechellia  caac---------aaactagc-agc----ca-aagcacttt---gggtggtccgtgcaaaa---------
        D. yakuba  caac---------aaaccagc-agc----ca-aagcacttt---gggtggtccgtgcaaaa---------
        D. erecta  caac---------aaaccagc-agc----ca-aagcacttt---gggtgggccgtgcaaaa---------
     D. biarmipes  caat-------agaaattagcgagctccgca-gtgaatgttggaggggagttagctcgata---------
   D. bipectinata  aaag---------gaaacaga-aac-------ggccgtttt---gggtggtctttgaaaat---------
    D. eugracilis  caacaactacaaaaaaacagc-agc----cagaagcactt---gggtggtct-tgcgaat---------
       D. elegans  catcaacaaca--aaaacagc-agc----caaaagcactt---gggtggtctcggcaaaa---------
      D. kikkawai  caacaatagca--aaagcggc-agc-----a-aaatacttc---g----------ccaaaaaccaaaatc
    D. takahashii  caac----acaaaaaagcagc-agc----cgaaagcactt---gggtggtctttgcgaaa---------
      D. rhopaloa  catcaacaaca--aaaacagc-agc----caaaagcactt---gggtggtctcgccaaaa---------
    D. ficusphila  caaaaaaa-----aaaacagc-agc----ca-aatctctt---aggtggtctcggcgaaa---------
D. pseudoobscura  -------------aaa------agc----ct-------tt---gggtggtcgtggcaaca---------
    D. persimilis  -------------aaa------agc----ct-------tt---gggtggtcgtggcaaca---------
       D. miranda  -------------aaa------agc----ct-------tt---gggtggtcgtggcaaca---------
    D. willistoni  ------------------gc-aac----ct-----cgctt---gggtggtc----caaac---------
       D. virilis  gaata--------acgccagc-ggg----cg-cag--tcgt---gtgtggtcttagctaaa---------
    D. mojavensis  caacaacaggaatacgcctgc-ggg----cg-caat-tcgt---gtgtggtctaagctaca---------
```

Enhancer 1EH

```
  D. melanogaster  gaaccc--aatggccaca------taactcc
      D. simulans  gaaccc--aatggccaca------taactcc
     D. sechellia  gaaccc--aatggccaca------taactcc
        D. yakuba  gagccc--aatggccaca------taactcc
        D. erecta  gaaccc--aatggccaca------taactcc
     D. biarmipes  gagccc--aatggccaca------taactcc
   D. bipectinata  gaaccc--aatggccacatacacgtaacacc
    D. eugracilis  gaaacc--aatggccaca------taactcc
       D. elegans  gacacc--aatggccaca------taactcc
      D. kikkawai  gagccccgaatggccacatata--taactcc
    D. takahashii  gaaccc--aatggccaca------taactcc
      D. rhopaloa  gatccc--agtggccaca------taactcc
    D. ficusphila  gagccc--aatggccaca------taactcc
D. pseudoobscura  ataccc--aatggccaca------taacacc
       D. miranda  ataccc--aatggccaca------taacacc
    D. willistoni  taaacc--tttgccccaca------taacacc
       D. virilis  ggagtc--agtgccccaaa------tagcacc

  D. melanogaster  acaac--------------------------gg------------cccgac------------------a
      D. simulans  acaac--------------------------gg------------cccgac------------------
     D. sechellia  acaac--------------------------gg------------cccgac------------------
        D. yakuba  acaac--------------------------gg------------ctcgac------------------
        D. erecta  acaac-------------------------ag------------cccgac------------------
     D. biarmipes  acaac--------------------------gg------------c-cgac------------------
   D. bipectinata  acgac--------------------------ga--------agaccccgaa------------------
    D. eugracilis  acaac--------------------------gg------------gccgacacataggGAGAGAGAGAGA
       D. elegans  acaac--------------------------gg------------cccaag------------------
      D. kikkawai  -------------------------------ga--------------gac------------------
    D. takahashii  acaac--------------------------gg------------cccgac------------------
      D. rhopaloa  acaac--------------------------gg------------cccGAG------------------
    D. ficusphila  acaac--------------------------gg------------cccgaa------------------
D. pseudoobscura  acagc--------------------------aa--------------caac------------------
       D. miranda  acagc--------------------------aa--------------caac------------------
    D. willistoni  acagc--------------------------agaaagaaAGAGAGAcccaa------------------
```


D. melanogaster gagagagagag----------agaa----------gaa
D. simulans ---agagagag----------agaa----------gaa
D. sechellia ---agagagag----------agaa----------gaa
D. yakuba ---agcgagag----------agaa----------gaa
D. erecta ---a--gagag----------agaa----------gaa
D. biarmipes ---agagagag----------agaa----------g-a
D. bipectinata --gagaaagag----------agaa----------gaa
D. eugracilis gagagagagagacagagagagagggagagagaaagaa----------gaa
D. elegans ---agagagagaca----------agaa----------gaa
D. kikkawai ---a-----ag----------agga----------gaa
D. takahashii ---agagagag----------agaa----------gaa
D. rhopaloa ---agagagag----------agaa----------g-a
D. ficusphila gaaagagagag----------agaa----------gaa
D. pseudoobscura ---a-aaagag----------agaa----------gaa
D. miranda ---a-aaagag----------agaa----------gaa
D. willistoni ---agaaaaaa----------agaa----------aaa
D. melanogaster cagtgacagagag------cgaatacc
D. simulans cagtgacagagag------cgaatacc
D. sechellia cagtgacagagag------cgaatacc
D. yakuba cagtgacagagag------cgaatacc
D. erecta cagtgacagagag------cgaatacc
D. biarmipes cagtgacagagag------cgaatacc
D. eugracilis cagtgacagagag------cgaatacc
D. elegans cagtgacagagag------cgaatacc
D. kikkawai cagtgacagagag------cgaatacc
D. takahashii cagtgacagagag------cgaatacc
D. rhopaloa cagtgacagagag------cgaatacc
D. ficusphila cggtgacagagag------cgaatacc
D. pseudoobscura gagtggaagagagggggcccgaatacc
D. miranda gagtggaagagagggggcccgaatacc
D. melanogaster ctagtcacgttttcggggcgccatataatt
D. simulans ctagtcacgttttgggggcgccataaatt
D. sechellia ctagtcacgttttgggggcgccataaatt
D. yakuba ctagtcacgttttgggggcgccatatatt
D. erecta ctagtcacgttttgggggcgccataaatt
D. biarmipes ctagtcacgtttc-ggggcgccataaatt
D. bipectinata ctagtcacgttttgggggcgccataaatt
D. eugracilis ctagtcacgttttgggggcgccataaatt
D. elegans ctagtcacgttttgggggcgccataaatt
D. kikkawai ctagtcacgttttgggggcgccataaatt
D. takahashii ctagtcacgttttgggggcgccataaatt
D. rhopaloa ctagtcacgtttcgggggcgccgtaaatt
D. ficusphila ctagtcacgttttgggggcgccataaatt
D. pseudoobscura ctagtcacgttttgggggcgccataaatc
D. miranda ctagtcacgttttgggggcgccataaatc
D. willistoni ctagtcacgttttgggggcgccataaatt
D. virilis ctggtcacgttttagggggcgccataaatt
D. mojavensis ccagtcacgtttaaggggcgccataaatt
D. melanogaster ct--acaagttttatggcacacg-c
D. simulans ct--acaagttttatggcacacg-c
D. sechellia ct--acaagttttatggcacacg-c
D. yakuba ct--acaagttttatggcacacg-c
D. erecta ct--acaagttttatggcacacg-c
D. biarmipes ct--acaagttttatggcacacg-c
D. bipectinata ct--acaagttttatggcacacg-c
D. eugracilis ct--acaagttttatggcacacg-c
D. elegans ct--acaagttttatggcacacg-t
D. kikkawai ct--acaagttttatggcacacg-t
D. takahashii ct--acaagttttatggcacacg-c
D. rhopaloa ct--acaagttttatggcacacg-c

```
   D. ficusphila  ct--acaagttttatggcacacg-c
D. pseudoobscura  ct--acaagttttatggcacacg-t
      D. miranda  ct--acaagttttatggcacacg-t
   D. willistoni  ctacacaggttttatggcacacgtc
      D. virilis  ct--acacgtttaatggcacaca-t
   D. mojavensis  ct--acacgttttatggcacact-t


 D. melanogaster  tcgtctgggtggtctgccatataga----------
     D. simulans  tcgtctgggtggtctgccatattga----------
    D. sechellia  tcgtctgggtggtctgccatattga----------
       D. yakuba  tcgtctgggtggtctgccatattga----------
       D. erecta  tcgtctgggtggtctgccatattga----------
    D. biarmipes  gtgtctgggtggtctgccatattga----------
  D. bipectinata  gcctctgggtggtctgccatattga----------
   D. eugracilis  atatctgggtggtctgccatattga----------
      D. elegans  ggatct-ggtggtctgtcatattga----------
   D. takahashii  gtgtctgggtggtctgccatattga----------
     D. rhopaloa  ggatctgggtggtctgccatattga----------
   D. ficusphila  ---tctgggtggtctgccatattga----------
D. pseudoobscura  gccttgggtggtctgccatattga----------
      D. miranda  gccttgggtggtctgccatattga----------
      D. virilis  tcggctgggtggtccgtcacgtcgc----------
   D. mojavensis  ttcactgggtggtccggcacgttgctgcttattga

 D. melanogaster  cggcataactgggtgcgctcatatt------taatggcc
     D. simulans  cggcataactgggtgcgctcatatt------ttatggcc
    D. sechellia  cggcataactgggtgcgctcatatt------ttatggcc
       D. yakuba  cggcataactgggtgcgctcatatt------ttatggcc
       D. erecta  cggcacaactgggtgcgctcatatt------ttatggcc
    D. biarmipes  cggcacaactgggtgcgctcatact------ttatggcc
  D. bipectinata  cggcataactgggtgcgctcatatt------ttatggcc
   D. eugracilis  cggcataactgggtgcgctcatatt------ttatggcc
      D. elegans  cggcataactgggtgcgctcgtatt------ttatggcc
     D. kikkawai  cggcataactgggtgcgctcatattttatggttatggcc
   D. takahashii  cggcataactgggtgcgctcatatt------ttatggcc
     D. rhopaloa  cggcataactgggtgcgctcatgtt------ttatggcc
   D. ficusphila  cggcataactgggtgcgctcatatt------ttatggac
D. pseudoobscura  cggcataactgggtgcgctcatatt------ttatggct
      D. miranda  cggcataactgggtgcgctcatatt------ttatggct

 D. melanogaster  agc-ggt--------------ccgtgtaattaagtagtcagg-cgctgtgggggcc
     D. simulans  agc-ggt--------------ccgtgtaattaagtagtcagt-cgctgtgggggcc
    D. sechellia  agc-ggt--------------ccgtgtaattaagtagtcagg-cgctgtgtgggcc
       D. yakuba  agc-ggt--------------ccgtgtaattaagtagtcagg-cgctgtgggggcc
       D. erecta  agc-ggt--------------ccgtgtaattaagtagtcggg-cgctctgggggcc
    D. biarmipes  agc-ggg--------------ccgagtaattaagtagtcaag-cgctgtgggggcc
  D. bipectinata  atc-ggg--------------tcgtgtaattaagtagtcatt-ctctgtgggggtt
   D. eugracilis  agc-ggt--------------ccgtgtaattaagtagtcaag-cgctgtgggggcc
      D. elegans  agc-ggc--------------ccgtgtaattaagtagtcaag-cgctgtgggggcc
     D. kikkawai  agc-gga--------------ccgtgtaattaagtagtcagatctctgtgggggta
   D. takahashii  agc-ggg--------------ccgtgtaattaagtagtcaag-cgctgtgggggcc
     D. rhopaloa  ----ggc--------------ccgtgtaattaagtagtcaag-cgctgtgggggcc
   D. ficusphila  agc-ggt--------------ccgtgtaattaagtagtcaac-caccgtgggggcc
D. pseudoobscura  atcgggg--------------ccgtgtaattaagtagtcaag-ctctgtgggggcc
      D. miranda  atcgggg--------------ccgtgtaattaagtagtcaag-ctctgtgggggcc
      D. virilis  ----------------------ccgtgtaattaagtagtcaaa-gcccgtgggggtg
   D. mojavensis  gtc-gactatttataatgcttcggtgtaattaagtagtcaaa-gcccgtgaggcatt

 D. melanogaster  gtcgga-tcttttgattgattgatgcgtg
     D. simulans  gtcgga-tcttttgattgattgatgcgcg
    D. sechellia  gtcgga-tcttttgattgattgatgcgtg
       D. yakuba  gtc-gg-tcttttgattgattgatgcgtg
       D. erecta  gtcggg-tcttttgattgattgatgcgtg
    D. biarmipes  gtcgga-tcttttgattgattgatgcgtg
  D. bipectinata  gtcgga-ta-tatgattgattgatgggga
   D. eugracilis  gtcgga-tcttttgattgattgatgcgtg
      D. elegans  gtcgga-ccttttgattgattgatgcgtg
     D. kikkawai  gtcgga-tc-tttgattgattgatgcgtg
   D. takahashii  gtcgga-tcttttgattgattgatgcgtg
     D. rhopaloa  gtcgga-tcttttgattgattgatgcgtg
```

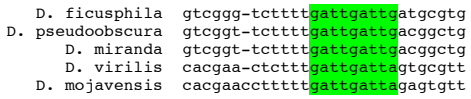


Figure 2.16 Sequence conservation of selected TF binding motifs in enhancers of *patched*

Sequence alignments were obtained from the UCSC Genome Browser, dm6 build. Selected conserved Tango/Spineless (Tgo/Ss) motifs are in gray; Distaless (Dll) in light blue; Engrailed (En) in green; Pannier (Pnr/dGATAe) in purple; Scalloped (Sd) in blue; Cubitus interruptus (Ci) in red; Odd-Skipped (Odd) in dark red; Forkhead (Fkh) in dark yellow; Sloppy-paired (Slp) in blue-green; Pleiohomeotic (Pho) in pink, GAGA Factor (GAF/Trl) in yellow. Blocks of sequence separated by line breaks are not necessarily contiguous.

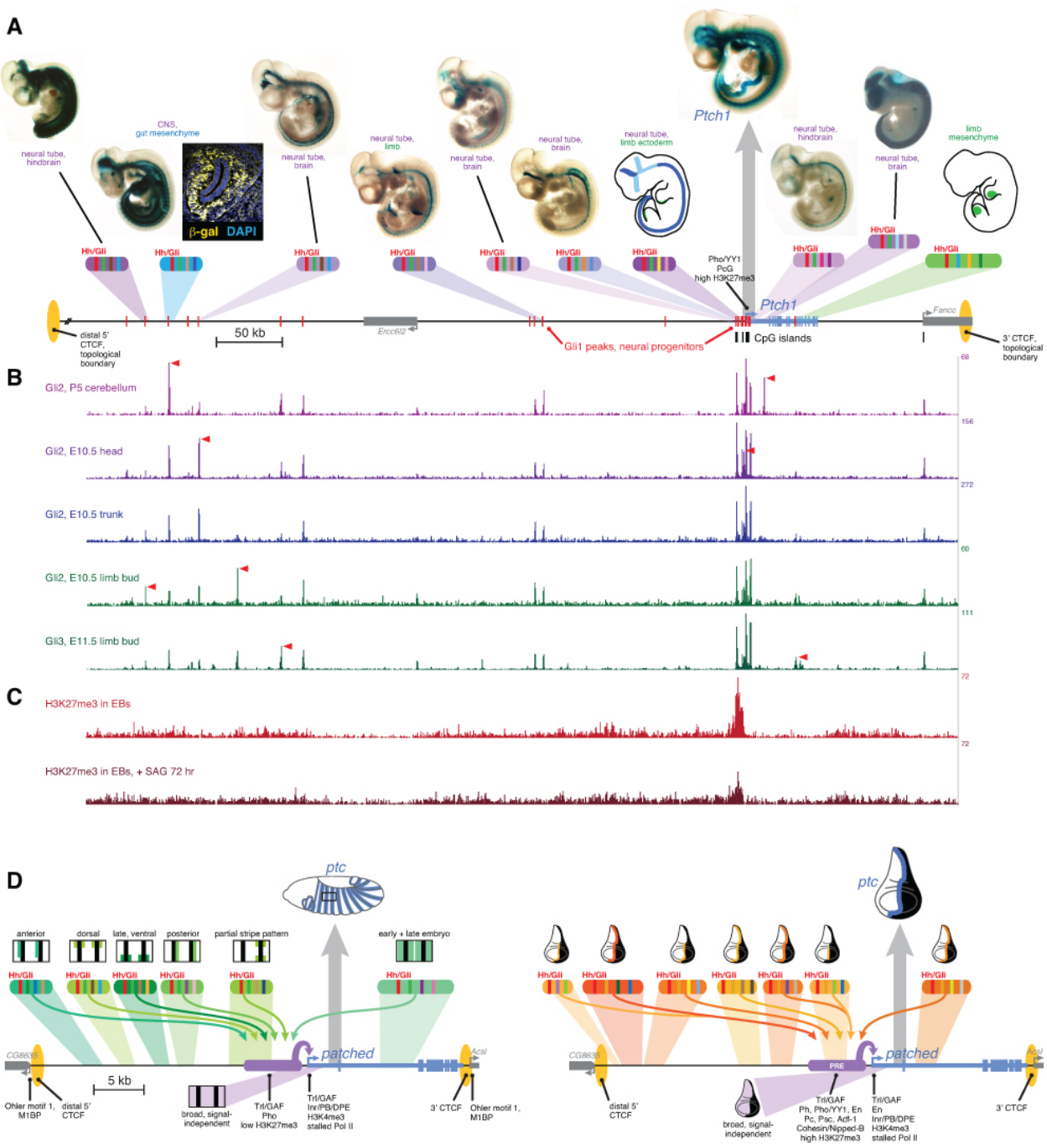


Figure 2.17 *Drosophila* and vertebrate *patched* orthologs share a similar *cis*-regulatory architecture and gene regulatory strategy

(A) X-gal-stained E10.5 mouse embryos carrying novel *Ptch1* enhancers as transient-transgenic lacZ reporters (Figure 2.22) are shown above a map of the mouse *Ptch1* region. Known *Ptch1* enhancers (Lopez-Rios et al., 2014; Vokes et al., 2007) are shown as cartoons. Red boxes along the gene locus schematic represent GLI1 occupancy in neural progenitors near *Ptch1* (as identified by ChIP-seq; (Peterson et al., 2012). (B) ChIP-seq of GLI2 and GLI3 near *Ptch1*. Red arrowheads highlight tissue-specific GLI

binding. (C) ChIP-seq shows H3K27me3 is enriched at the proximal promoter but reduced upon pathway stimulation, consistent with relief of Polycomb-mediated repression. (D) Summary of the *ptc* response to Hh in the *Drosophila* embryonic ectoderm. Multiple Hh-responsive enhancers combine Ci/GLI sites (red bands) and tissue-specific inputs (colored bands), generating overlapping subsets of the embryonic *ptc* response which are integrated and amplified by the promoter-proximal region, resulting in a complete segment-polarity stripe pattern (top). The promoter-proximal region also generates the basal signal-independent *ptc* expression (bottom, purple). (E) Summary of the Hh response of *ptc* in the wing disc. A large battery of tissue-specific GLI-responsive enhancers, driving wing stripes of different widths, intensities, and dorso-ventral extents, interact with a Polycomb/Trithorax-bound promoter-proximal region to provide a complete wing stripe pattern (top). The PRE-containing promoter-proximal region also produces signal-independent *ptc* expression in the anterior (bottom, purple). Identification of flanking CTCF insulator sites for both fly and mouse *patched* were determined using ChIP-seq and ChIP-CHIP analyses available from the UCSC genome browser and modENCODE databases (Figure 2.18).

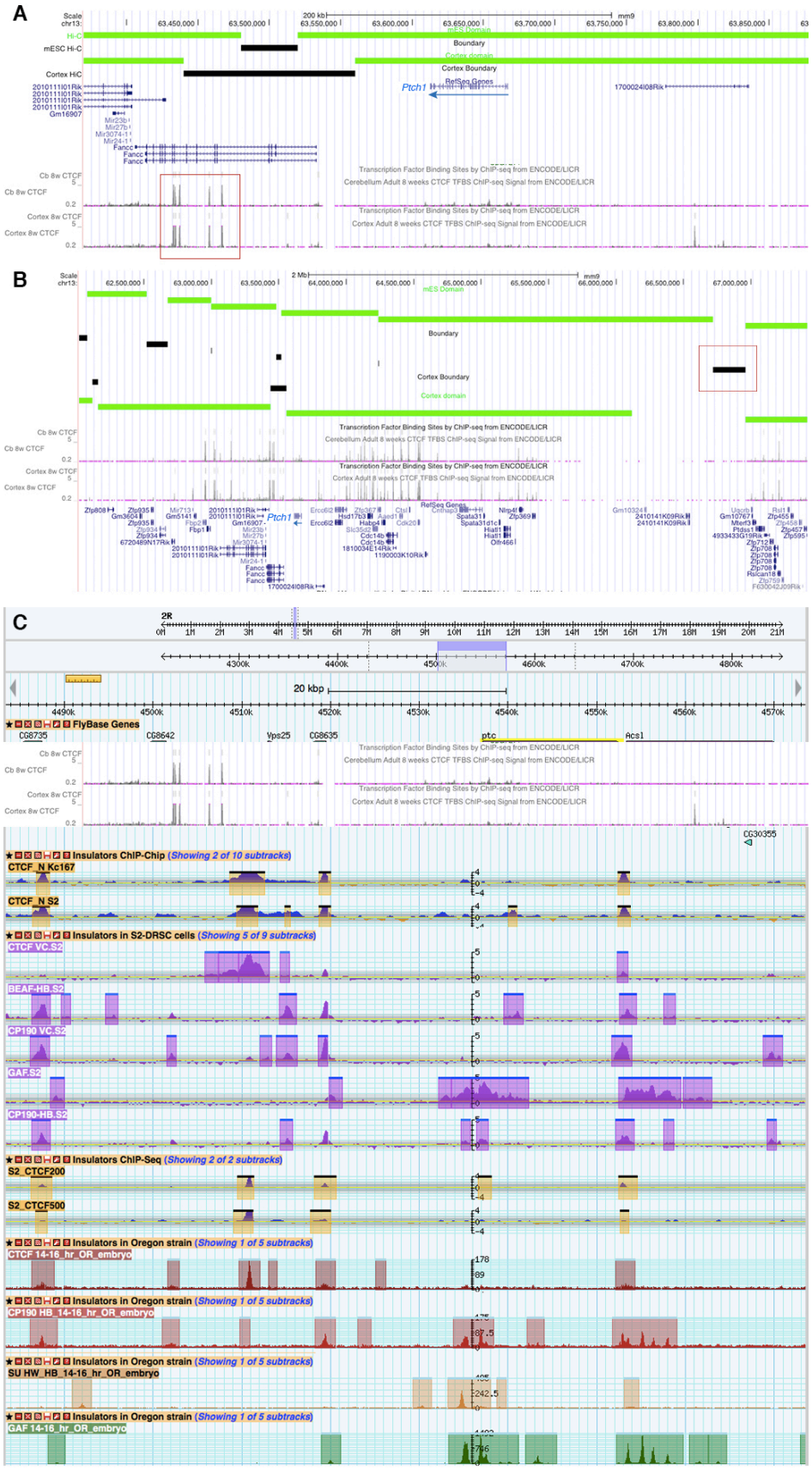

A
B
C
Ptch1
Boundary
Cortex Boundary
Cortex domain
mES Domain
Transcription Factor Binding Sites by ChIP-seq from ENCODE/LICR
Cerebellum Adult 8 weeks CTCF TFBS ChIP-seq Signal from ENCODE/LICR
Cortex Adult 8 weeks CTCF TFBS ChIP-seq Signal from ENCODE/LICR
FlyBase Genes
Insulators ChIP-Chip (Showing 2 of 10 subtracks)
CTCF_N_Kc167
CTCF_N_S2
Insulators in S2-DRSC cells (Showing 5 of 9 subtracks)
CTCF_VC_S2
BEAF-HB_S2
CP190_VC_S2
GAF_S2
CP190-HB_S2
Insulators ChIP-Seq (Showing 2 of 2 subtracks)
S2_CTCF200
S2_CTCF500
Insulators in Oregon strain (Showing 1 of 5 subtracks)
CTCF_14-16_hr_OR_embryo
CP190_HB_14-16_hr_OR_embryo
SU_HW_HB_14-16_hr_OR_embryo
GAF_14-16_hr_OR_embryo

Figure 2.18 Topological boundaries and chromatin insulators at mouse *PTCH1* and fly *patched*

(A) Map of the immediate vicinity of mouse *Ptch1*, displayed in the UCSC Genome Browser (genome.ucsc.edu), showing the topological domains and boundary regions in mouse ES cells and cortex, as determined by Hi-C chromatin interactions (Dixon et al., 2012), and ChIP-seq binding data for the chromatin insulator/boundary factor CTCF in embryonic mouse brain(Wang et al., 2012). Red box indicates a cluster of CTCF peaks in the Fancc gene, downstream of *Ptch1*. (B) Larger-scale view (~5.5 Mb) of a chromosomal region including *Ptch1*. Red box shows a boundary region far upstream of *Ptch1*. The topological domain containing *Ptch1* extends from Fancc over 2 Mb to the 5' of *Ptch1* (to the right in this view). (C) Summary of modENCODE *Drosophila* genomic data (Celniker et al., 2009) in the vicinity of the *ptc* locus (~90 kb), showing binding of CTCF and other insulators and boundary-associated factors. As in the mouse, the nearest 3' CTCF boundary to fly *ptc* is nearby (between the 3' end of *ptc* and the promoter of the adjacent gene Ascl), while the nearest 5' boundary is further away (~18 kb upstream, at the CG8635 gene promoter). All characterized *ptc* enhancers fall between these two CTCF sites.

Primer	Sequence	Restriction Site
A_F	CACCctcgagCTCGCCTAATGAAGTTGTTGGC	XhoI
A_R	**GTAaccggt**GTTCTGTGATATCTATCTTGT	AgeI
B_F	CACCctcgagAACAAACTTGGTATTTTCTCGG	XhoI
B_R	**GTActcgag**CGCGCGCCGCTGTTTCTC	XhoI
GC_F	CACCGGCGCGCCCTATGGGCCAATGACAAATG	AscI
GC_R	**GTACCGCGG**GGGTGATTCAGGAGTTTTTCC	SacII
HF_F	CACCGGCGCGCCTCCCACTTCATAACCCTC	AscI
HF_R	**GTACCGCGG**CGCTCTCTCTTTCGGGGAGAA	SacII
JH_F	CACCGGCGCGCCTACGTACTCTTATTACTCCACTC	AscI
JH_R	**GTACCGCGG**GCTATTGCATTTGTCATTGGC	SacII
LJ_F	CACCGGCGCGCCCTACTTGGTTTGATAAAT	AscI
LK_R	**GTACCGCGG**ACGGTGTGTGTGAACCCAACTAATTG	SacII
LJ_R	**GTACCGCGG**ATAAGTACAGTGCTGGTCATA	SacII
ML_F	CACCGGCGCGCCCGAGGCGAGATGGCTTCG	AscI
ML_R	**GTACCGCGG**GCAGCGACTGCTGAGCCAGTC	SacII
NM_F	CACCGGCGCGCCTACAATATCTATTATCTA	AscI
NM_R	**GTACCGCGG**AGTTCCCATTCAGCTTTGACA	SacII
PN_F	CACCGGCGCGCCCGCCAAAATGAAAATTATTACAAAG	AscI
PN_R	**GTACCGCGG**CGTCCCTCTGTCGTCGCTGGG	SacII
RP_F	CACCGGCGCGCCAGCATTAACAGCCGAAGC	AscI
RP_R	**GTACCGCGG**ACCGAGCCGTACAAATATAAC	SacII
VR_F	CACCGGCGCGCCACAGACGGGGCTACACTGAG	AscI
VR_R	**GTACCGCGG**ATTCGGAGACATTCGCAGAC	SacII
VT_R	**GTACCGCGG**GACAGACCCCGAAGACTGAG	SacII
YU_F	CACCGGCGCGCCGCCCTGTCGTCTTTGTCTTC	AscI
YX_R	**GTACCGCGG**GTATTAATAGTGGGAGCTCTG	SacII
YV_R	**GTACCGCGG**AAATTTGCATAATAATAAGTC	SacII
WV_F	CACCGGCGCGCCTACAATTTTGGAAAAGG	AscI
U_F	CACCGGCGCGCCTCGCTGCGTGTTGCTGTG	AscI
YU_R	**GTACCGCGG**AACCATCAAACCCACGAAAC	SacII
ZY_F	CACCGGCGCGCCCCGGATCGACCTAGGTAAGG	AscI
ZY_R	**GTACCGCGG**TTCTTCTGGGTGTGTGCTTG	SacII
VT_F	CACCGGCGCGCCCCGCATACCCTATGATATGACTC	AscI
DB_F	CACCGGCGCGCCATGCATGCGCAGCCTGCCAC	AscI
DB_R	**GTACCGCGG**CGCGCGCCGCTGTTTCTC	SacII
1AC_F	CACCGGCGCGCCGGTGAGTGCCCAACTACAG	AscI
1AC_R	**GTACCGCGG**GTTGGGCGCCATATGTTTAC	SacII
1BC_F	CACCGGCGCGCCTCCGTCCAGCAGTCGAAAG	AscI
1BC_R	**GTACCGCGG**CCCATTCGATATACCCTCAAG	SacII
1CE_F	CACCGGCGCGCCTCGGTTGTAACGTGCTTTTG	AscI
1CE_R	**GTACCGCGG**TGCAGATGGCAGATCAAGTC	SacII
1EH_F	CACCGGCGCGCCGAAGTGCTTAACAAGTTAAC	AscI
1EH_R	**GTACCGCGG**CACGACAACCAATGAGATCG	SacII
1GH_F	CACCGGCGCGCCGGTAATAAATGCGGCAGACG	AscI
1HI_F	CACCGGCGCGCCGATTGATTGATGCGTGATGC	AscI
1HI_R	**GTACCGCGG**CTGAAAAATGCAACAAAATA	SacII

Figure 2.19 PCR primers used to amplify *Drosophila* genomic sequences

The promoter analysis primers A_F and B_F contain a 5' tag of CACCctcgag – CACC is used to facilitate TOPO cloning (Life Technologies; see Methods) for ease of sequencing, and a XhoI site (ctcgag) for cloning purposes. The A_R primer contains a

5’ tag of GTAaccggt. The GTA serves as a buffer for ease of restriction digest and an AgeI restriction site (accggt) for cloning purposes. The B_R primer contains a 5’ tag of GTActcgag. The GTA serves as a buffer for ease of restriction digest and a XhoI restriction site (ctcgag) for bidirectional cloning purposes. All other primers were designed for TOPO cloning of enhancer sequences: Each enhancer’s forward primer contains a 5’ tag of CACCCCGCGG. CACC again was used to facilitate TOPO cloning while CCGCGG is a SacII restriction site for screening purposes. Each Reverse primer contains a 5’ tag of GTACCGCGG, the GTA again serves as a buffer for ease of restriction digest and a SacII restriction site (CCGCGG) was used for cloning purposes.

Primer	Sequence
LK_93	AATCGTGCTC**TGCCACCCA**CTCACACTTG
LK_93ko	AATCGTGCTC**TGCAAAACA**CTCACACTTG
LK_47	AAGTGGACGA**GGCCACCAA**ATGACAACAC
LK_47ko	AAGTGGACGA**GACAAAAAA**ATGACAACAC
VT_93	AATGGCATGT**TGCCACCCA**AGTTGCCAGT
VT_93ko	AATGGCATGT**TGCAAAACA**AGTTGCCAGT
VT_93hi	AATGGCATGT**GACCACCCA**AGTTGCCAGT
VT_141	AGTTGCCAGT**TGGCTGGCC**AAGACTTAAT
VT_141ko	AGTTGCCAGT**TGTTTTGCC**AAGACTTAAT
VT_141hi	AGTTGCCAGT**TGGGTGGTC**AAGACTTAAT
1EH_82	GCGGCATAAC**TGGGTGCGC**TCATATTTAA
1EH_82ko	GCGGCATAAC**TGTTTTCGC**TCATATTTAA
1EH_82hi	GCGGCATAAC**TGGGTGGTC**TCATATTTAA
1EH_22	TCAGGCGCTG**TGGGGGGCC**GGAGAAAATT
1EH_22ko	TCAGGCGCTG**TGTTGTGCC**GGAGAAAATT
1EH_22hi	TCAGGCGCTG**TGGGTGGTC**GGAGAAAATT
DB_1	TTGGGTAGGG**GACCACCCA**CATCGCTTGG
DB_1KO	TTGGGTAGGG**GACAAAACA**CATCGCTTGG
DB_2	CTGCAGTGTG**GACCACCCA**AAAATAAATT
DB_3	TTTTTGCATG**GACCACCCA**AAGTGCCTTG
ZY_1	TCGAATTAAG**GGCCACCCA**CGAAGAGACA
ZY_1KO	TCGAATTAAG**GGCAAAACA**CGAAGAGACA
ZY_2	GGCTAATATG**GACCACCCA**AGAAGATGAC
ZY_2KO	GGCTAATATG**GACAAAACA**AGAAGATGAC
ZY_3	CACACACGTC**GATCGCCCA**GGAGCGCTGG
ZY_3KO	CACACACGTC**GATAGAACA**GGAGCGCTGG
ZY_4	GGATGTGGAG**GGCCACACA**CGTCGATCGC
ZY_4KO	GGATGTGGAG**GGCAAAACA**CGTCGATCGC

Figure 2.20 EMSA oligonucleotides used in GBS competition binding assays

The top strand of each double stranded oligonucleotide is shown. The GBS being tested is marked in bold.

Enhancer location (relative to Ptch1 promoter)	Coordinates (mm9)	Source
-456kb	chr13:64122052-64122757	this study
-438kb	chr13:64104376-64104670	this study
-415kb	chr13:64081569-64081954	this study
-154kb	chr13:63820845-63821332	this study
-8.1kb	chr13:63674651-63674885	this study
-3.7kb	chr13:63670185-63671020	this study
-0.95kb	chr13:63667602-63667781	Vokes et al., 2007
+2.6kb	chr13:63664151-63664519	this study
+11kb	chr13:63655468-63657042	Visel et al., 2007
+45kb	chr13:63621577-63634154	Lopez-Rios et al., 2014

Figure 2.21 Mouse *Ptch1* enhancer coordinates

Sequence coordinates, corresponding to the mm9 build of the Mus musculus genome, are provided for all enhancers tested as lacZ reporters in E10.5 mouse embryos in Figure 6A. Enhancers are named for their position relative to the *Ptch1* promoter. References are provided for previously published enhancers.

Chapter 3

Non-GLI Transcription Factors Regulating *patched*

3.1 Abstract

The constitutive response of *ptc* to Hh signaling is regulated by an array of context specific enhancers controlled by the GLI family of TFs in both flies and mice. It is important to remember that no gene or enhancer is regulated by a single TF, which implies that there is still much to learn about other factors acting on the enhancers already discussed. Multiple tissue-specific TFs contribute to the regulation of these enhancers in different developmental and adult contexts, for an example of a very well characterized enhancer requiring multiple inputs, see (Swanson et al., 2010). Since only a handful of Hh-responsive enhancers have been identified (Ramos and Barolo, 2013), there has not been a major, comprehensive effort to identify the TFs that regulate Hh-responsive enhancers, aside from GLI. After identifying more than 30 context specific, GLI-responsive enhancers (Chapter 2), I have begun to identify and characterize tissue-specific inputs that also regulate *ptc.* In this chapter, I have assessed several novel transcriptional inputs predicted to contribute to *ptc* regulation.

3.2 Introduction

Although many components of the Hh signaling pathway were originally identified in flies (Ingham and McMahon, 2001; Ingham et al., 1991; Nüsslein-Volhard and Wieschaus, 1980), most of what we know about additional tissue specific transcription

factors controlling Hh target genes have been identified in mice (Hoffmann et al., 2014; Oosterveen et al., 2012; Peterson et al., 2012; Ribes et al., 2010). Here, I have compiled most of what is known about these additional non-GLI inputs identified by other groups and complete new experiments to address novel tissue specific TFs required for Hh-responsive enhancer function in flies.

One of the first and best characterized transcription factors to be identified as a context-specific input into Hh target gene enhancers is SOX2, a key determinant of neural progenitor identity (Peterson et al., 2012; Pevny and Placzek, 2005). The experiments leading to this discovery were carried out by tagging murine GLI1 with a FLAG epitope at the C-terminus (Vokes et al., 2007) and using ChIP-seq in differentiated neural progenitors to identify regions of the genome occupied by GLI. Computational analysis of these GLI-bound regions demonstrated that the most significantly enriched binding site was that of GLI, as expected. Interestingly, there were many other motifs identified that were shared and overrepresented among the newly identified GLI-binding regions. Of these, one of the top non-GLI predicted factors was SOX2. Peterson and colleagues then performed SOX2 ChIP-seq in the same cells to further test their hypothesis that SOX2 also regulates GLI-responsive enhancers and showed that SOX2 was bound to about 12% of the regions also occupied by GLI. Furthermore, these binding regions aligned with active enhancer marks like H3K27Ac upon SHH stimulation, suggesting that these regions were most likely active enhancers bound by both GLI and SOX2. A subset of these co-bound enhancers were tested as transient transgenic reporters with all predicted SOX2 binding sites mutated (similar to how I mutated the GBS in Chapter 2). Each enhancer relied on SOX2 binding sites for

activity in this assay, further demonstrating that SOX2 activates Hh-responsive enhancers.

While SOX2 is one of the best characterized non-GLI inputs into Hh regulated enhancers, the ChIP-seq analysis of GLI1 targets by Peterson et al. also predicted other context specific inputs based on sequence alone. The top candidates included FOXA2, PBX, EVX, and RFX, although they have not yet been confirmed with transgenic reporters. FOXA2 was of particular interest because independent work published concurrently also identified FOXA2 as a potential co-regulator of SHH target enhancers (Metzakopian et al., 2012). These experiments set out to identify all targets of FOXA2 in the neural tube and one of their main findings was that many of the FOXA2 ChIP-seq peaks overlapped known GLI-bound peaks, several of which had already been tested as SHH regulated enhancers in reporter assays. These included enhancers of the *Ptch1*, *GLI1*, *Nkx2.9*, and *HHip* genes (Agren et al., 2004; Dai et al., 1999; Hallikas et al., 2006; Peterson et al., 2012; Santagati et al., 2003; Sasaki et al., 1997; Vokes et al., 2007). While this group made a convincing argument for the co-regulation of these enhancers by FOXA2 and GLI, they did not actually make transgenic animals following up on the role of FOXA2 in these contexts. Therefore, while there is much evidence implying that FOXA2 is a co-regulator, it has not been directly confirmed.

In addition to these datasets, Oosterveen et al., completed another study that identified GLI regulated enhancers for SHH targets in the neural tube (Oosterveen et al., 2012). Their analysis examined sequence conservation and the presence of GLI binding sites and then tested predicted enhancers as minimal *lacZ* reporter constructs in the chicken neural tube. Using this experimental set up, they identified enhancers for the

Nkx2.2, Nkx2.9, Olig2, Nkx6.1, Nkx6.2, Dbx2, Dbx1, and *Pax6* genes. They also observed that many of these enhancers correlated to regions occupied by SOX proteins by ChIP analysis in neural tissue, where these enhancers are active. SOX proteins are a class of bHLH transcriptional activators known to regulate neural tube development, so it was plausible that SOX regulated these enhancers (Bergsland et al., 2011). To confirm this, binding sites for SOXB1 were mutagenized resulting in a loss of activation in the neural tube. This result was repeated in murine cell lines using luciferase assays to measure transcriptional activity as well. Together, this group identified several SHH responsive enhancers in the neural tube that depend on GLI binding sites of various affinities (see Chapters 1 and 2 for a more thorough discussion about the importance of low affinity binding sites) and also identified SOXB1 as a non-GLI input that is essential for mediating the SHH response in the neural tube.

Limited progress has been made in identifying co-regulators of SHH responsive enhancers outside of the neural progenitor cell types discussed so far. However, in the heart, another tissue that requires SHH signaling for normal development, at least one additional factor called TBX5, a T-box transcription factor, has been identified (Goddeeris et al., 2008; Hoffmann et al., 2009). While studying the role of TBX5 in heart development, Xie and colleagues demonstrated via in situ hybridization that both TBX5 and GLI1 transcription factors are expressed in the same tissue at the same time. Inhibiting expression of either factor with a dominant negative or a genetic deletion led to impaired development (Xie et al., 2012). With these data, Hoffman and colleagues set out to identify the mechanism underlying how these two factors might work together and went on to identify a single region of overlap between GLI and TBX5 ChIP-seq in

heart tissue, suggesting the presence of an enhancer occupied by both TFs in this context. Transient transgenic *lacZ* reporter mice carrying either an enhancer with wild type sequence or with mutated GLI or TBX5 binding sites demonstrated a requirement for both GLI and TBX5 to activate expression in the developing heart, suggesting that TBX5 is another tissue-specific input into SHH responsive enhancers.

While a handful of context-specific inputs have been shown to regulate SHH responsive enhancers, the list is still short, partly because of the lack validated GLI dependent enhancers. While there are many GLI bound regions identified by ChIP-seq, the overwhelming majority have not been tested for functionality, and at this time are still predicted enhancers. Because I now have identified and characterized more than 30 tissue specific enhancers of *Drosophila ptc*, I have the unique opportunity to identify other inputs into these regions, and therefore, into *ptc*. I began this analysis by combing through the sequences of the enhancers identified in Chapter 2 (see Figure 2.3 for complete list) using several criteria (Figure 3.1). Initially, I looked for large blocks of well-conserved sequences among related species of *Drosophila*. This was carried out in two ways: First, I used the UCSC genome browser to identify conservation within *Drosophilids*. Second, I used a freely available online comparative genomics tool called EvoprinterHD (https://evoprinter.ninds.nih.gov/evoprintprogramHD/drosophila.html) that aligns multiple genomes simultaneously to search for highly conserved orthologous sequences that might be of regulatory importance (Yavatkar et al., 2008) (Figure 3.1). After identifying conserved blocks of sequence within each enhancer, I searched for predicted transcription factor binding sites within these shorter blocks using several different computer programs. These included algorithms that searched entire databases

worth of TFBS positional weight matrices and predicted binding sites, such as CisBP (Weirauch et al., 2014), WeederWeb (Pavesi et al., 2004), Fly Factor Survey (http://mccb.umassmed.edu/ffs/DownloadData.php), or the JASPAR transcription factor binding database (Mathelier et al., 2016), http://jaspar.genereg.net/). While these programs use different algorithms, they all examine a particular sequence and identify predicted binding motifs that were generated from previously completed bacterial 1-hybrid screens, yeast 1 hybrid screens, and any additional published literature or datasets including ChIP-seq motif prediction. These programs predicted several binding sites for different TFs. I then examined any available expression data for that TF either by conducting a literature search or by using the Berkeley *Drosophila* Genome Project (Stapleton et al., 2002) http://www.fruitfly.org/index.html), which provides in situ expression data in an array of differently staged embryos for many *Drosophila* genes. These expression data were critical in making an assessment of whether or not to proceed. Even if there was a perfect binding site for a transcription factor using any of the prediction methods discussed above, if that TF is not expressed in that tissue, it cannot be regulating expression. Furthermore, I examined ChIP-seq datasets for the predicted transcription factors that had been published previously to see if any binding occurred in the enhancer of interest. With this information, I then used PCR mutagenesis to ablate these predicted binding and then made transgenic animals to visualize any change in expression in that tissue.

Although this work is still in progress, I have identified multiple novel TF-inputs into these enhancers. First, I describe Pannier, a *Drosophila* GATA TF, which provides activation in a subset of *ptc* positive cells (Herranz and Morata, 2001). I also have

strong evidence that 1EH requires GAGA input (first discussed in Chapter 2). Interestingly, there are at least two enhancers that were predicted to require additional transcriptional inputs from the repressors Sloppy Paired 1 (Slp1) and Odd-Skipped (Odd) using these criteria, but neither had a large effect on expression in our assay. I will discuss alternative explanations for the tissue specific responses of these enhancers as well. Finally, I have demonstrated a requirement for RFX (Regulatory Factor X) as an additional input into a previously characterized mouse *Ptch1* enhancer and suggest a role for RFX in flies as well (Vokes et al., 2007). Importantly, there are RFX transcription factors in *Drosophila* that mainly regulate ciliogenesis in sensory neuron differentiation (Dubruille, 2002) and recently Hh signaling has been shown to be active in *Drosophila* cilia, although a direct mechanism of target gene activation has not yet been identified (Kuzhandaivel et al., 2014). The reporter data I present here suggests a novel link between mouse and fly *patched* regulation and expands our knowledge of how Hh target enhancers combine multiple inputs to respond to signaling in a variety of different contexts.

3.3 Results

3.3.1 GATA: a novel regulator of *Drosophila ptc*

The *ptc* enhancers identified in Chapter 2 did not recapitulate a complete response to Hh signaling. Many of these enhancers were highly cell type specific, not only restricted to a developmental stage like the embryonic ectoderm, but spacially restricted within that particular tissue. PN, for example, was active only at the dorsal most cells of the embryo and this expression was not mediated by the predicted GBS in

this enhancer (Figure 3.2A-C). I reasoned that this highly restricted patterning might be the result of non-GLI tissue specific inputs that regulate expression in a subset of cells.

I then used the methods described in the introduction (Figure 3.1) to identify a set of highly conserved Pannier transcription factor binding sites near the 5' end of enhancer PN (Figure 3.2D-E). Pannier (Pnr or dGATAe), a *Drosophila* GATA zinc finger transcription factor, is known to be a critical regulator of embryonic development (Calleja et al., 2000). It is critical for the embryonic process called dorsal closure, which is one of the last major morphogenetic changes during embryogenesis and is responsible for closing the cellular hole left over after germ band retraction (Herranz and Morata, 2001; Jacinto et al., 2002). Since these newly identified dGATAe binding sites in this enhancer are conserved (Figure 3.2D) and the expression of this TF matches the expression patterns generated by the PN enhancer (Figure 3.2B,F), I used PCR mutagenesis to ablate these two predicted dGATAe binding sites found in this enhancer. Mutation of these sites greatly reduced activation, indicating that these dGATAe binding sites contribute to enhancer activity (Figure 3.2G-H) and suggesting that dGATAe regulates this enhancer in the dorsal embryonic ectoderm while also highlighting a potential role for Hh signaling in the process of dorsal closure.

3.3.2 The case for GAGA factor input into *ptc* enhancers

In addition to enhancer PN, I also noticed that the expression pattern of 1EH also did not appropriately recapitulate the *ptc* response; rather than producing a more restricted pattern, 1EH produced too broad of an expression pattern (Figure 3.3). I wanted to determine whether a cluster of GBS at the 3' end of the enhancer was primarily responsible for the activity of 1EH. To test this, I generated a minimal

enhancer construct, 1GH, consisting of this cluster of conserved sites (Figure 3.3A). Interestingly, although 1GH contains most of the well-conserved GBS (Chapter 2), it was not sufficient to respond to Hh signaling (Figure 3.3B,C) even using the endogenous *ptc* promoter, BA. I then hypothesized that 1EH required information from its 5' end, specifically within 1EF. By looking at sequence conservation, I found several blocks of well conserved sequence, and within those blocks, noticed several GAGAG repeats, which are characteristic binding sites for GAGA Factor, or GAF (Oh et al., 2013; Schaaf et al., 2013; van Steensel et al., 2003; Figure 3.3A, green lines). Furthermore, these GAF binding sites are located over a region of DNA predicted to be highly accessible by DNAse-hypersensitivity at the same stage of embryonic development as when this enhancer is active (Figure 3.3A, blue track). The initial experiment of generating the 1GH transgenic reporter fragment was a fortuitous choice: by isolating the cluster of GLI binding sites in 1GH, I had also excluded the predicted GAF binding sites in 1EF. Examining the same results with this new information, the drastic reduction in activation and patterning might be explained by the loss of those GAF binding sites (Figure 3.3B,C). Using the native *ptc*BA promoter in this experiment was not sufficient to rescue activation even though the BA region contains GAF binding sites, sits on a region of accessible chromatin, and has predicted GAF occupation in embryos (Figure 3.3A, 2.12).

3.3.3 Regulatory Factor X: another link between flies and mice

In addition to examining the fly enhancers identified in Chapter 2, I also examined the sequences of several mouse enhancers using the same basic strategy outlined in Figure 3.1. This resulted in the identification of a highly conserved binding

site for a TF called Regulatory Factor X (RFX). The RFX binding motif had independently been shown to be over-represented in a previous GLI1 ChIP-seq dataset, suggesting that it might be a significant player in regulating Hh target genes (Peterson et al., 2012). We identified a high quality RFX binding site in a previously characterized mouse *Ptch1* enhancer, called Peak2 (Vokes et al., 2007). This small, 180bp promoter proximal *Ptch1* enhancer is active in the murine neural tube, a well-characterized site of SHH signaling. Much of the sequence in this enhancer is well conserved, often to zebrafish, suggesting functional relevance. The RFX binding site we found sits in one of these well-conserved regions (Figure 3.5A-C). While there are no available ChIP-seq data in mice for RFX, there has been ChIP-seq analysis in human cells looking at the binding of RFX5, one of 8 members of the family the RFX family of transcription factors (Choksi et al., 2014). Interestingly, there is a strong peak of RFX5 binding to Peak2 in these cells, suggesting that this might be a functional binding site (ENCODE Project Consortium, 2012). With this information, I generated a luciferase construct containing a wild type Peak2 enhancer and compared it to a version with its RFX binding site mutated. In NIH3T3 cells, wild type Peak2 responds to SHH signaling via GLI1 induction, but this response is weakened when the RFX binding site is mutagenized, demonstrating this binding site is required for maximum levels of activation upon SHH pathway stimulation (Figure 3.4D).

Because Peak2 was active in murine cells as well as the neural tube, I wanted to see if this enhancer could also respond to Hh signaling in *Drosophila*. Therefore, I examined Peak2 as a GFP reporter in transgenic flies. I found that Peak2-GFP was active in pupal and larval salivary glands (Figure 3.5E-E'') and depended on its RFX

binding site for achieving maximum levels of activation in this context (Figure 3.5 F-F''). The minimal promoter alone weakly activated GFP in these contexts, but was weaker than the Peak2 RFXWT construct (Figure 3.5 G-G'' compared to 3E-E"). Peak2 failed to respond in any other context examined, including known sites of Hh signaling like imaginal discs, which is not totally unexpected, since there is no RFX protein normally expressed in these contexts (data not shown). There is, however, RFX expressed in salivary glands during late embryogenesis (Hammonds et al., 2013; Tomancak et al., 2002; Tomancak et al., 2007). Together, these data suggest that Peak2 requires RFX input to respond to Hh signaling and hints at a strong evolutionarily conserved mechanism for Hh activation, fitting the evolutionary themes discussed in Chapter 2.

3.3.4 Predicted TFs that weakly regulate or do not contribute to *ptc* expression in the embryonic ectoderm

It is also important to note that not all predicted tissue specific inputs contributed to *ptc* expression. For instance, enhancer LK produced a highly restricted expression pattern in the embryonic ectoderm in posterior cells while enhancer VT was active in the anterior region of each segment (Figure 2.5E,F). After analyzing the sequences of these enhancers, I found conserved binding sites for at least one transcriptional repressor in each enhancer. First, I found a Slp1 binding site in enhancer LK (Figure 3.5A-C). Slp1 is a transcriptional repressor, and is active specifically in the anterior cells proximal to the Engrailed positive, Hh secreting cells (Gebelein et al., 2004**;** Figure 3.5D). Importantly, these are cells in which *ptc* is expressed, but LK fails to respond. I hypothesized that LK was being repressed in the anterior by Slp1. Upon mutation of this Slp1 binding site, I saw a weak de-repression in these cells (Figure 3.5E-F, arrows point out derepression),

suggesting that Slp1 might be partially repressing *ptc* in these cells directly through enhancer LK.

I next examined the conserved regions of enhancer VT, where I found an Odd-Skipped (Odd) binding site using Fly Factor Survey that was well-conserved in other *Drosophilid* genomes (Figure 3.6B,C). Odd is a transcriptional repressor that is expressed specifically in the posterior of Hh positive cells in the embryonic ectoderm (Mulinari and Häcker, 2009, Figure 3.6D). When I mutated this predicted binding site in the context of VT, however, I did not see de-repression in the posterior, as expected. The GFP expression pattern was largely unchanged compared to wild type expression (Figure 3.6E,F), indicating that even though the sequence contained a predicted binding site for Odd, it is not functional in this context. This experiment does not rule of the possibility of Odd regulating expression of *ptc* through other binding sites in this, or any other enhancer in the locus, as there might be other (potentially lower affinity) binding sites missed in the analysis.

3.4 Discussion

In this chapter, I have characterized several novel TFs that contribute to the tissue specific response of *patched* to Hh signaling. While GLI binding site affinity informs the tissue specific response of these enhancers, GLI is not the sole contributor to enhancer function. Here I present evidence for novel additional inputs that contribute to patterning created by Hh signaling in different tissues and developmental contexts.

It is interesting to note that enhancer PN is the only embryonic ectoderm enhancer identified in Chapter 2 that was not affected by removing its GBS (Figure 3.2B,C). This could potentially be accounted for by two, non-exclusive explanations.

First, there could be even lower affinity GBS within this enhancer that were not examined in this analysis. When I mapped the low affinity GBS onto the *ptc* enhancers, I chose a cutoff of the top 200 sites (Gurdziel et al., 2015; Hallikas et al., 2006; Lorberbaum et al., 2016). Lower affinity GBS might be key to regulating expression in dorsal cells. Second, enhancer PN might work together with other enhancers in the locus to generate the full response of *ptc* to Hh. Testing this would require a locus wide manipulation using targeted genomic engineering like CRIPSR/Cas9, rather than the reporter experiments mostly used in this dissertation. We are actively working on both of these explanations to better understand how Hh targets are regulated (see Chapter 5 for more details of these ongoing experiments).

I then demonstrated that enhancer PN requires input from dGATAe in the dorsal most cells of the embryonic ectoderm (Figure 3.2), suggesting a novel role for dGATAe in the regulation of Hh target genes. These data hint at a direct mechanism by which Hh signaling might be playing a role in dorsal closure, which is significant because many of the components of Hh signaling are expressed and functional in the cells that undergo dorsal closure where dGATAe is most active (Yavari et al., 2010; Figure 3.2F). The process of dorsal closure in flies is essentially two epithelial sheets of cells that are “sewn” or “zipped” together to close an open region of the embryo (Martin and Parkhurst, 2004). The process of epithelial tissues coming together is important in development, since failure of successful fusion of tissues in mammalian development often leads to defects such as spina bifida and cleft lip/palate (Millard and Martin, 2008). These defects are also two common defects associated with defective SHH signaling as

well (Cobourne et al., 2009; Teglund and Toftgård, 2010). Enhancer PN's requirement of dGATAe might lend additional insight to the role of Hh signaling in epithelial fusion.

The potential regulation of *ptc* by GAF has already been extensively discussed in Chapter 2, but mostly in the context of the *ptc* promoter. It is not uncommon for Trithorax group proteins (TrxG) to mediate epigenetic regulation of the locus from non-promoter regions (Bowman et al., 2014; Brown and Kassis, 2013; Kassis and Brown, 2013; Schaaf et al., 2013). Here I introduce another piece of evidence suggesting that GAF regulates *ptc* in this way. In the case of 1EH, it seems that GAF activates its target, since when you remove most of the well-conserved GAF binding sites from the enhancer, expression is attenuated (Figure 3.3B,C). Further studies that target those GAF sites directly would provide valuable information about how GAF regulates 1EH. In the above experiment, I did not mutagenize target sites, rather, I removed a large region of the enhancer that could have additional binding sites for other activating transcription factors besides GAF. Thus we cannot be certain that all of the activation seen in 1EH vs 1GH mediated by the GAF binding sites.

It is also interesting that the GAF input into the promoter region, BA, was not able to provide enough activation to 1GH to generate a more complete Hh response (Figure 3.3A,C), even though I have already demonstrated that region B contains significant activation potential in the embryo (Figure 2.11C). This could be explained by either of two explanations. First, like discussed earlier, there might be additional activating factors binding to this enhancer in the 1EF region. By removing this entire sequence, I might have lost additional factors required to respond. Second, this enhancer is found in the intron of *ptc*, therefore when it presumably is looped to the promoter in its native

genomic context to activate *ptc* expression, the orientation of its GAF sites are at a different position than when examined in the context of the reporter tested here. This is also in agreement with previous data because the other embryonic enhancer that showed increased activation when examined with the BA promoter (LK, Figure 2.11) is located upstream of the promoter, so it may not change 5' to 3' orientation when it interacts with the promoter. We could test this by completing a 3C (chromatin conformation capture), experiment to identify how these enhancers are physically interacting with the promoter to give us a clue about how orientation of enhancers might affect expression (Dekker et al., 2002).

The identification of RFX input into the mouse *Ptch1* enhancer highlights an exciting new input into Hh signaling. It is known that SHH signaling in mammals often requires cilia to function properly (Carpenter et al., 2015; Choksi et al., 2014) and a large number of developmental defects associated with cilia dysfunction, known as ciliopathies, have already been described (Hildebrandt et al., 2011). Hence, the RFX input into *Ptch1* regulation described here provides at least one potentially direct trancriptional link between cilia and SHH target gene regulation. Furthermore, many RFX transcription factors are required for ciliogenesis (Ashique et al., 2009; Dubruille, 2002; Swoboda et al., 2000), so a link between *patched* regulation and RFX implicates either a role for SHH signaling in building and maintaining cilia, or a role for RFX in regulating expression of genes that are processed in cilia.

Another exciting result is the transcriptional activity of Peak2 in the salivary glands of *Drosophila*. Because this is a mouse enhancer that is able to respond to Hh signaling in the context of a fly, it is consistent with our previous claim that there is a

deeply conserved *cis*-regulatory strategy regulating *ptc* during development. Additionally, Peak2 only responds in a very restricted region of the fly, but not in the regions of the fly best characterized by Hh signaling, like the imaginal discs focused on in Chapter 2. This is not that surprising because RFX is not expressed in that context, probably because these are non-ciliated tissues. In fact, flies only use cilia in a few settings, most notably in sensory neuron differentiation where a new role for Hh signaling has recently been described (Kuzhandaivel et al., 2014), and where RFX has been shown to be active as well (Cachero et al., 2011). If Hh is required in the sensory neurons, then *ptc* should respond since it is a broadly responsive target gene (Chapter 2). Following this logic, then, it is likely that one of the enhancers characterized in Chapter 2 responds in sensory neurons. Since we have a large number of enhancers to examine, we may be able to find one that responds to Hh in this region and then search the sequence for potential RFX binding sites.

Peak2 requires its RFX binding site in the context of the salivary gland, raising the possibility that this tissue might be ciliated in flies. While this has never been demonstrated, it has been shown that mammalian salivary glands require cilia or normal development and function (Bernfield et al., 1972). Future work would be required to demonstrate if *Drosophila* salivary glands are ciliated, which could be done by using known markers of cilia, like Arl13B, for example (Joiner et al., 2015; Vieillard et al., 2015). We could then attempt to identify the fly *patched* enhancer(s) that respond in the salivary glands and then see how RFX might contribute to this expression. An RFX ChIP-seq experiment would also be very helpful to identify when and where this ciliary

factor binds DNA, although isolating sensory neuron in flies might be challenging since they are found in such a heterogeneous cell population in the embryo.

Not all of our predicted transcription factors regulated *patched* enhancers exactly as we expected. Our predicted Slp1 and Odd inputs into LK and VT, respectively, did not contribute much patterning information to *ptc* expression in our transgenic readout. These data could mean that, although both binding sites are highly similar in sequence to the motifs for Odd and Slp1, there might be more, lower affinity binding sites for these TFs within these enhancers, similar to how Ci requires low affinity sites in the embryo. These particular positional weight matrices (PWMs) were generated by yeast one hybrid screens. So, while these TFs are able to recognize this consensus, the PWM allows for some nucleotide variation (Figure 3.5C,H) and I may have incorrectly predicted these binding sites. To account for this, I could more carefully search the sequences of these enhancers for lower affinity sites and also test the predicted binding sites using gel shift assays, or EMSAs to validate binding events. Furthermore, I could also use ChIP-qPCR to identify if either Odd or Slp1 recognizes these enhancers in their native cellular context.

The partial derepression seen upon removal of the Slp1 binding site in enhancer LK does suggest Slp1 might play a role in *ptc* regulation (Figure 3.6). However, the response was weak when compared to nearly all of the binding site mutations completed with GLI binding sites in Chapter 2, and with dGATAe (Figure 3.2). This weak de-repression might still be functionally relevant, but much more work is needed to confirm the potential Slp1 contribution. In addition to the ChIP-qPCR already discussed, I would also like to more carefully analyze this sequence to search for lower

affinity Slp1 binding sites, since I only looked for the highest affinity sites in this set of experiments. It is also important to note that when the Slp1 protein is deleted in a genetic knockout, *ptc* is not able to respond to Hh signaling, although, it was additionally shown that Hh itself is not produced at its normal levels either, so Slp1 may not be regulating *ptc* directly (Cadigan et al., 1994).

Another interpretation of my concerning enhancers VT and LK is that neither Slp1 nor Odd are regulate these enhancers. This highlights the possibility that enhancer VT is required for activation in the anterior, while LK is required for activation in the posterior and they work together in their native context to generate a full response to Hh signaling. This mechanism could be tested by examining expression of *patched* in the context of the endogenous locus, and manipulating each enhancer individually, using CRISPR/Cas9. Alternatively, if I made a transgene containing both VT and LK as a single enhancer driving GFP, they might generate a more full response to Hh signaling. One of the pitfalls of this is that this experiment violates the endogenous spacing of these enhancers, which might prevent the proper recognition of TFs with their cognate binding sites in the DNA.

While there is much work to be done to highlight the function of these elements, I have identified a novel set of tissue specific transcription factors including dGATAe, GAGA factor (GAF), and RFX that contribute to *patched* regulation and have laid out a successful framework with which we can predict context specific transcription factors that might be regulating known enhancers (summarized in Figure 3.1).

3.5 Materials and Methods

DNA sequence alignments

Sequences and multi-species alignments were obtained from the UCSC Genome Browser (www.genome.ucsc.edu) or evoPrinterHD (https://evoprinter.ninds.nih.gov/evoprintprogramHD/evphd.html).

DNA cloning and mutagenesis

Wild-type *ptc* enhancers were amplified by PCR (Roche Expand High Fidelity PCR System) from BAC DNA (CH322-170A-12 or CH322-188E13). PCR primers are provided in Table 3.1. Enhancer constructs were sub-cloned into the pENTR/D-TOPO plasmid (Life Technologies) by TOPO cloning. Enhancers tested with the hspmin promoter, taken from the *D. melanogaster* Hsp70 gene, were subsequently cloned into the pHPdesteGFP transgenesis vector via LR Cloning (Life Technologies). Enhancers tested with the endogenous *ptc* promoter were cloned by traditional methods into the pStinger transgenesis vector (Barolo et al., 2000). Targeted GBS mutations were created by overlap-extension PCR (Swanson et al., 2010b). All enhancers and promoters were screened by restriction digest and sequencing. Peak2 mouse enhancer sequence was taken from the mm9 genome build and was ordered as a GeneBlock fragment to facilitate cloning (Table 3.2**;** details about ordering these fragments can be found at https://www.idtdna.com/pages/products/genes/gblocks-gene-fragments).

Drosophila transgenesis

P-element transformation was performed as previously described (Swanson et al., 2010a) in the w1118 strain. Site-directed transformation by embryo injection was performed as described by (Bischof et al., 2007), with reporter transgenes integrated into a ΦC31 landing site at genomic position 86Fb.

Immunohistochemistry and confocal microscopy

Primary antibodies used included rabbit anti-EGFP (Invitrogen 1:100), mouse anti-Engrailed (Developmental Studies Hybridoma Bank 4D9, 1:50). In *Drosophila* embryos, EGFP antibodies were used to visualize reporter expression; in salivary glands from both pupa and larvae, native GFP fluorescence was imaged directly. Antibodies obtained from the Developmental Studies Hybridoma Bank were developed under the auspices of the National Institute of Child Health and Human Development and maintained by the Department of Biological Sciences, The University of Iowa (Iowa City, IA). AlexaFluor488, AlexaFluor555, and AlexaFluo568 conjugates with secondary antibodies from Invitrogen were used at 1:2000 dilutions. DAPI was included in the Prolong Gold antifade mountant (Life technologies). Confocal images were captured on an Olympus FluoView 500 Laser Scanning Confocal Microscope mounted on an Olympus IX-71 inverted microscope, and on a Nikon A1 confocal miscroscope. Samples to be directly compared were fixed, prepared, and imaged under identical confocal microscopy conditions and settings.

Cell culture assay

NIH/3T3 cells were cultured at 37°C, 5% CO2, 95% humidity in Dulbecco's modified eagle medium (DMEM; Gibco, cat. #11965-092) containing 10% bovine calf serum (ATCC; cat. #30-2030) and penicillin/streptomycin/glutamine (Gibco, cat. #10378-016). Luciferase assays were performed by plating 2.5 x 104 cells/well in 24 well plates. The next day, cells were co-transfected using Lipofectamine 2000 with the DNA constructs indicated in each experiment in addition to *ptc*Δ136-GL3 (Chen et al., 1999; Nybakken et al., 2005) and pSV-Beta-galactosidase (Promega) constructs to report Hh pathway activation and normalize transfections, respectively. GLI1 cDNA was added where relevant to activate the Hh pathway. Cells were changed to low-serum media (DMEM supplemented with 0.5% bovine calf serum and penicillin/streptomycin/glutamine) 48 hours after transfection and cultured at 37°C in 5%CO2 for an additional 48 hours. Cells were harvested and luciferase and beta-galactosidase activities were measured using Luciferase Assay System (Promega, cat. # E1501) and BetaFluor -gal assay kit (Novagen, cat. #70979-3). Multiple assays were performed and each treatment group was assayed in triplicate. Two-tailed t tests for two samples with unequal variances were used to compare samples.

Encode Project Consortium

Peak2 was analyzed with information from ENCODE (ENCODE Project Consortium, 2012) with Data Coordination Center (DCC) and GEO accession numbers of the datasets,DCC accession: ENCSR037HRJ; GEO accession: GSE30567.

3.6 Acknowledgements

I would like to thank Scott Barolo for his guidance in this chapter, as he contributed a great amount to the identification of binding sites that were examined here. In addition, Ben Allen provided hours of helpful discussions as well, particularly in regard to RFX and cilia. My colleague Brandon Carpenter generously completed the luciferase assays examining Peak2 in the NIH3T3 cells.

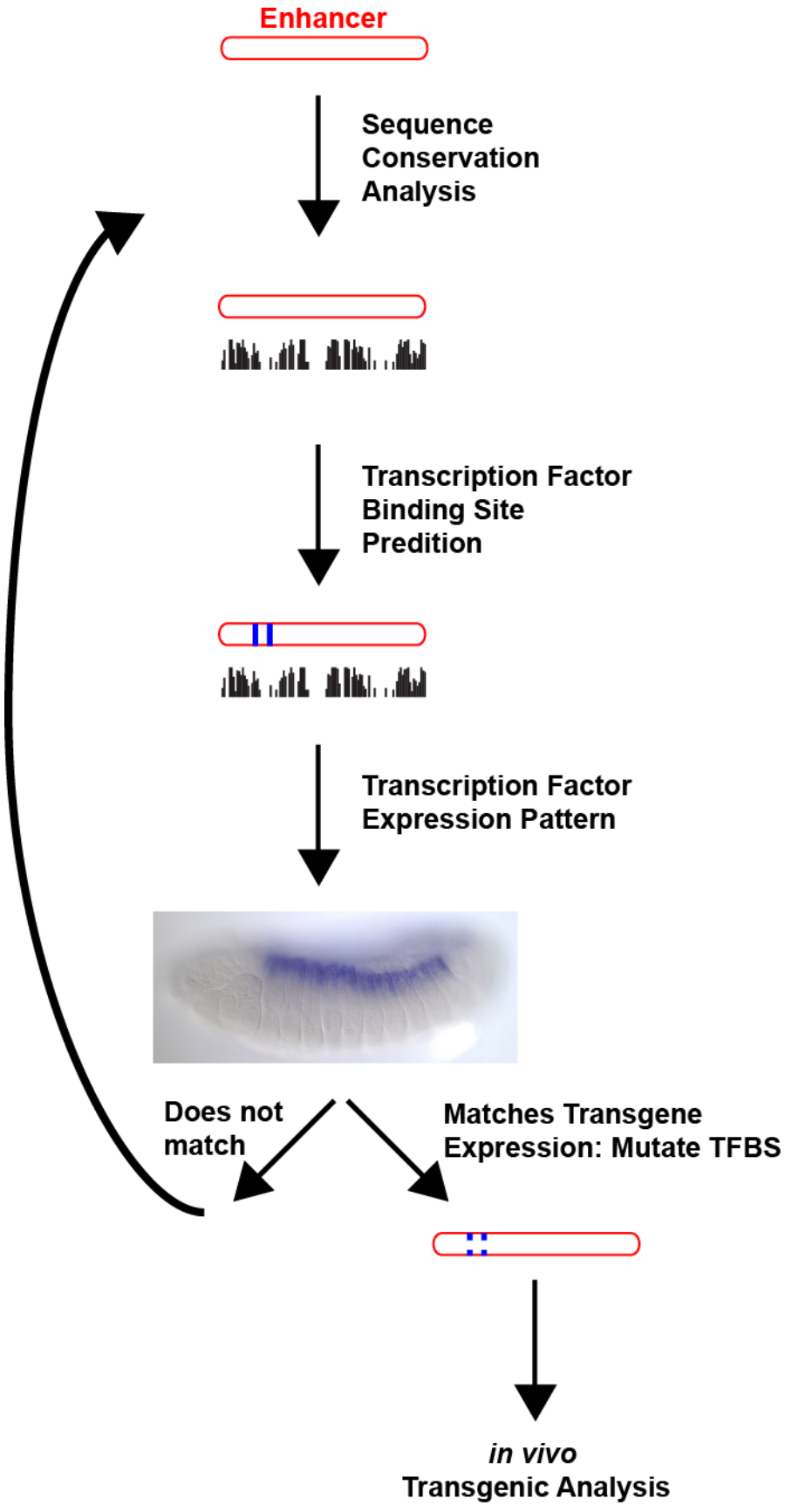

Enhancer
Sequence
Conservation
Analysis
Transcription Factor
Binding Site
Predition
Transcription Factor
Expression Pattern
Does not
match
Matches Transgene
Expression: Mutate TFBS
in vivo
Transgenic Analysis

Figure 3.1 Predicting non-GLI transcriptional regulators of *patched*

After a predicted *ptc* enhancer had been identified in Chapter 2, the sequence was examined for regions that were well conserved using the UCSC genome browser (black bars underneath enhancer). These conserved regions were then examined with binding site prediction software. High-ranking binding sites were followed up by examining the tissue specific expression pattern of that TF using either an *in situ* database or by searching through published literature. If the expression pattern of this TF matched the expression pattern of the enhancer-reporter construct, the predicted binding sites were mutated using overlap extension PCR and examined in transgenic animals (bottom right). If the expression pattern of the predicted TF did not match the enhancer sequence, I did not follow up on it and repeated the process (bottom left).

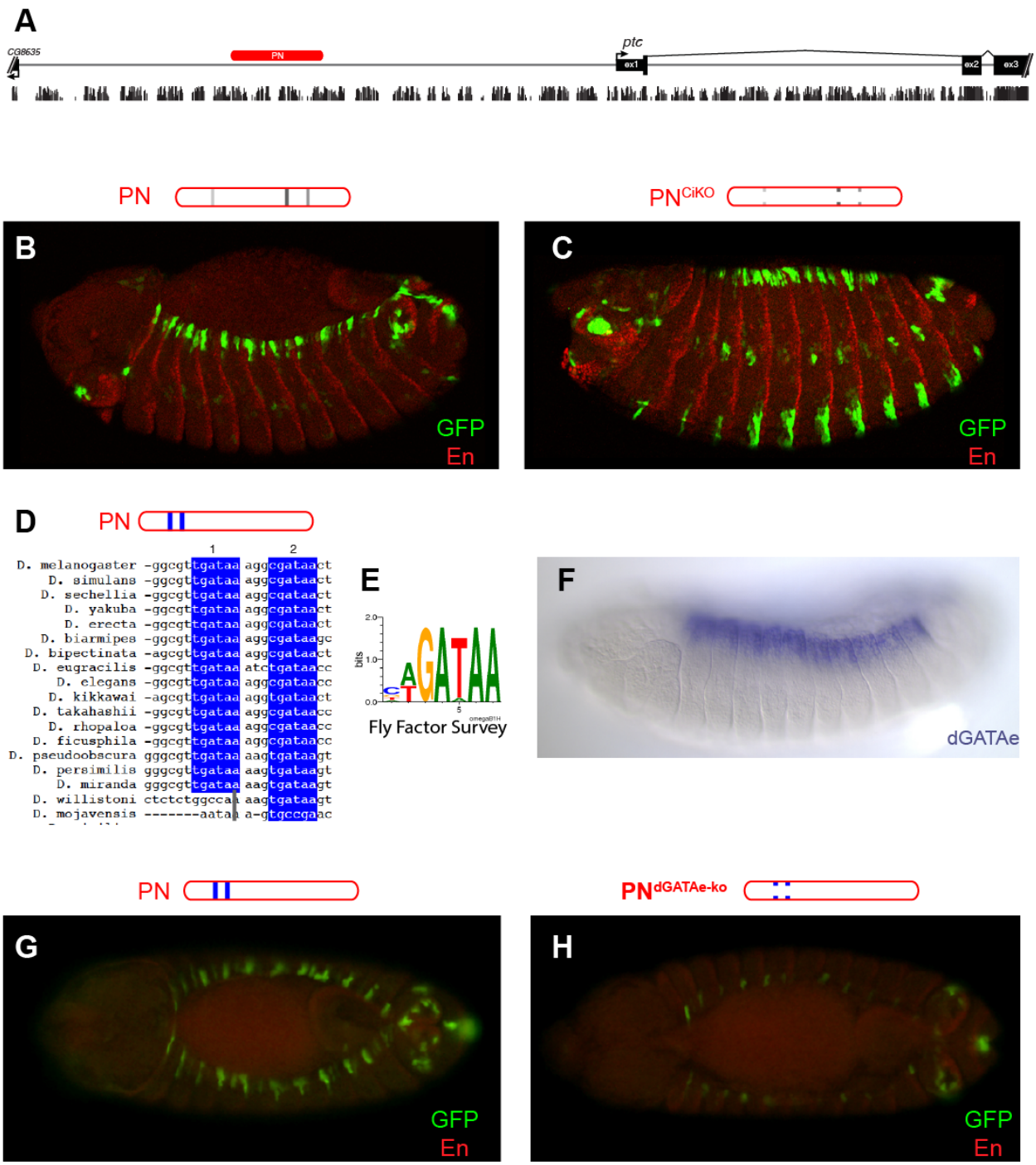


Figure 3.2 dGATAe contributes to *ptc* expression via enhancer PN

(A) Enhancer PN, in red, is shown in relation to *ptc.* (B) Wild type PN drives GFP expression. (C) PN lacking best predicted GLI binding sites as a GFP reporter. (D) Conservation of predicted dGATAe binding sites in enhancer PN. (E) Fly Factor Survey motif associated with dGATAe. (F) *In situ* showing expression of dGATAe in appropriately staged embryo (from the Berkeley Drosophila Genome Project). (G) Wild

type PN drives GFP expression, dorsal view. (H) PN lacking best predicted dGATAe binding sites as a GFP reporter.

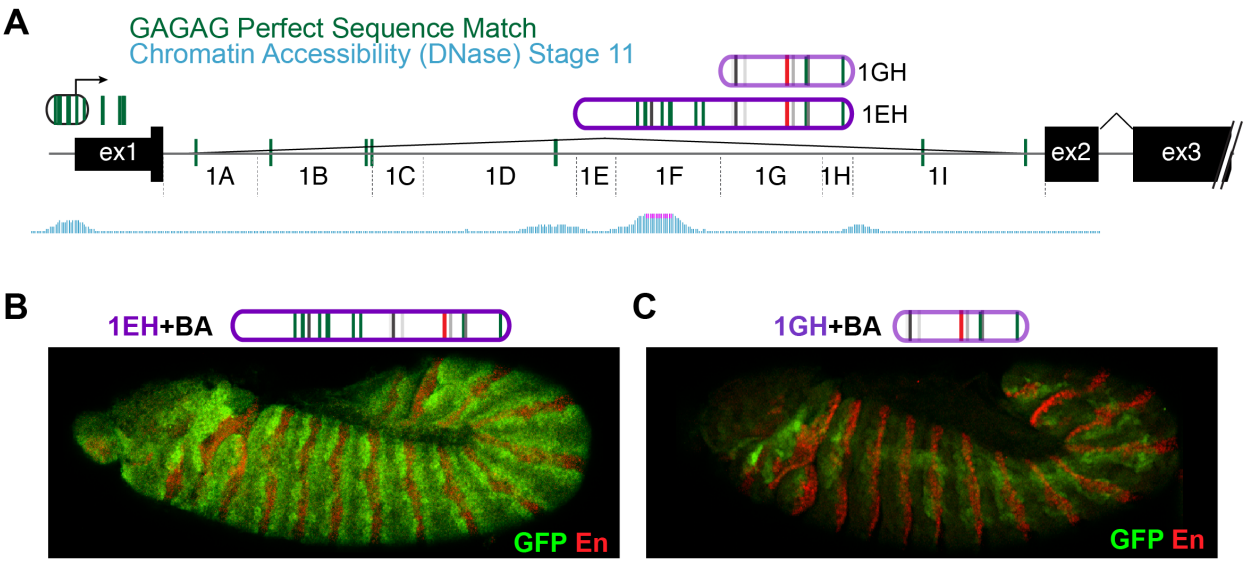


Figure 3.3 GAF contributes to *ptc* expression via enhancer 1EH

(A) Enhancers 1EH and 1GH are found in the first intron of *ptc.* DNAse hypersensitivity is shown as the blue track underneath the gene diagram (taken from UCSC genome browser, dm3, red bar over the blue peak in 1F indicates statistical significance). Green lines in the enhancers and locus indicate predicted GAGAG binding sites based upon sequence alone. (B) Wild type 1EH drives GFP expression. (C) Wild type 1GH drives GFP expression.

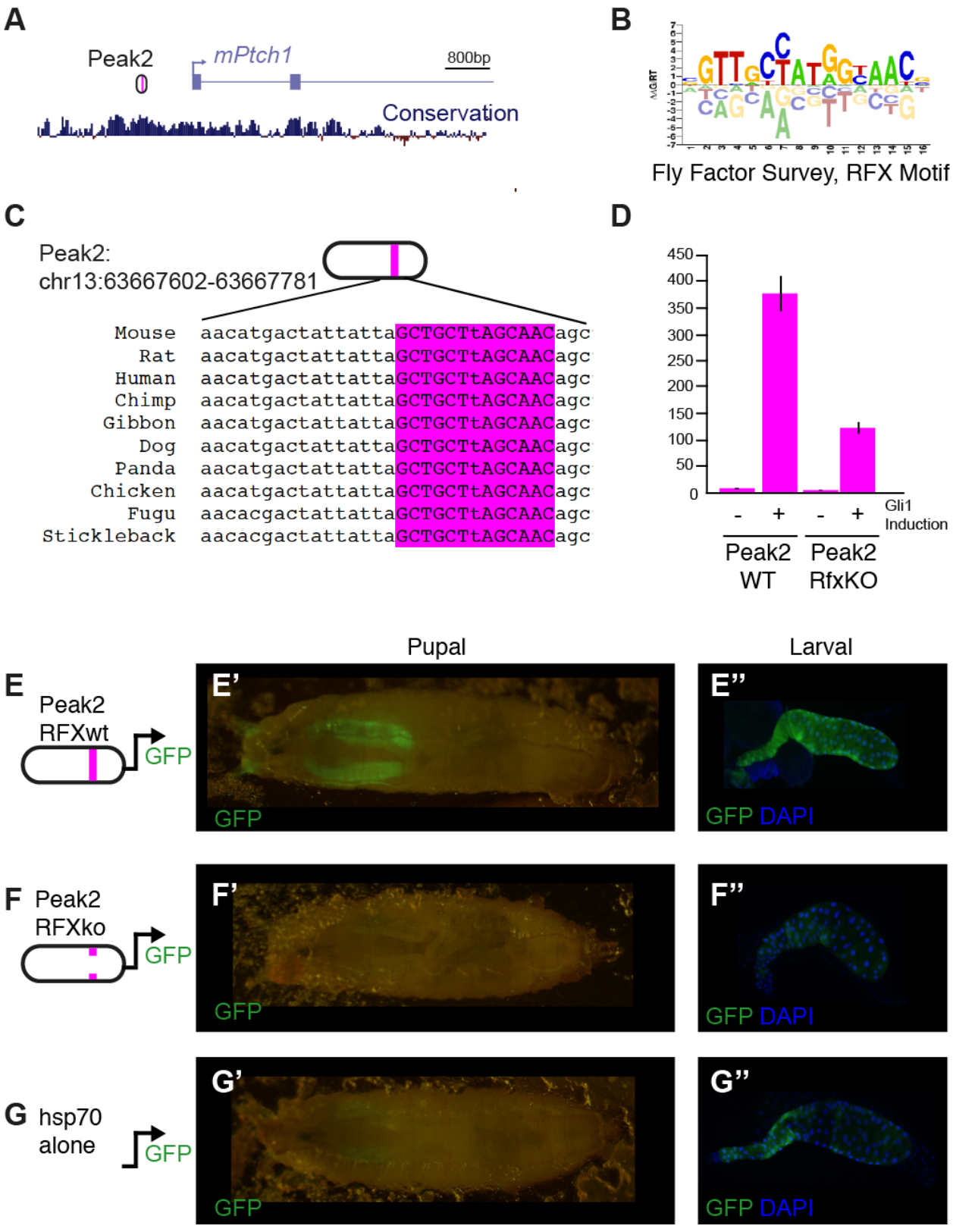


Figure 3.4 RFX regulates *Ptch1* in mice and might contribute to *ptc* regulation in *Drosophila* salivary glands

(A) The Peak2 enhancer is depicted upstream of mouse *Ptch1*. Conservation of sequence is shown as a blue track underneath the gene diagram, all taken from the UCSC genome browser. (B) Positional weight matrix of RFX, from Fly Factor Survey. (C) Conservation of RFX binding site in Peak2 from mouse to fish. (D) GLI1-responsiveness of Peak2 enhancer in NIH/3T3 cells (wild type on left, Peak2-RFXbinding site KO on right). (E-G") fly transgenic GFP reporter constructs examine in pupal (E'-G') and larval (E"-G") salivary glands.

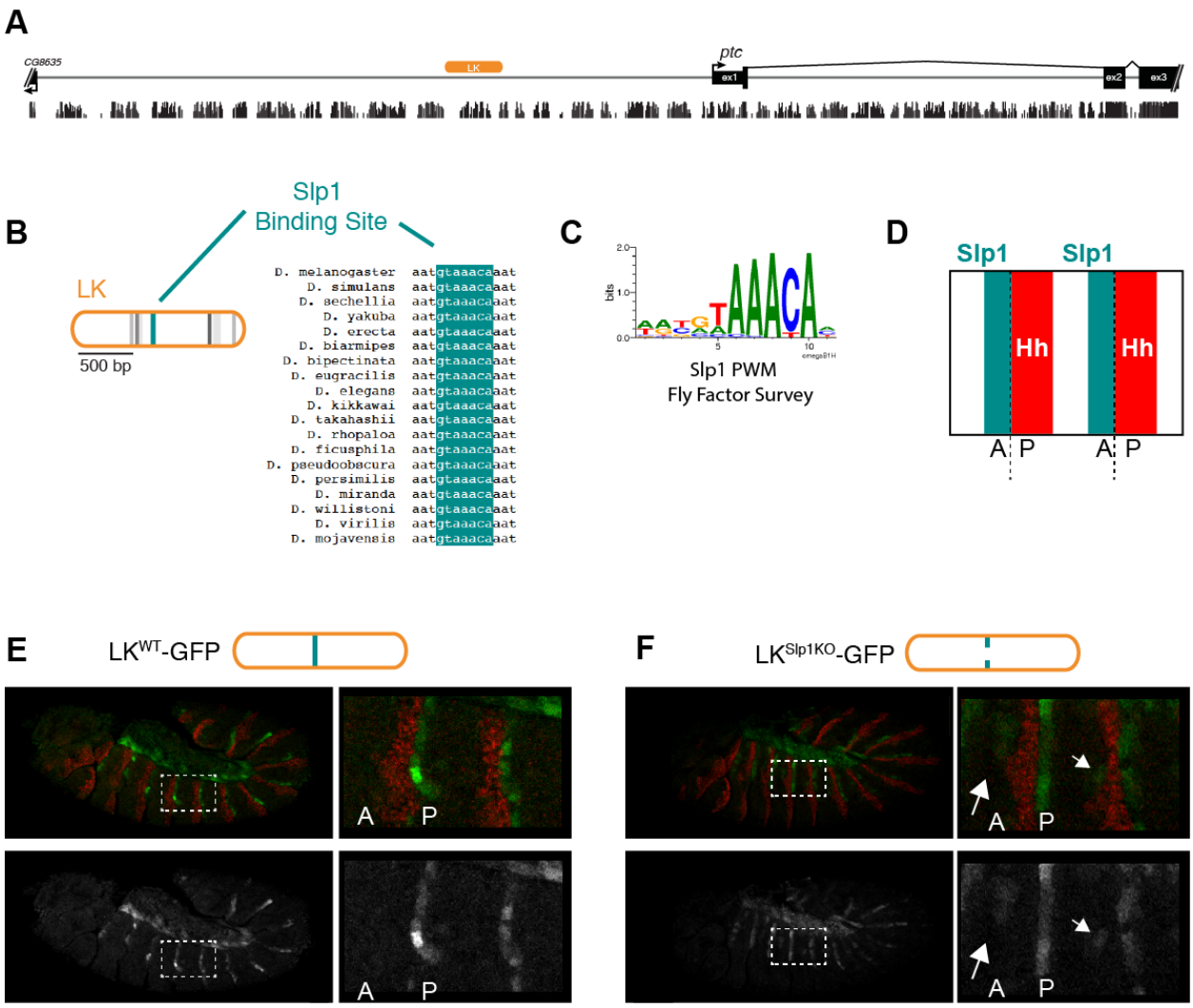


Figure 3.5 Slp1 provides minimal contribution to repression of enhancer LK

(A) Enhancer LK, in orange, is shown in relation to *ptc.* (B) Conservation of predicted Slp1 binding sites in enhancer LK. (C) Fly factor Survey motif associated with Slp1. (D) Expression of Slp1 in embryos (adapted from Gebelein et al., 2004). (E) Wild type LK drives GFP expression (GFP is green, En is Red). Bottom is GFP channel alone. (F) LK lacking best predicted Slp1 binding site as a GFP reporter (GFP is green, En is Red). Bottom is GFP channel alone. Arrows highlight de-repression in the anterior.

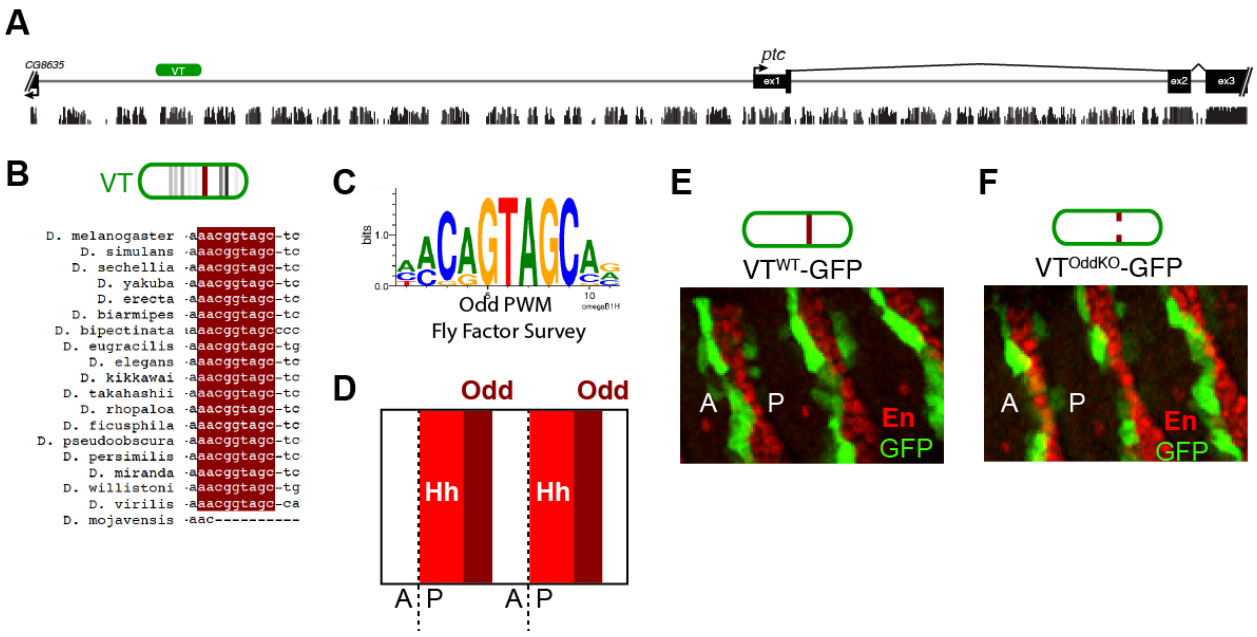


Figure 3.6 Odd does not contribute to VT repression through predicted binding site

(A) Enhancer VT, in green, is shown in relation to *ptc.* (B) Conservation of predicted Odd binding site in enhancer VT. (C) Fly Factor Survey motif associated with Odd. (D) Expression of Odd in embryos (adapted from Mulinari and Häcker, 2009). (E) Wild type VT drives GFP expression mainly in anterior. (F) VT lacking best predicted Odd binding site as a GFP reporter.

Primer	Sequence	Restriction Site
LK_F	CACCGGCGCGCCCTACTTGGTTTGATAAAT	AscI
LK_R	**GTACCGCGG**ACGGTGTGTGTGAACCCAACTAATTG	SacII
PN_F	CACCGGCGCGCCCGCCAAAATGAAAATTATTACAAAG	AscI
PN_R	**GTACCGCGG**CGTCCCTCTGTCGTCGCTGGG	SacII
VT_F	CACCGGCGCGCCACAGACGGGGCTACACTGAG	AscI
VT_R	**GTACCGCGG**GACAGACCCCGAAGACTGAG	SacII
1EH_F	CACCGGCGCGCCGAAGTGCTTAACAAGTTAAC	AscI
1EH_R	**GTACCGCGG**CACGACAACCAATGAGATCG	SacII
1GH_F	CACCGGCGCGCCGGTAATAAATGCGGCAGACG	AscI

Table 3.1 Primers used to generate enhancer-reporter constructs to detect tissue specific inputs into *ptc* enhancers

All primers are written from 5' to 3' orientation. F refers to forward primer, R refers to reverse primer. Additional binding site to facilitate cloning are depicted in colors.

Peak2 geneblock	Sequence
Peak2-RFXwt geneblock	CACCGGCGCGCCAGCGAGCGAAAGCGAGCGAAAGAGAAAA AGGCTGGAACTCCCGCCCCCGGGGCTGTCAGATGGCTTGG GTTTCTGCAACGCGATTGGCTCTTGGAGGGCAGAAATTACTC AGCAAACATGACTATTATTA**GCTGCTTAGCAAC**AGCTCACCAA AGTAGAGAGACCACCCAGGTAGGCAACCCCCGTGTGTGCAT CCCAGGCTTGGGGGTGGGGGGGCGCTCGCTCAGCGCCAAC CCTCT**CCGCGGTAC**
Peak2-RFXko geneblock	CACCGGCGCGCCAGCGAGCGAAAGCGAGCGAAAGAGAAAA AGGCTGGAACTCCCGCCCCCGGGGCTGTCAGATGGCTTGG GTTTCTGCAACGCGATTGGCTCTTGGAGGGCAGAAATTACTC AGCAAACATGACTATTATTA**CGGCAGtaCcTTG**AGCTCACCAA AGTAGAGAGACCACCCAGGTAGGCAACCCCCGTGTGTGCAT CCCAGGCTTGGGGGTGGGGGGGCGCTCGCTCAGCGCCAAC CCTCT**CCGCGGTAC**
	Predicted RFX Binding Site

Table 3.2 Geneblock sequences used to generate Peak2 enhancer-reporter constructs

Chapter 4

Computational Prediction of Hedgehog Responsive Enhancers

4.1 Abstract

The Hedgehog (Hh) signaling pathway directs a multitude of cellular responses during embryogenesis and adult tissue homeostasis. Stimulation of the pathway results in activation of Hh target genes by the transcription factor Ci/GLI, which binds to specific motifs in genomic enhancers. In *Drosophila*, only a few enhancers (*patched*, *decapentaplegic*, *wingless*, *stripe*, *knot*, *hairy*, *orthodenticle*) have been shown by in vivo functional assays to depend on direct Ci/GLI regulation. All but one (*orthodenticle*) contain more than one Ci/GLI site, prompting us to directly test whether homotypic clustering of Ci/GLI binding sites is sufficient to define a Hh-regulated enhancer. We therefore developed a computational algorithm to identify Ci/GLI clusters that are enriched over random expectation, within a given region of the genome. Candidate genomic regions containing Ci/GLI clusters were functionally tested in chicken neural tube electroporation assays and in transgenic flies. Of the 22 Ci/GLI clusters tested, seven novel enhancers (and the previously known *patched* enhancer) were identified as Hh-responsive and Ci/GLI-dependent in one or both of these assays, including: *Cuticular protein 100A* (*Cpr100A*); *invected* (*inv*), which encodes an *engrailed*-related transcription factor expressed at the anterior/posterior wing disc boundary; *roadkill (rdx)*, the fly homolog of vertebrate *Spop*; the segment polarity gene *gooseberry (gsb)*; and

two previously untested regions of the Hh receptor-encoding *patched (ptc)* gene. We conclude that homotypic Ci/GLI clustering is not sufficient information to ensure Hh-responsiveness; however, it can provide a clue for enhancer recognition within putative Hedgehog target gene loci.

4.2 Introduction

The Hedgehog (Hh) signaling pathway plays multiple roles in embryonic organ development and adult tissue homeostasis across animal phyla (Briscoe and Thérond, 2013; Jiang and Hui, 2008; McMahon et al., 2003). Hh signaling directs specific cell fate choices, controls tissue patterning and governs cell proliferation. Several human developmental diseases are caused by altered Hh signaling, including spina bifida, exencephaly (Murdoch and Copp, 2010), holoprosencephaly (Schachter and Krauss, 2008), cleft lip/palate (Lipinski et al., 2010), and a host of malformations in vertebral, anal, cardiac, tracheal, esophageal, renal, and limb tissues (together known as VACTERL Association;(Ngan et al., 2013)). Aberrant Hh signaling is also responsible for several cancers, including basal cell carcinoma, medulloblastoma and rhabdomyosarcoma (Teglund and Toftgård, 2010). Recently, cancers of the pancreas, colon, ovary, stomach and lung have also been associated with increased Hh signaling (Barakat et al., 2010; Teglund and Toftgård, 2010), prompting initiation of clinical trials with Hh antagonists for some of these conditions (de Sauvage, 2007; Scales and de Sauvage, 2009; Sekulic et al., 2012; Tang et al., 2012).

The Hh-regulated GLI family transcription factors (including Cubitus interruptus (Ci) in the fly and GLI1-3 in mammals) are highly conserved across metazoans, as is the sequence of the preferred consensus Ci/GLI binding site (Hallikas et al., 2006;

Winklmayr et al., 2010). Despite the functional importance and high conservation of the Hh pathway, surprisingly little is known about its target genes in any organism. These target genes and their associated enhancers, which are responsible for the genomic response to Hh in development and disease, have significant potential therapeutic and diagnostic value.

One method for identifying putative enhancers is chromatin immunoprecipitation (ChIP) (Biehs et al., 2010; Lee et al., 2010; Vokes et al., 2007; Vokes et al., 2008), though such data are subject to the spatiotemporal limitations of the analyzed cells or tissues and can be diluted by a high number of false positive binding sites. While many potential murine Hedgehog-responsive enhancers have been pinpointed in this manner, relatively few have been functionally verified by mutagenesis of transcription factor binding sites (Lee et al., 2010; Peterson et al., 2012; Vokes et al., 2007; Vokes et al., 2008). In *Drosophila*, an alternative approach, DamID which fuses a DNA interacting protein to DNA adenine methyltransferase leading to methylation near binding locations, identified 52 potential Ci/GLI target enhancers, though none of these were functionally verified by mutagenesis of Ci/GLI sites (Biehs et al., 2010). To date, only seven *Drosophila* enhancers have been shown by mutational analysis to be Ci/GLI-dependent (Alexandre et al., 1996; Blanco et al., 2009; Hersh and Carroll, 2005; Müller and Basler, 2000; Ohlen et al., 1997; Piepenburg et al., 2000; Ramos and Barolo, 2013), which limits our understanding of the basic rules that govern their activity and context specificity.

Analysis of the known *Drosophila* Hh enhancers reveals that three (regulating *ptc, wg,* and *knot*) contain clusters of three or more Ci/GLI binding sites, while the

remaining enhancers (of the *dpp*, *stripe* and *hairy* genes) contain two sites (Alexandre et al., 1996; Blanco et al., 2009; Hersh and Carroll, 2005; Müller and Basler, 2000; Ohlen et al., 1997; Piepenburg et al., 2000; Ramos and Barolo, 2013). These examples, and findings in other systems (Berman et al., 2002; Gotea et al., 2010; Lifanov et al., 2003; Markstein et al., 2002; Rebeiz et al., 2002) suggest that homotypic clustering might be a relevant indicator of Hh enhancer activity in the fly. To test this, we computationally identified regions of the fly genome in which the density of Ci/GLI binding sites is enriched relative to chance expectation. We then tested the ability of these regions to: 1) drive Hh-dependent activity in the developing chicken neural tube, and 2) direct tissue-specific gene expression in a *Drosophila* transgenic reporter model. Importantly, the functional significance of the Ci/GLI binding motifs was also tested by mutation of these sites within each active enhancer. Of the 17 top clusters, four (23%) drove reporter expression in a known Hh domain and/or in a Ci/GLI-dependent fashion in one or both assays. Thus, while some Hh-regulated enhancers indeed contain homotypic clusters of Ci/GLI motifs, not all such clusters function as enhancers in vivo.

We also asked whether Ci/GLI site clustering could be used to predict the location of enhancers in genes that are known or putative targets of Hh signaling. Using the strategy outlined in Figure 4.1, we identified five such Ci/GLI site clusters, four of which were subsequently validated as Hh enhancers by functional assays (80%). Thus, altogether, our analysis of clustered Ci/GLI sites identified eight Hh enhancers, including seven novel enhancers and one previously identified *ptc* enhancer. These findings double the number of functionally verified Hh enhancers.

4.3 Results

4.3.1 Computational identification of clustered Ci/GLI sites across the *Drosophila* genome

To test if clustering of Ci/GLI sites could be used to predict Hh enhancers, we developed a computational strategy to identify all regions of the genome that contain clusters of 3-10 Ci/GLI sites that are enriched above chance expectation. Since the Ci/GLI binding sequence is highly GC rich, these sites are more likely to occur by chance in GC rich regions of the genome. Thus, to achieve an unbiased assessment of clustering likelihood, it was important to utilize a background model with a GC landscape similar to that of the native genome. Three different background models were examined (see Materials and Methods for details). The three models were compared by mapping all predicted Ci/GLI sites (MSS ≥0.75) and examining the GC content of the genomic sequence surrounding each predicted Ci/GLI site (Figure 4.2). Importantly, the randomized (Model 1) and shuffled 3-mer (Model 2) strategies significantly change (i.e., homogenize) the GC context around Ci/GLI sites, while the Flip GC/AT model (Model 3), by its nature, faithfully replicates the GC context of Ci/GLI sites in the real genome; thus, this model was selected for use.

An accurate assessment of the relative density of Ci/GLI clusters found in the native genome also requires that the background genomes contain a similar composition (number and type) of Ci sites as the native genome. After generating background genomes using the Flip GC/AT method, we noticed that the total number of predicted Ci/GLI binding sites on each chromosome was consistently reduced compared to the native Dm genome (Figure 4.3A). Left uncorrected, this deficit in total sites would lead to an artificial enrichment of clusters of Ci/GLI sites in the Dm genome

when compared to the background model. To correct for this discrepancy, we re-built background chromosomes (see detail in Materials and Methods) so that they contained the same number of each type of Ci/GLI binding site (based on matrix similarity score) found in the Dm genome (Figure 4.3B). Relative enrichment of Ci/GLI clusters in each genomic region was then assessed across the native genome by direct comparison to the 100 rebuilt background chromosomes (Figure 4.3C).

4.3.2 Ci/GLI cluster analysis in *Drosophila melanogaster*

Clusters of 3-10 Ci/GLI sites (with a maximum end-to-end distance of1000 bp) were identified in the native Dm and Dp genomes. Background modeling and background correction was performed separately for Dm and Dp. For each putative cluster, a cluster coefficient (CC) was defined as the number of Ci/GLI sites in a given genomic region divided by the average number of Ci/GLI sites in the same genomic location in 100 control genomes (schematically illustrated in Figure 4.3C). Only clusters with a CC of greater than or equal to 4 and at least one Ci/GLI site with a MSS of 0.81 or greater were chosen for subsequent analysis. These filters (1kb length; CC ≥ 4; one site ≥ MSS of 0.81) were designed to increase the likelihood that functional enhancers would be identified. As an additional stringency filter, we required that Ci/GLI site clusters be present in orthologous regions of both Dp and Dm genomes (see Materials and Methods for details). Table 4.1 lists all selected Dm clusters with a CC greater than or equal to 4 (ranked by order of Ci/GLI site density and average MSS). We sorted these results by average MSS (high to low), to strengthen the likelihood that all of the Ci/GLI sites located within any putative cluster were capable of binding Ci/GLI, and observed that sites in a known Hh-regulated enhancer of the *ptc* gene (Alexandre et al.,

1996) had the maximum average MSS of 1 (Table 4.1). In addition to this known enhancer region, we selected the next 16 putative Hh enhancer regions for functional validation (Table 4.2).

4.3.3 Functional verification of Ci/GLI-driven enhancers in a chicken neural tube assay

We first screened for possible enhancer function of the 16 novel genomic regions (Table 4.2) in the developing chicken neural tube, one of the best-studied sites of Hh signaling (Dessaud et al., 2008). In this assay, Hamburger-Hamilton stage 11 embryos are electroporated with DNA reporter constructs in which the putative enhancer is cloned upstream of a minimal promoter driving EGFP expression (see Materials and Methods). This assay has been previously used to validate enhancers for multiple signaling pathways (Lang et al., 2010; Lei et al., 2006; Oosterveen et al., 2012; Peterson et al., 2012; Timmer et al., 2001; Uchikawa et al., 2003; Vadasz et al., 2013; Wang et al., 2011). Endogenous Sonic Hedgehog (SHH) produced by the notochord and floorplate drives expression of Hh-dependent enhancers in the ventral half of the neural tube (Dessaud et al., 2008). Additionally, to further increase the sensitivity of our assay, we co-electroporated a constitutively active form of *Smoothened* (*SmoM2*) (Xie et al., 1998), which activates Hh signaling throughout the neural tube. Successful activation of Hh signaling by *SmoM2* is readily detectable as an expansion of the expression domain of the known Hh target gene, NKX6.1(Oosterveen et al., 2012; Peterson et al., 2012), on the electroporated side of the neural tube. An RFP-expressing plasmid (pCIT) was co-electroporated to confirm the success of the electroporation. For those enhancers that demonstrated apparent Hh activation (expression of the enhancer-containing construct, but not the enhancer-less construct, in the presence of

SmoM2), Ci/GLI-dependent activity was further confirmed by mutagenesis of the Ci/GLI binding sites.

Of the 16 computationally predicted enhancers tested in this way, four drove Hh-enhancer dependent expression in the chicken neural tube assay (Figure 4.4). An intronic sequence of the *invected* (*inv*) gene harbors two of these active regions, each containing a cluster of four Ci/GLI sites with MSS ≥0.81. Both regions drive expression in the presence of co-electroporated *SmoM2* and mutagenesis of the Ci/GLI binding sites abrogates this response in both cases (Figure 4.4B,C).

Two additional predicted enhancers, located near the genes *Cpr100A* and *Plc21C*, also showed expression in the chicken neural tube assay (Figure 4.4D,E). However, mutation of the Ci/GLI sites abrogated EGFP expression only in the putative *Cpr100A* enhancer (Figure 4.4D), but not in the *Plc21C* enhancer (Figure 4.4E). Thus, only the former behaved as a direct Hh target; the *Plc21C* enhancer is responsive to Hh pathway activation, but this activity does not depend upon the Ci/GLI binding sites. Thus, altogether, in addition to the top scoring, previously validated proximal *ptc* enhancer, three of the 16 novel predicted enhancers were validated by the chicken *in ovo* electroporation assay, for an overall success rate of 4/17 or 23%.

We next tested whether additional information would further improve prediction of Hh enhancers. We searched the list of clusters in Table 4.1 for regions annotated to genes that are known or likely Hh targets or that participate in Hh-regulated developmental events, and chose regions linked to *roadkill* (*rdx*), *retinal homeobox* (*Rx*), *gooseberry* (*gsb*) (Kent, 2006; Li and Noll, 1993; Ohlen et al., 1997), and two additional regions of the *patched* gene ($ptc^{-2.8}$ and $ptc^{+5.3}$) for testing. Of these five cluster regions,

only *rdx* tested positive in the chicken neural tube assay (Figure 4.4F), reflecting a similar 20% success rate.

To learn more about the sensitivity of the chicken neural tube assay, we also tested 18 clusters with Ci/GLI sites of low MSS (0.75-0.8). These may represent clusters of sites of low affinity Ci/GLI binding. The regions tested included the known enhancers regulating the *wingless* (*wg)* and *decapentaplegic* (*dpp)* loci (Table 4.3). However, none of these showed activity in the chicken neural tube.

Having identified two closely associated novel regions of the *inv* gene that both act as Hh enhancers in the chicken electroporation assay (Figure 4.4B,C), we next utilized this assay to further examine these regions. While both enhancers respond to SmoM2 co-electroporation, only one ($inv^{+18.6}$), drives EGFP expression in response to endogenous levels of Hh signaling (i.e. in the absence of *SmoM2* co-electroporation) (Figure 4.5).

Notably, though it is not in the top 16 predictions, Table 4.1 lists a third cluster in this region of the *inv* locus, lying between the two active regions tested above. Thus, we also tested a fragment spanning all three of these predicted *inv* Ci/GLI clusters, containing a total of 12 Ci/GLI binding sites (inv^{long}) (Figure 4.6A). This larger construct is activated both by endogenous SHH in the ventral neural tube and by co-electroporation of *SmoM2* (Figure 4.6B). Furthermore, a construct (inv^{long}-CiKO) containing mutations in 10 of the 12 Ci/GLI binding sites identified computationally (only the two Ci/GLI sites with lowest MSS were left intact) fails to activate EGFP expression, even when co-expressed with *SmoM2* (Figure 4.6C), confirming the Hh-dependent activity of this large complex enhancer. Further selective mutagenesis of Ci/GLI sites

within the larger fragment demonstrates that, in the absence of the *inv*$^{+16.8}$ and *inv*$^{+18.6}$ Ci/GLI clusters, the central cluster of Ci/GLI binding sites is unable to drive enhancer activity in the chicken neural tube (Construct D, Figure 4.6C).

4.3.4 Functional verification of Ci-driven enhancers in transgenic *Drosophila*

To further verify enhancer function in *Drosophila*, we generated transgenic reporter flies in which EGFP was driven by predicted enhancers and examined gene expression in two of the best-studied Hh-responsive contexts: the stage 9-13 embryo (when Hh signaling is active during development of a variety of tissues) and the anterior/posterior boundary of the larval wing imaginal disc (Alexandre et al., 1996; Ohlen and Hooper, 1997; Ramos and Barolo, 2013). The top computational hit, upstream of the *ptc* gene (Table 4.2, Figure 4.7A) has three consensus Ci/GLI binding sites and was previously shown to harbor enhancer activity (Alexandre et al., 1996). This conserved cluster was examined as a minimal fragment, (*ptc*$^{-0.6}$), which was able to respond to Hh signaling in the wing (Figure 4.7A). When the three consensus Ci/GLI binding sites were mutated, enhancer activity was abrogated (Figure 4.7B), confirming that enhancer activity directly depends upon function of the Ci/GLI binding motifs. This region was also found to have enhancer activity in a recent unbiased search for imaginal disc enhancers (Jory et al., 2012).

We next examined the other novel 16 top computationally predicted enhancers in *Drosophila* and found that three regions exhibited enhancer activity in the fly assay. Although *inv*$^{+16.8}$ and *inv*$^{+18.6}$ were both active in the chicken neural tube assay when co-electroporated with *SmoM2* (Figure 4.7B,C), only *inv*$^{+18.6}$ responded in the wing imaginal disc (Figure 4.7 E,G). This *inv*$^{+18.6}$ enhancer was also the only enhancer to demonstrate

positive activity in the chicken ventral neural tube in the absence of *SmoM2*, in response to endogenous Hh expression (Figure 4.5). When the four predicted Ci/GLI binding sites with higher MSS were mutated in $inv^{+18.6}$, this enhancer was no longer able to respond to Hh signaling in the wing imaginal disc (Figure 4.7F), demonstrating its Hh-dependent activity. A larger construct (inv^{long}), encompassing the four Ci/GLI sites in $inv^{+16.8}$, the four in $inv^{+18.6}$, and the intervening cluster of four predicted sites that was tested in the chicken assay in Figure 4.6A, was also able to drive expression in Hh-responsive cells of the wing imaginal discs of transgenic flies (Figure 4.7C). The in vivo activity of this genomic fragment depended on the predicted Ci/GLI sites (Figure 4.7D), confirming it as a direct Ci/GLI target enhancer.

In addition to confirming direct Hh-responsiveness of the *ptc* and *inv* enhancers, we also examined the other predicted enhancers in Table 4.2 in transgenic fly assays. Both *hth* and *Plc21C* showed enhancer activity in the transgenic fly assay, but neither was Hh-dependent (Figure 4.8). *Hth* exhibited a segmented expression pattern in the fly embryo, which remained unaltered after mutagenesis of the Ci/GLI binding sites (Figure 4.8A,B). *Plc21C* was expressed in the fly gut and expression persisted after mutation of the Ci/GLI binding sites (Figure 4.8C,D), consistent with the results in the chicken neural tube assay (Figure 4.4E).

Examination of the five additional Ci/GLI clusters selected from known or suspected Hh target genes yielded four potential Hh-responsive enhancers: *rdx*, $ptc^{-2.8}$, $ptc^{+5.3}$ and *gsb*. A Ci/GLI cluster in the intron of *roadkill* (*rdx)* was active at the A/P boundary of the wing imaginal disc in Hh-responsive cells (Figure 4.7H). Mutating the predicted Ci/GLI sites within this cluster abrogated its activity (Figure 4.7I). *rdx* had

previously been shown to be genetically downstream of Hh signaling (Kent, 2006), but the enhancer that mediates this response had not been identified. The *rdx* enhancer identified here also responds to Hh in the chicken neural tube assay (Figure 4.4F).

Within the *ptc* locus, two other Ci/GLI binding site clusters are computationally predicted in addition to the previously identified promoter-proximal enhancer that topped the list. The first of these, $ptc^{-2.8}$, is found 2.8 kb upstream of *ptc*, and contains 5 predicted Ci/GLI binding sites. When examined in the wing imaginal disc, $ptc^{-2.8}$ responds with a stripe of expression largely overlapping Ptc positive cells (Figure 4.7J). Upon mutation of the predicted Ci/GLI binding sites in this novel enhancer, its ability to respond to Hh is greatly reduced (Figure 4.7K). A second cluster of Ci/GLI sites in the first intron of *ptc* ($ptc^{+5.3}$) is also predicted. This putative enhancer contains 7 predicted Ci/GLI binding sites, one of which matches the optimal consensus site recognized by Ci/GLI. In flies containing this transgene, *ptc*-like reporter gene expression is seen in the embryo (Figure 4.7L), but not the wing disc (data not shown). Two stripes of enhancer expression are detected, proximal to cells secreting Hh ligand, marked by En, in all segments of the embryonic ectoderm. After mutation of the predicted Ci/GLI binding sites contained within this enhancer, the segmentally repeated stripes are lost (Figure 4.7M).

Finally, a region with several clusters of Ci/GLI binding sites was identified downstream of the *gooseberry (gsb)* coding region. *Gooseberry*, a segment polarity gene, is part of the Hh-Wnt segmentation network, but no direct Ci/GLI target enhancer has been identified (Li and Noll, 1993). One of the known enhancers of *gsb*, which does not appear to be regulated by Ci/GLI, is 5' of the gene (Bouchard et al., 2000). The 3'

enhancer identified by our analysis contains five predicted Ci/GLI binding sites and is active in segmental stripes in the embryonic ectoderm of transgenic *Drosophila*, posterior to each stripe of Hh-secreting cells at stage 11 (Figure 4.7N). Upon mutation of the Ci/GLI binding sites, activity is attenuated, suggesting that the *gsb* enhancer requires direct Ci/GLI input in order to respond to Hh signaling in the embryo (Figure 4.7O).

Overall, the fly assay functionally verified six Hh-dependent enhancers out of 22 tested, for a success rate of 27%. The genomic locations of these enhancers, relative to the gene locus, are presented in Figure 4.9. One additional enhancer, *cpr100A*, was demonstrated to be Hh-dependent in the chicken, but had no activity in the fly assay; thus, it must be considered a potential regulatory element. This result suggested that *cpr100A* might have been a false-negative in the fly assay, and prompted us to examine it, along with all of the other predicted enhancers, in a third site of Hh signaling, the adult testis. Although the testis depends on Hh signaling, none of the predicted enhancers were active in this tissue. It is possible, however, that the *cpr100A* cluster (or any other predicted enhancer that is negative in the chicken and/or fly assays) may be active in another tissue that was not examined (Michel et al., 2012). Altogether, both assays established 7/22 (31.8%) of tested Ci/GLI clusters as Hh enhancers, six of which are novel (the potential *cpr100A* element is not included in this count).

4.4 Discussion

Homotypic clustering of transcription factor binding sites has been observed in multiple settings and has been successfully used to identify potential enhancers (Berman et al., 2004; Gotea et al., 2010; Lifanov et al., 2003; Markstein et al., 2002;

Suryamohan and Halfon, 2015). Since all but one of the known *Drosophila* Hh-driven enhancers contain two or more Ci/GLI sites, we assessed the extent to which clustering of Ci/GLI sites can be used to predict the location of Hh-dependent enhancers, a question that has not previously been directly tested. To do this, we utilized a background correction method that preserves local nucleotide topography to allow us to identified genomic regions that appear to have unusually dense Ci/GLI binding site representation and tested the extent to which these regions can function as Hh-dependent enhancers.

To establish background genomes for comparison of GLI density, we used a strategy that randomly flips each base to its complimentary partner. This approach maintains the GC/AT landscape of the native *Drosophila* chromosomes. Overall, only 43% of the *D. melanogaster* contains G or C bases while the consensus Ci/GLI binding site itself is 67% GC rich (Hallikas et al., 2006; Keightley et al., 2009). Distribution of GC content has been strongly correlated with gene density and other genomic features and the importance of maintaining the original properties of the native sequence when generating a background comparison has been discussed previously (Fitch, 1983). Other background generation methods that preserve dinucleotide frequencies also exist (Coward, 1999; Fitch, 1983). Additional comparisons would be needed to determine which background strategy best strengthens enhancer detection.

The success rate of functional enhancer identification based on the approach used here was 23%, suggesting that clustering of Ci/GLI sites alone is not sufficient to effectively predict Hh-regulated enhancers. However, this success rate increased to 80% when examining Ci/GLI clusters associated with known or suspected Hh target

genes. Together, these data indicate that Ci/GLI clustering is not, by itself, an effective means to predict Hh-regulated enhancers. While some Hh enhancers can be identified by virtue of Ci/GLI homotypic clustering, not all homotypic clusters function as enhancers. Since one of the previously identified Hh enhancers (in orthodenticle) only has one Ci/GLI site (Ramos and Barolo, 2013), it is also clear that the presence of clustered Ci/GLI sites is not a requirement for functional Hh enhancers. However, in the context of additional information, clustering can be used as one criterion to predict enhancers within a suspected Hh target gene locus. Future studies will be necessary to determine whether the presence of multiple Ci/GLI sites are more effective predictors of Hh-regulated enhancers associated with putative Hh target genes, or whether a single Ci/GLI site is equally likely to drive Hh-dependent target gene expression.

Given that Ci/GLI binding site clustering alone is not sufficient to identify Hh-regulated enhancers, this raises the question: what is an effective method to identify Hh-regulated enhancers? One possibility is to pair Ci/GLI binding sites with sites for other transcriptional co-activators or co-repressors. *De novo* motif analysis has been performed previously as part of ChIP-chip analysis of GLI repressor binding in the developing limb (Vokes et al., 2008). More recent studies suggest that GLI proteins cooperate with SOXB1 proteins to drive Hh-regulated gene expression during spinal cord development (Oosterveen et al., 2012; Peterson et al., 2012). However, specific co-factor identification may yield only tissue-specific Hh-regulated enhancers. Thus, other approaches include: 1) examining Ci/GLI binding site association with active or repressive chromatin modifications, which has been recently used to investigate Hh-regulated enhancers in the developing neural tube (Nishi et al., 2015), and 2)

investigating Ci/GLI binding site location near sites of open chromatin using techniques such as DNAse I hypersensitivity and FAIRE (Giresi et al., 2007; McKay and Lieb, 2013). It is likely that a combination of these methods will be required to effectively identify a more complete set of Hh-regulated enhancers on a genome wide basis

One intriguing finding from this work is the identification of multiple discrepancies between the chicken neural tube and transgenic fly assays (Table 4.2). These data emphasize the importance of testing putative enhancers in diverse assay systems to provide several different contexts in which an enhancer can show activity. The chicken neural tube assay is a quick and inexpensive strategy that, in a large-scale study, could improve throughput. It has been successfully used previously to identify Hh-regulated mouse enhancers (Oosterveen et al., 2012; Peterson et al., 2012), and is used here to validate Hh-regulated fly enhancers. However, because some enhancers may require additional species-specific information that is not present in the chicken neural tube, false negative calls are a limitation of this assay. Further, the requirement for context-specific information may also restrict the utility of this assay in the identification of general Hh-regulated enhancers (Vokes et al., 2007). Along these lines, analysis of 18 clusters containing Ci/GLI sites of lower predicted affinity, including the known Hh enhancers in the *wg* and *dpp* loci (Müller and Basler, 2000; Ohlen and Hooper, 1997), showed no activity in the chicken neural tube (Table 4.3). Thus, this assay may only detect Hh enhancers with high affinity Ci/GLI binding sites, thereby missing some true positives (Ramos and Barolo, 2013). Nevertheless, the assay can be useful to dissect enhancer activity in the context of a complex developing tissue (Figure 4.5).

The computational study presented here can be compared with a recent analysis of potential Ci/GLI-driven enhancers in *Drosophila*, by Biehs et al., who fused Ci^{ACT} (activator) and Ci^{REP} (repressor) proteins with DNA adenine methyltransferase (Dam) domains to define chromatin regions in stage 10-11 embryos that are occupied by Ci/GLI in vivo (Biehs et al., 2010). That study listed 1743 sites bound by Dam-Ci fusion proteins; of these, 55 sites (3%, listed in Table 4.4) were represented in clusters that were selected by our computational analysis. This limited overlap is likely due to two factors. First, since the computational study was limited to analysis of larger clusters, enhancers that are driven by one or two Ci/GLI sites were not selected, by design. Second, because the DamID study was performed in 2-6 hour embryos, Ci/GLI binding events were likely limited to chromatin regions that were accessible at that developmental stage. Of the seven previously known Hh/GLI-regulated enhancers, the DamID approach identified Ci/GLI binding to two (*ptc* and *wg*), while the computational strategy described here detected three (*ptc*, *wg* and *knot*). The other four previously known enhancers (*stripe, hairy*, *dpp* and *orthodenticle*) were not detected computationally because those enhancers have only two Ci/GLI sites (our filters selected clusters of 3-10). Of the new enhancers functionally confirmed in our study, none were found to harbor protected regions in the DamID assay. Biehs et al. used expression assays to identify 147 genes whose expression appeared to correlate with Hh signaling activity. They then asked, of these 147 genes, how many had protected regions within or adjacent to the transcription unit? Protected regions were identified as DamAct or DamRep protection and consisted of a total 2108 protected regions. They identified 52 genomic regions that were DamID-protected and showed expression

changes when Hh signaling was modulated. Thus, 35% of the genes that appear to be targets (as assessed by their expression modulation) showed some DamID protection, but only 2.5% of the total DamID protected regions were found to be probable Hh targets (Biehs et al., 2010). Four of these 52—but none of the validated enhancers—can be found in the list of 55 sites common to the two studies.

An important aspect of the present study is that the direct Hh dependency of all enhancers was verified by Ci/GLI binding site mutagenesis. While expression assays such as those used by Biehs et al. clearly demonstrate a Hh response, they do not establish whether this response is direct or indirect and do not confirm that the response is mediated through the Ci/GLI binding sites in the candidate enhancers. Indeed, of the top 17 clusters detected computationally, we found four direct targets and two additional enhancers that showed apparent expression in *ptc*-expressing cells, but this expression persisted after mutation of the Ci/GLI sites (Figure 4.8) suggesting that other factors might be responsible for this enhancer activity. This raises a cautionary note about assigning potential Hh, or any signaling cascade, responsiveness in the absence of functional verification (Halfon et al., 2011).

Using homotypic Ci/GLI site clustering as a criterion together with functional analyses, we have doubled the number of previously verified *Drosophila* Ci/GLI-dependent enhancers, including multiple distinct enhancers that regulate a single Hh-responsive gene (i.e., *ptc*, *inv*, and *gsb*). Further testing of other candidate clusters identified in this study might further enlarge the pool of known Hh-responsive enhancers that are active in diverse tissues and organs, providing a robust substrate for the future dissection of the rules that underlie context-specific enhancer function.

4.5 Materials and Methods

Computing resources

Except where otherwise indicated, all computational steps were performed using custom Perl scripts, which are available for download at https://github.com/um-gurdziel/GurdzielUdagerLorberbaum2015. Overlap between coordinates in bed file format were performed using the UCSC Table Browser.

Definition of putative Ci/GLI binding sites

A mono-nucleotide distribution matrix for Ci binding sites, derived from *in vitro* competitive DNA binding assays with recombinant Ci protein and labeled oligonucleotides, was obtained via the Genomatix Software Suite (www.genomatix.de; Genomatix, Germany) (Hallikas et al., 2006). The consensus index vector for such a matrix reflects the degree of nucleotide preference at each position; values range from 0, indicating equal preference for any of the four nucleotides, to 100, indicating strict preference for a single nucleotide (Quandt et al., 1995). The matrix similarity score (MSS) for a given site is calculated as the ratio of its matrix-vector product to that of the consensus site, as described previously (Quandt et al., 1995), and MSS values range from 0 to 1 (where 1 equals an exact match to the consensus site). The first nine of the eleven positions in the Ci matrix have consensus index vector values greater than 70, suggesting that they contain a high degree of specific information about potential Ci binding. Thus, these matrix positions were used to define a set of 211 9-mers (422 in sense and antisense directions) that pass a minimum level (0.75) of overall matrix similarity (i.e. with a MSS ≥ 0.75) to the optimal consensus Ci site (GACCACCCA)

(Table 4.5) (Hallikas et al., 2006; Quandt et al., 1995) and also contain concordant (C and C or G and G) nucleotides in the 4^{th} and 6^{th} position, which are critical for Ci binding (Winklmayr et al., 2010).

Identification and annotation of predicted Ci binding sites in genomic sequence

Genomic sequence files (chromFa) for *D. melanogaster* (Dm) and *D. pseudoobscura* (Dp) were downloaded from UCSC Genome browser (genome.ucsc.edu) build dm3 (Celniker et al., 2002; Fujita et al., 2011; Richards et al., 2005). The genomic coordinates of predicted Ci/GLI binding sites were identified for chr2R, chr2L, chr3R, chr3L, chr4, and chrX (build dm3); and chr2, chr3, chr4 and chrX (build dp3). Each putative Ci/GLI binding site was annotated for nearest gene/transcript, distance to nearest gene/transcript, and associated gene/transcript feature transcript using refFlat files obtained from UCSC Genome Bioinformatics. Ci/GLI clusters were defined as regions containing at least three and at most ten putative Ci/GLI binding sites within a maximum distance of 1000 base pairs (bp) (measured from the outside ends of the flanking sites (Fujita et al., 2011; McQuilton et al., 2012). Predicted sites were also annotated with respect to the nearest CTCF boundary region (Holohan et al., 2007). Cluster regions that contained predicted Ci binding sites that mapped to exons or repeat regions were excluded. Repeat regions often have regulatory function (Sawaya et al., 2013; Taher et al., 2015). However, testing the regulatory activity of Ci binding motifs in repetitive sequences, and the effect of their clustering in these regions, was beyond the scope of this study.

Background Modeling

To identify regions of the genome that exhibit a higher density of Ci/GLI sites than would be expected by chance, we compared the actual distribution of Ci/GLI sites to a randomized background model. Three different modes of background modeling were examined. For Model 1 (Random), all bases in the genome were randomized, as was done in a previous analysis of clustered binding sites for Suppressor of Hairless (Rebeiz et al., 2002). For Model 2 (Shuffle 3mer), the genome was parsed into contiguous 3-mers and these were then shuffled to create the background. In Model 3 (Flip GC/AT), each base was randomly flipped between itself and its complementary base pair (e.g., G will randomly become G or C; A will become A or T; C will become C or G; T will become T or A). On the basis of the data shown in Results, only the Flip GC/AT model generates background genomes that most closely represent the GC content surrounding Ci/GLI sites in the native genome. Since GC rich Ci/GLI sites will occur by chance more often in GC rich than AT rich regions, use of a randomization model that homogenizes the AT/GC landscape would artificially reduce the density of expected Ci/GLI sites in GC rich areas and increase this density in AT rich regions. Therefore, using the Flip GC/AT strategy, background models were generated separately for the Dm and Dp genomes for comparison to each native genome.

Generation of artificial genomic sequence and random genomic distributions of binding sites

On a chromosome-by-chromosome basis, 1000 sets of background genomic sequences were generated using the Flip GC/AT method. However, base flipping

resulted in fewer Ci/GLI sites in the randomized chromosomes, relative to the native Dm or Dp genome. To correct for this, putative Ci/GLI binding sites were identified in each of the 1000 background genomic sequences and the genomic coordinates of each site was recorded. Site motifs, tagged with their location coordinates, were pooled into a master list of possible site positions. This master list was used to re-create 100 background chromosomes for each chromosome, such that each background chromosome contained the same composition of Ci/GLI sites (overall number and motif) as the native Dm or Dp chromosome (see Results).

Assessment of relative Ci/GLI binding site clustering

Ci/GLI site clusters were defined as regions containing at least three and at most ten putative Ci/GLI sites within a maximum distance of 1000 base pairs (bp; measured from the outside ends of the flanking sites). The genomic coordinates of each cluster were cataloged, and clusters were subsequently filtered for the presence of at least one predicted binding site with a MSS ≥0.81. This was done to decrease the number of clusters comprised entirely of low scoring sites, substantial portions of which are predicted to be non-functional. Clusters that contained exon or repeat elements were excluded. Clusters for which the Ci/GLI binding sites themselves accounted for more than 25% of the end-to-end cluster length were also excluded, since the majority of such clusters were composed of repetitive sequence. For each cluster, the number of binding sites expected to be present by chance for that specific genomic region was determined from 100 control reconstructed genomes as described in Results. A clustering coefficient (CC) was defined as the number of Ci/GLI sites observed in a

given interval of the native genome (at a given location) divided by the average number of Ci/GLI sites in the same region of the background genome (at the same location). To enrich for clusters likely to represent enhancers, we selected a CC cutoff of four which captured all of the previously known clustered Hh enhancers. Importantly, the CC score was used as a filter, and not as a ranking tool.

Orthologous enrichment of Ci clusters

Clusters were identified and annotated in the Dp genome exactly as described above for Dm. Background modeling for the Dp genome was done by Flip GC/AT; 1000 randomized genomes were generated and corrected as outlined above for number and affinity class to make 100 randomized, corrected Dp genomes for comparison to the native Dp genome. Clusters identified in the Dp genome were selected according to the same criteria as for the Dm genome (cluster size ≤ 1000; 3-10 Ci/GLI sites; CC ≥ 4; at least one site with MSS ≥ 0.81). The coordinates for enriched clusters of Ci/GLI binding sites (CC ≥ 4) were determined for Dm and Dp and compared using the LiftOver tool available from UCSC Genome Bioinformatics (Fujita et al., 2011). All clusters that were present in orthologous positions of the Dm and Dp genomes (i.e., with an overlap of one or more bases, irrespective of sequence identity) were selected for further analysis.

Cloning of putative enhancer regions for testing

Putative enhancer regions in the Dm genome were visualized in the UCSC Genome Browser, and using the Conservation track (12 Flies, Mosquito, Honeybee, Beetle Multiz Alignments & phastCons Scores), the ends of an individual enhancer element were

extended to include contiguous highly conserved sequence (Blanchette et al., 2004). Putative enhancers were amplified from w^{1118} genomic DNA using template-specific PCR primers (Table 4.6). A CACC extension was added to the end of one primer to facilitate directional cloning. PCR fragments were cloned into the pENTR/D-TOPO vector using the standard kit (Invitrogen) and then shuttled into either Ganesh-G2 (Swanson et al., 2008) or HP-desteGFP (Boy et al., 2010) vectors using the Gateway® cloning system (Invitrogen). Ci binding site mutations (C4A) were introduced by overlap extension PCR, as previously described (Swanson et al., 2010a). QuikChange mutagenesis (Stratagene) was also used to mutate some Ci binding sites. pCIT was generated by replacing eGFP in pCIG (Megason and McMahon, 2002) with TdTOMATO, which was cloned into the location between the third PmlI site and the NotI site in pCIG. *SmoM2*-pCIT was generated by cloning rat *SmoM2* into the XhoI and ClaI sites of pCIT.

Drosophila transgenesis

Transformation was achieved by injection of w^{1118} or *ZH-attP-86Fb* embryos, essentially as described previously (Bischof et al., 2007; Rubin and Spradling, 1982). A current protocol is available at: http://sitemaker.umich.edu/barolo/injection. For w^{1118} transgenesis, at least three independent lines were examined; one or more lines were examined for *ZH-attP-86Fb* transgenesis.

Drosophila tissue analysis

Since Hh is active in a variety of tissue contexts in the embryo (brain, gut, muscle, segmental stripes etc.), we utilized embryos at stages 9-13 to gain an unbiased view of all of these contexts. Additionally, we specifically examined the wing imaginal disc since this is a well-known and well-characterized expression domain for Hh signaling. Of the 22 genes selected for analysis (Table 4.2), 17 are expressed in the embryo or imaginal disc (Hammonds et al., 2013; Tomancak et al., 2002; Tomancak et al., 2007). There are no data on two (CG5475, CG4704) and three others (beat-IV, BDGP, HGTX) are not reported to be expressed in these sites, but these have been incompletely studied. For imaginal disc analysis, 3rd instar wandering larvae were collected from vials, and discs were dissected fresh and fixed in 4% paraformaldehyde. For embryo analysis, embryos were collected in 6-hour batches at 25°C, dechorionated in 100% bleach, fixed in 4% paraformaldehyde, and devitellinized by shaking in methanol and heptane.

Chicken in ovo electroporations

Chicken neural tube electroporations were performed essentially as described previously (Tenzen et al., 2006). Briefly, 500 ng/ml of reporter vector and 500 ng/ml of either pCIT or *SmoM2*-pCIT was dissolved in PBS with 50 ng/ml of Fast Green and injected into the neural tubes of Hamburger-Hamilton stage 10–12 chicken embryos. Approximately 48 hours following electroporation embryos were recovered and fixed in 4% paraformaldehyde for subsequent immunofluorescent analysis. Fertile eggs were obtained from the Michigan State University Poultry Farm.

Immunofluorescence and microscopy

Drosophila embryos and imaginal discs were blocked with 10% BSA in phosphate-buffered saline (PBS) with 0.1% Triton X-100. The following primary antibodies were used overnight at 4°C: rabbit anti-GFP IgG antibody (1:200; Life Technologies A11122), mouse anti-Ptc (1:50, DSHB; APA1) and mouse anti-En (1:50, DSHB; 4D9). Samples were then incubated in the following secondary antibodies for 2 hours at room temperatures, Alexa Fluor 488-conjugated goat anti-rabbit IgG antibody (1:2,000; Life Technologies A11008) and/or Alexa Fluor 468-conjugated goat anti-mouse IgG antibody (1:2,000; Life Technologies A11004). Embryos and imaginal discs were mounted on glass slides using ProLong Gold with DAPI and imaged on an Olympus BX-51 upright microscope, Nikon A1 confocal with Ti-E microscope or Olympus FluoView 500 Laser Scanning Confocal Microscope. For direct comparisons, wild type and mutant constructs were processed in parallel including being imaged on the same day, using the same exposure settings.

Immunofluorescent analyses of chicken neural tubes were performed essentially as described previously (Jeong and McMahon, 2005). The antibodies used were as follows: 1:20 Mouse IgG1 anti-NKX6.1 (DSHB; F55A10). DAPI (Life Technologies) was used at a dilution of 1:30,000. All secondary antibodies (Alexa Fluor; Life Technologies) were used at a dilution of 1:500. Primary antibodies were incubated overnight at 4°C, followed by incubation with secondary antibodies for one hour at room temperature. Images were collected with a Leica SP5X confocal microscope.

4.6 Acknowledgements

Like Chapter 2, this work was highly collaborative. Katherine Gurdziel and Aaron Udager, former graduate students in the Gumucio lab, completed all of the computational work in this chapter, making the in vivo analysis possible. Jane Song, a former technician in Ben Allen's Lab completed the chicken neural tube assays. In addition, Neil Richards, David S. Parker, Lisa A. Johnson, and Scott Barolo all significantly contributed to this work.

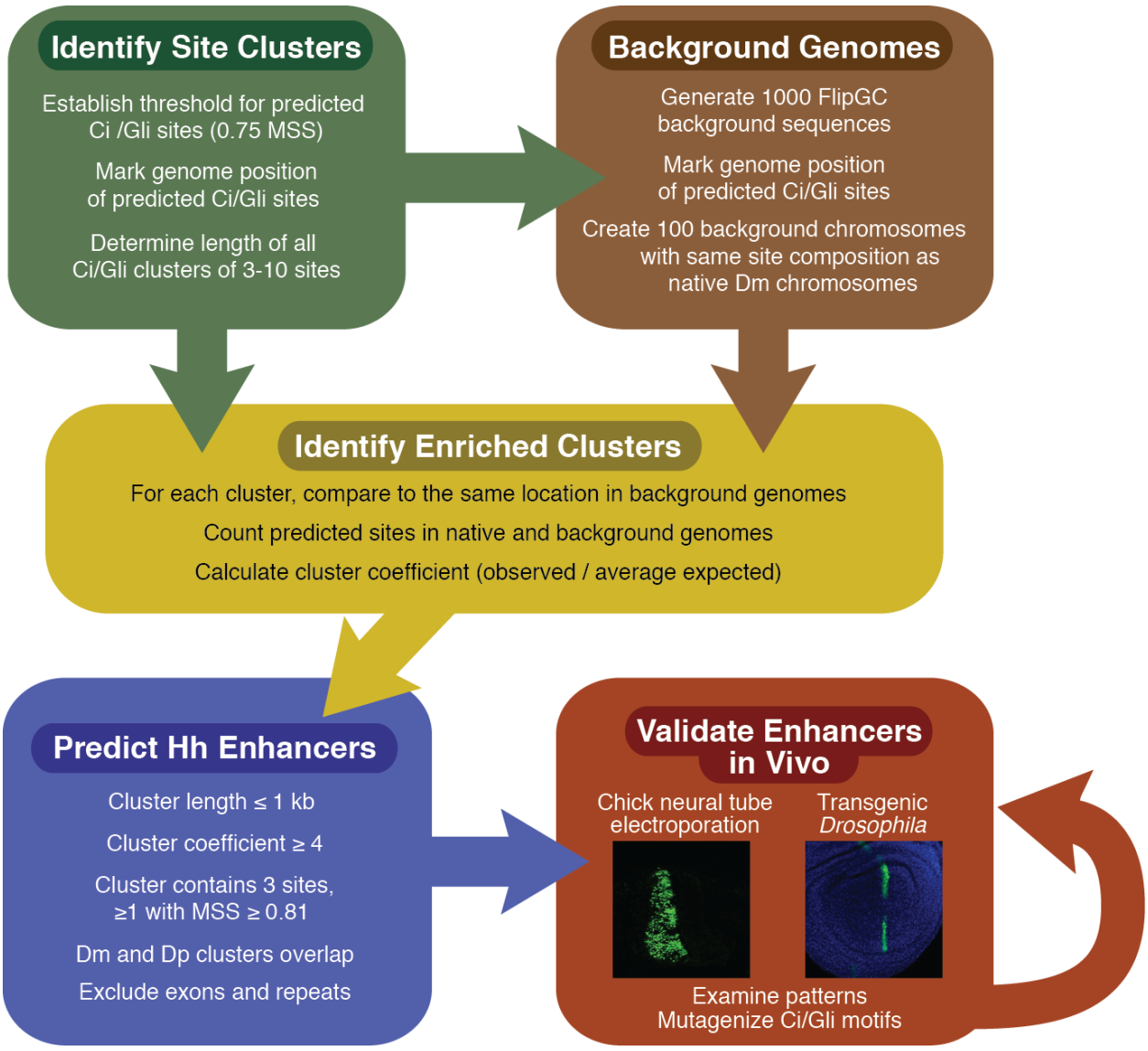


Figure 4.1 Pipeline for detection and validation of Hh-responsive enhancers

Ci/GLI cluster identification and background genome generation were performed as outlined in Figure 4.3. The cluster (CC) for a given genomic region was calculated as the total number of sites observed in the Dm or Dp genome (observed) divided by the average number of sites per background genome for that species (expected). Clusters of Ci/GLI sites with a (CC) ≥ 4 were further filtered as follows: Clusters were required to contain at least one Ci/GLI site of ≥0.81 MSS; Dm Clusters were required to overlap in position (but not sequence) with a cluster in Dp; Clusters in exon or repeat regions were excluded. The entire table of selected clusters, sorted by chromosomal location, is provided in Table 4.4. The list of clusters was then ranked by average MSS of the predicted Ci/GLI sites and the top 17 were examined functionally (these included 16 novel hits and one known enhancer, *ptc*-0.6). The Hh-responsive enhancer activity of genomic regions containing selected clusters was functionally evaluated by means of a transgenic fly assay as well as by chicken neural tube electroporation. For genomic

regions that showed apparent Hh responsiveness, Ci/GLI sites were mutated and re-assayed to confirm direct Ci/GLI regulation.

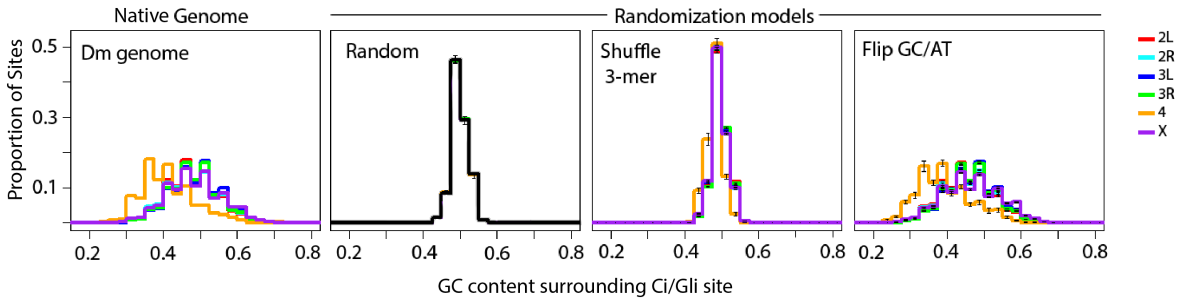


Figure 4.2 Assessment of GC content surrounding Ci/GLI sites in the *Drosophila melanogaster* genome

Proportion of sequence that is GC in the 50 bp surrounding predicted Ci/GLI sites on each Dm chromosome is displayed in the left panel. Chromosomes are color-coded: 2L (red); 2R (light blue); 3L (blue); 3R (green); 4 (yellow), and X (purple). Notably, all *Drosophila* chromosomes have similar distributions of GC content surrounding the predicted existing Ci sites, except for Chromosome 4, which is considerably more AT rich in regions surrounding predicted Ci sites (yellow line). Three different models (Flip GC/AT, Shuffle 3-mer and Random, see Materials and Methods for details) were used to create three background sequences for each chromosome and the GC content in 50 bp surrounding each Ci/GLI site was compared among the models. Black is used for the randomized model since all chromosomes collapse on the same distribution. Error bars show standard error of the mean for the 100 chromosomes in each model. Only the Flip GC/AT model recapitulates the GC profile of the native Dm genome.

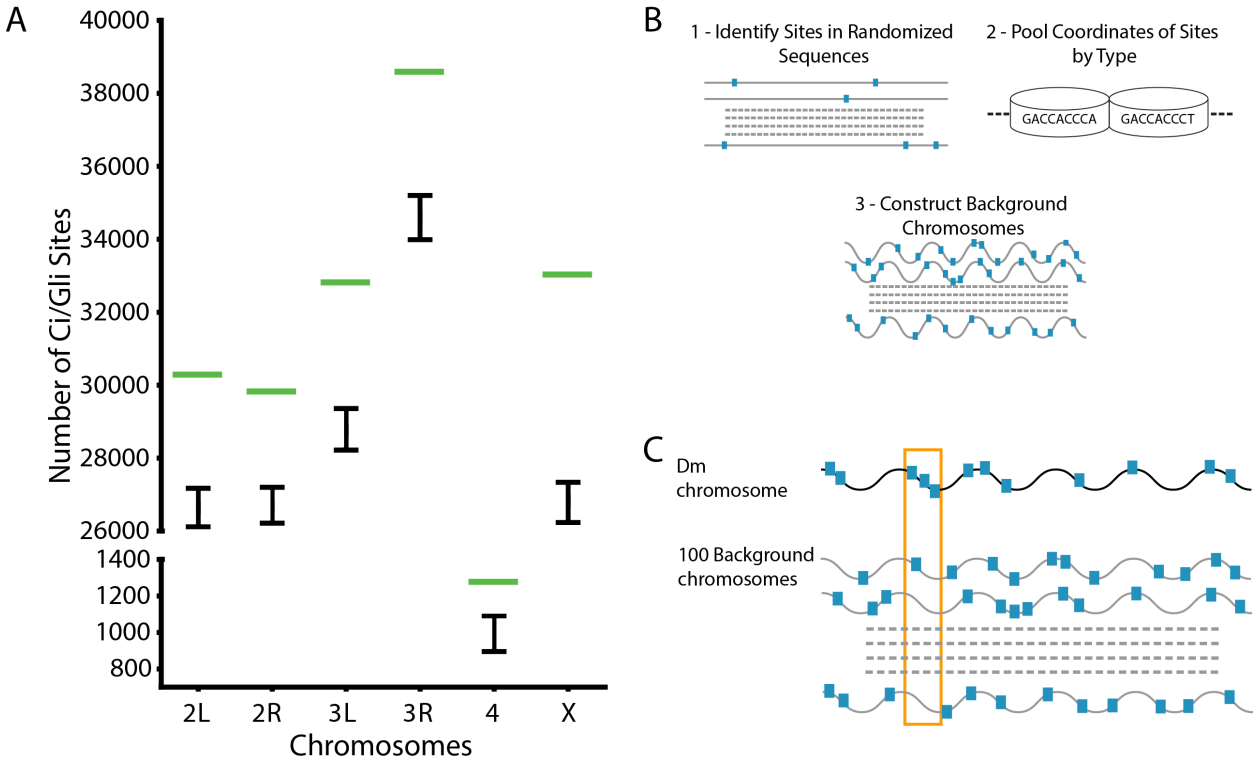


Figure 4.3 Construction of background genomes and determination of cluster enrichment

(A) The actual number of predicted Ci/GLI sites (≥0.75 MSS) determined in each Dm chromosome is shown by the green lines. The Flip GC/AT method was used to create 1000 background sequences and the number of predicted Ci/GLI sites was tallied for each sequence. Box plots show that randomized chromosomes contain substantially fewer predicted Ci/GLI sites. Brackets represent the range in total number of Ci/GLI sites across the background sequences for each chromosome. (B) To correct for the depleted number of predicted Ci/GLI sites and create background chromosomes that would closely mirror the native Dm genome, the location (coordinates) and type (sequence) of all predicted Ci/GLI sites in each of the 1000 background sequences were recorded and pooled. Background genomes were then constructed by randomly selecting coordinates from the pools so that the composition (number and site type) matched that of the corresponding Dm chromosome. (C) Enrichment of clusters of 3-10 Ci/GLI sites relative to the background chromosomes was then determined. The example shows analysis of enrichment for clusters of 3 Ci/GLI sites (blue boxes). The Dm chromosome (black line) is compared with 100 background chromosomes (grey lines); the diagram shows only three of the 100 background chromosomes. In a moving window, each group of three Ci/GLI sites was delineated in the Dm chromosome (one such cluster is outlined in orange) and the average number of Ci/GLI sites was determined within that same genomic space in each of the 100 background chromosomes. The cluster outlined by the orange box is considered enriched if the average number of sites in the Dm chromosome is ≥4 fold more than the average number of Ci/GLI sites per background chromosome.

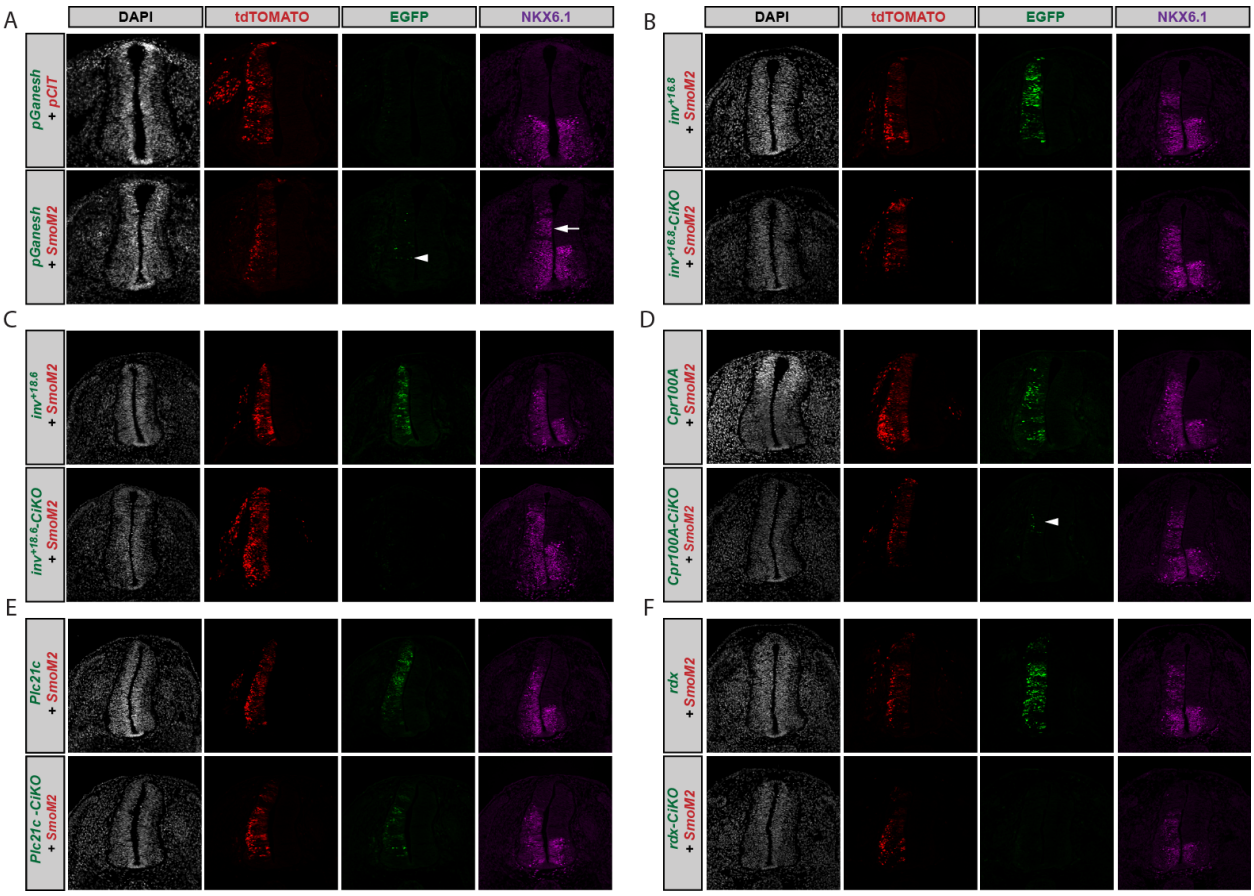
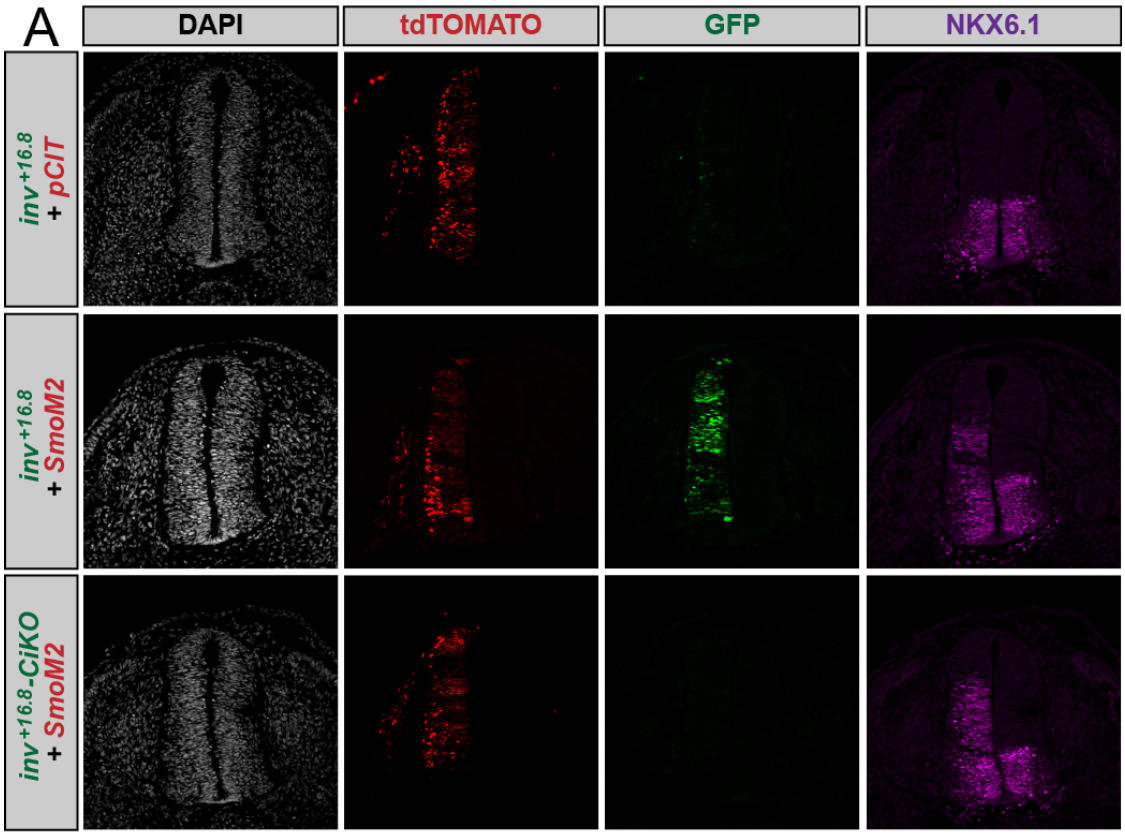


Figure 4.4 Validation of predicted Hh-responsive enhancers in the chicken neural tube

Transverse sections of Hamburger-Hamilton stage 21-22 chicken embryos are shown. DAPI (grayscale, far left column) depicts nuclei. tdTOMATO (red, middle left column) marks cells electroporated with pCIT or SmoM2. GFP (green, middle right column) reports enhancer activation. Anti-NKX6.1 antibody staining (magenta, far right column) denotes Hh-responsive cells. (A) Chicken embryos co-electroporated with an enhancerless pGanesh construct (containing only an Hsp70 minimal promoter) and either pCIT or a constitutively active SmoM2. An arrowhead (middle right column; bottom row) depicts a few GFP positive cells in pGanesh electroporated embryos. Note the ectopic NKX6.1 expression (far right column) indicative of overactive Hh signaling in electroporated cells (white arrow). (B-E) Candidate Hh-responsive *inv*+16.8 (B top row), *inv*+18.6 (C top row), *cpr100A* (D top row), and *plc21C* (E top row) constructs all exhibit GFP expression in cells in which Hh is activated by co-electroporation of SmoM2. However, chicken embryos co-electroporated with SmoM2 in combination with a Ci/GLI-binding deficient mutant (CiKO) of each candidate (bottom rows) show a complete absence of GFP expression in the case of *inv*+16.8-CiKO (B) and *inv*+18.6-CiKO (C), despite ectopic NKX6.1 expression in both conditions (far right column). *cpr100A*-CiKO (D) has a greatly diminished expression pattern with only a few GFP positive cells (white arrowhead) remaining (middle right column; bottom row). *plc21C*-CiKO (E) does not show loss of GFP expression, indicating that it is not a direct Hh target, since its response to Hh signaling is not Ci/GLI dependent. *rdx* (F top row) GFP expression

corresponds to Hh expressing cells and shows no expression once Ci/GLI sites are mutated (*rdx*-CiKO bottom row).


A
DAPI
tdTOMATO
GFP
NKX6.1
inv+16.8 + pCIT
inv+16.8 + SmoM2
inv+16.8-CiKO + SmoM2

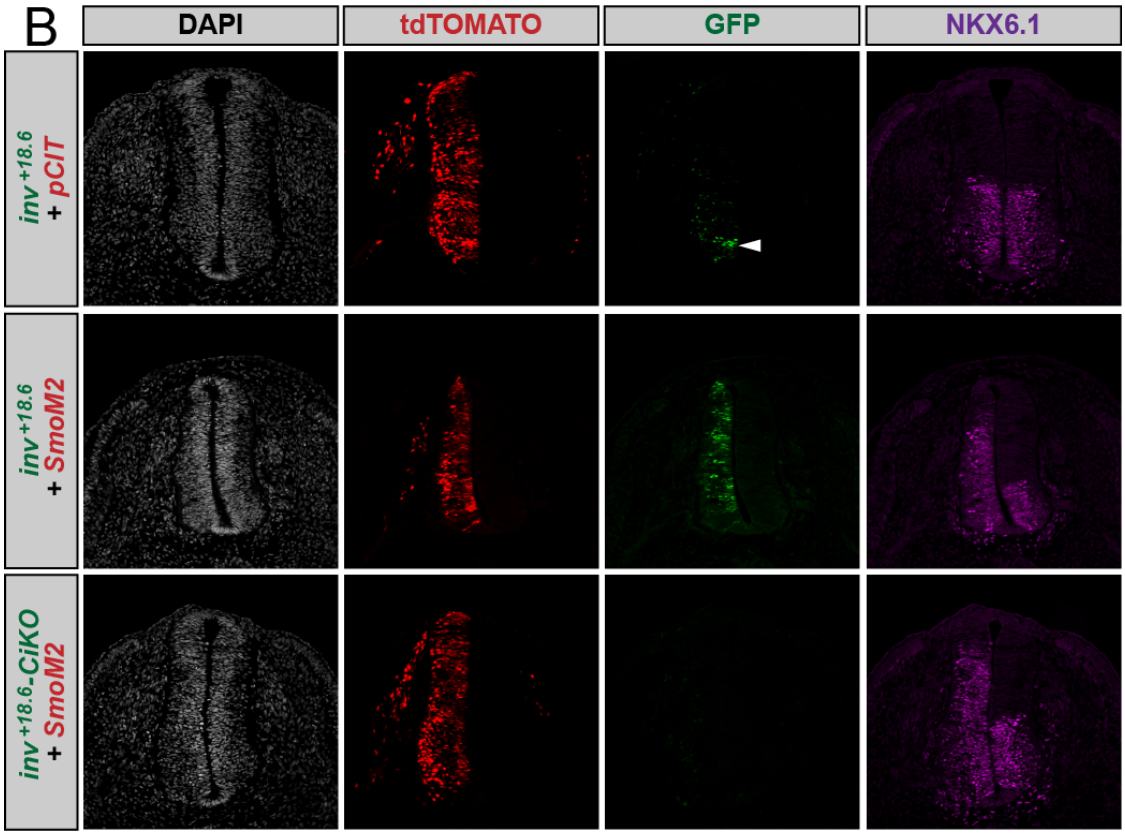

B
DAPI
tdTOMATO
GFP
NKX6.1
inv+18.6 + pCIT
inv+18.6 + SmoM2
inv+18.6-CiKO + SmoM2

Figure 4.5 Endogenous expression of *inv*+16.8 and *inv*+18.6 in the chicken neural tube

Transverse sections of Hamburger-Hamilton stage 21-22 chicken embryos are shown. DAPI (grayscale, far left column) depicts nuclei. tdTOMATO (red, middle left column) marks cells electroporated with pCIT. GFP (green, middle right column) reports enhancer activation. Anti-NKX6.1 antibody staining (magenta, far right column) denotes Hh-responsive cells. (A) Chicken embryos electroporated with *inv*+16.8 show no GFP expression in the chicken neural tube. (B) Chicken embryos electroporated with *inv*+18.6 exhibit GFP expression (white arrowhead).

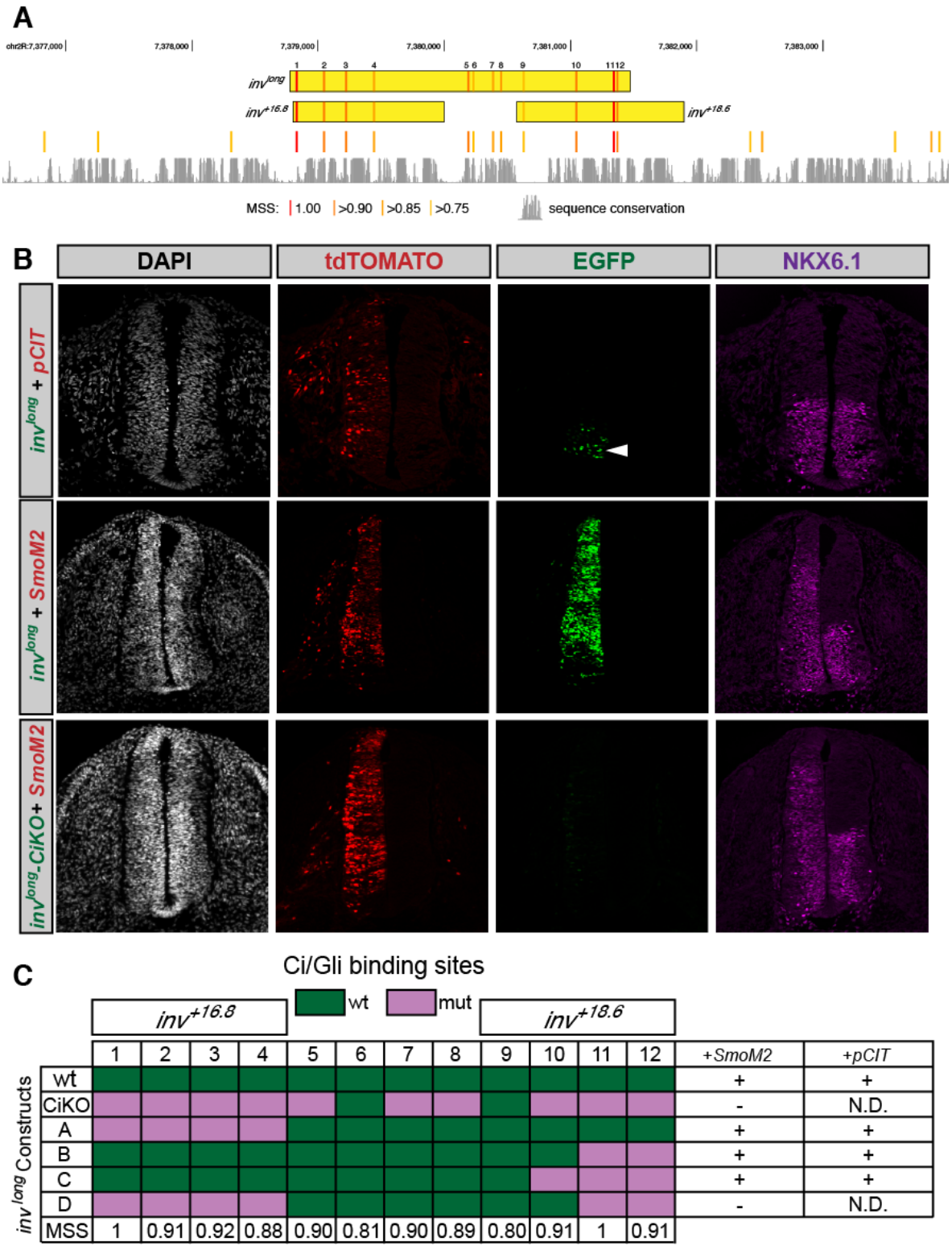


Figure 4.6 Expression of a complex *inv* enhancer in the chicken neural tube and *Drosophila* wing imaginal disc

(A) Genomic landscape of the *inv* locus depicting the *inv*long, *inv*+16.8 and *inv*+18.6 constructs. Ci/GLI binding sites are shown as red/orange bars; the intensity of red coloration indicates the MSS. Sequence conservation is indicated by the track at bottom of the panel. (B) Transverse sections of Hamburger-Hamilton stage 21-22 chicken embryos are shown as in Fig. 5. DAPI (gray, far left column) depicts nuclei. tdTOMATO (red, middle left column) marks cells electroporated with pCIT or SmoM2. GFP (green, middle right column) reports enhancer activation. Anti-NKX6.1 antibody staining (magenta, far right column) marks Hh-responsive cells. The *inv*long (top row) enhancer demonstrates GFP expression in the ventral neural tube (white arrowhead). The expression of *inv*long is strengthened and broadened with co-electroporation of SmoM2 (middle row). Mutagenesis of Ci/GLI binding sites demonstrates that enhancer activity is Ci/GLI dependent (bottom row). (C) Tabulation of activity in the chicken neural tube of invlong constructs containing different Ci/GLI site compositions. Green boxes indicate wild type Ci/GLI sequences; purple boxes indicate mutated Ci/GLI sites. Constructs that have functional Ci/GLI sites that correspond to *inv*+18.6 (construct A) or *inv*+16.8 (construct B and C) exhibit GFP expression in the neural tube. However, the central Ci/GLI binding sites are insufficient to drive enhancer activity alone (construct D).

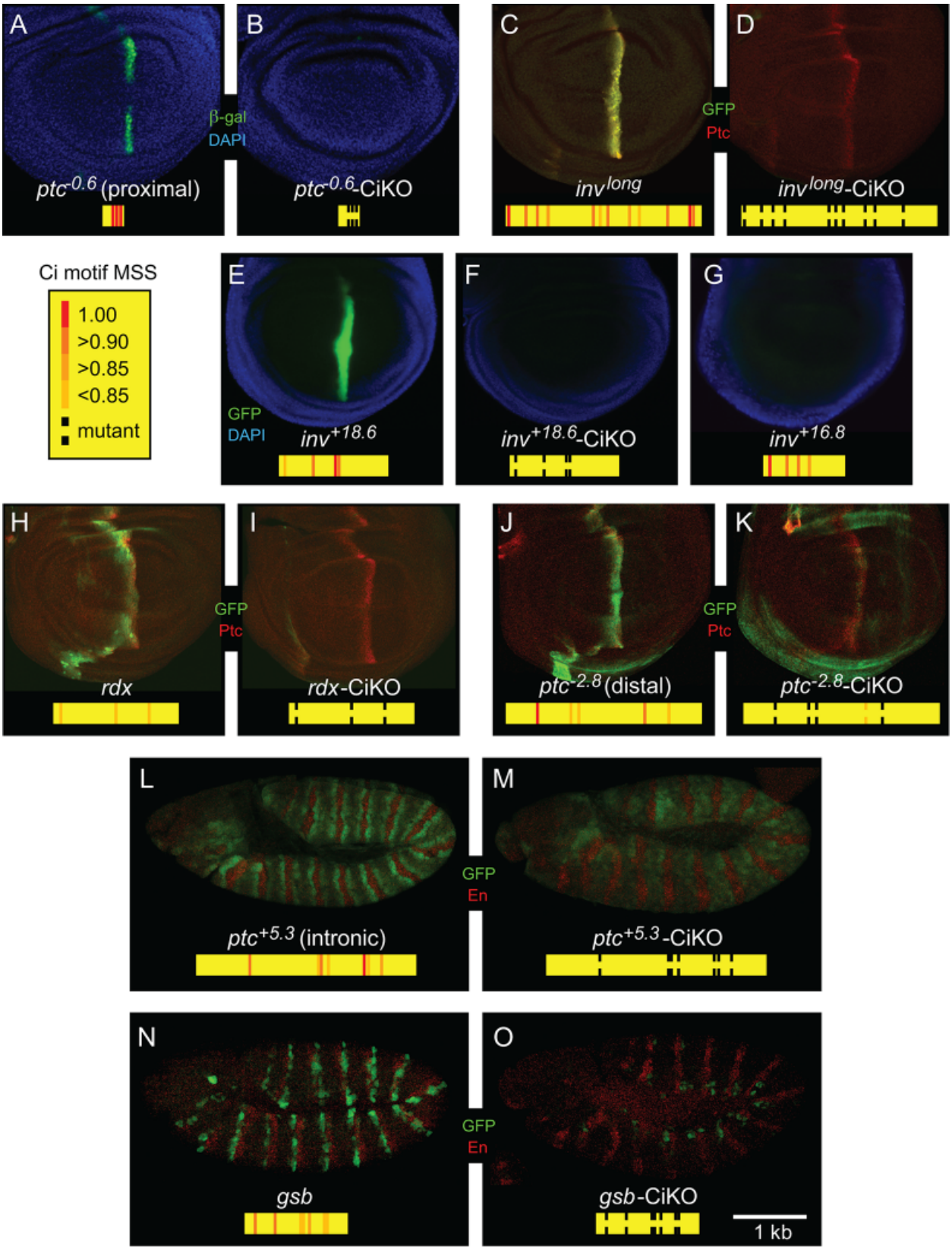

A
B
β-gal
DAPI
ptc-0.6 (proximal)
ptc-0.6-CiKO
C
D
GFP
Ptc
invlong
invlong-CiKO
Ci motif MSS
1.00
>0.90
>0.85
<0.85
mutant
E
F
G
GFP
DAPI
inv+18.6
inv+18.6-CiKO
inv+16.8
H
I
J
K
GFP
Ptc
rdx
rdx-CiKO
ptc-2.8 (distal)
ptc-2.8-CiKO
L
M
GFP
En
ptc+5.3 (intronic)
ptc+5.3-CiKO
N
O
GFP
En
gsb
gsb-CiKO
1 kb

Figure 4.7 Novel enhancers directly respond to Hh signaling in the wing imaginal disc and embryo

(A-K) β-galactosidase or GFP marks the expression of enhancers in the pouch of the wing imaginal disc. A diagram of the fragments tested and location and MSS for all Ci/GLI sites is shown for each candidate (yellow rectangles). Each wild type enhancer responds to Hh signaling along the anterior-posterior compartment boundary of the wing disc, with the exception of *inv*+16.8(G). Active enhancers lose Hh responsiveness in the wing imaginal disc when predicted Ci/GLI binding sites are mutated, as shown in the right of each panel. (L-O) GFP marks the expression of the noted enhancers in the embryo. En expression (red) marks cells producing Hh ligand. When the predicted Ci/GLI binding sites in these enhancers are mutated (M-O), activity in Hh-responsive cells is severely reduced.

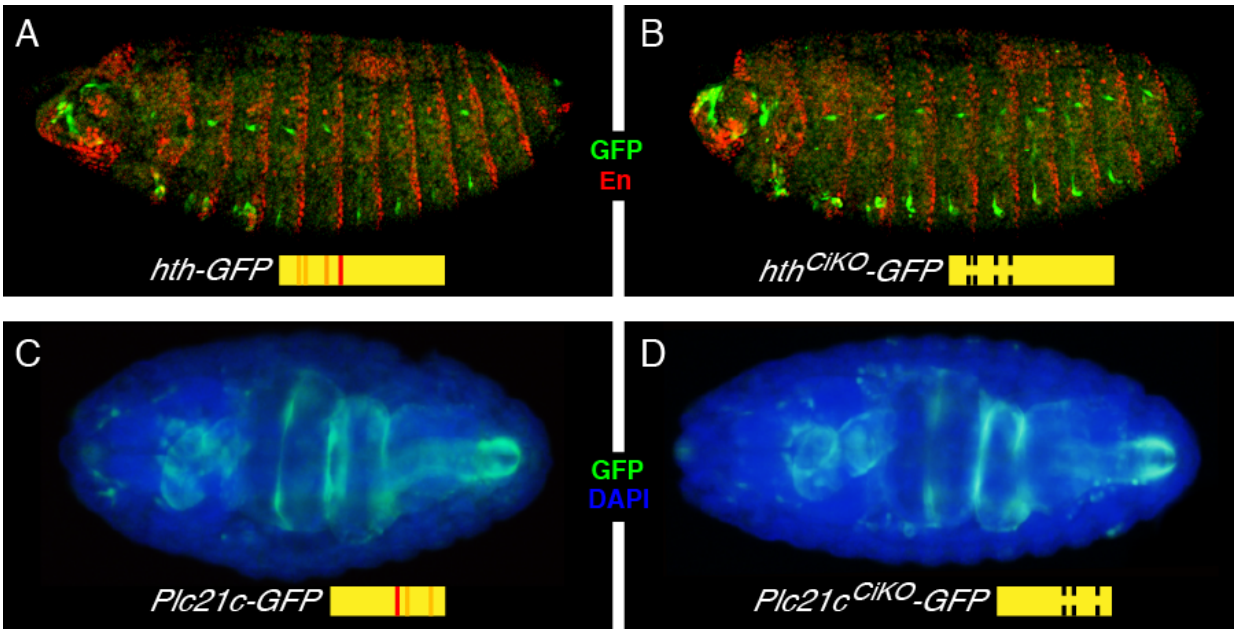


Figure 4.8 Expression of *hth* and *plc21C* regions in the fly are not Ci/GLI-dependent

Both *hth* and *Plc21C* drive GFP expression in the fly embryo. Hth exhibits expression in the brain as well as a punctate segmental pattern parallel but outside of En expression (shown in red) which marks cells that produce and secrete Hh ligand (A,B). *plc21C* expresses throughout the gut (C). Expression for both constructs is not Hh dependent since it persists after mutation of Ci/GLI binding sites (B and D).

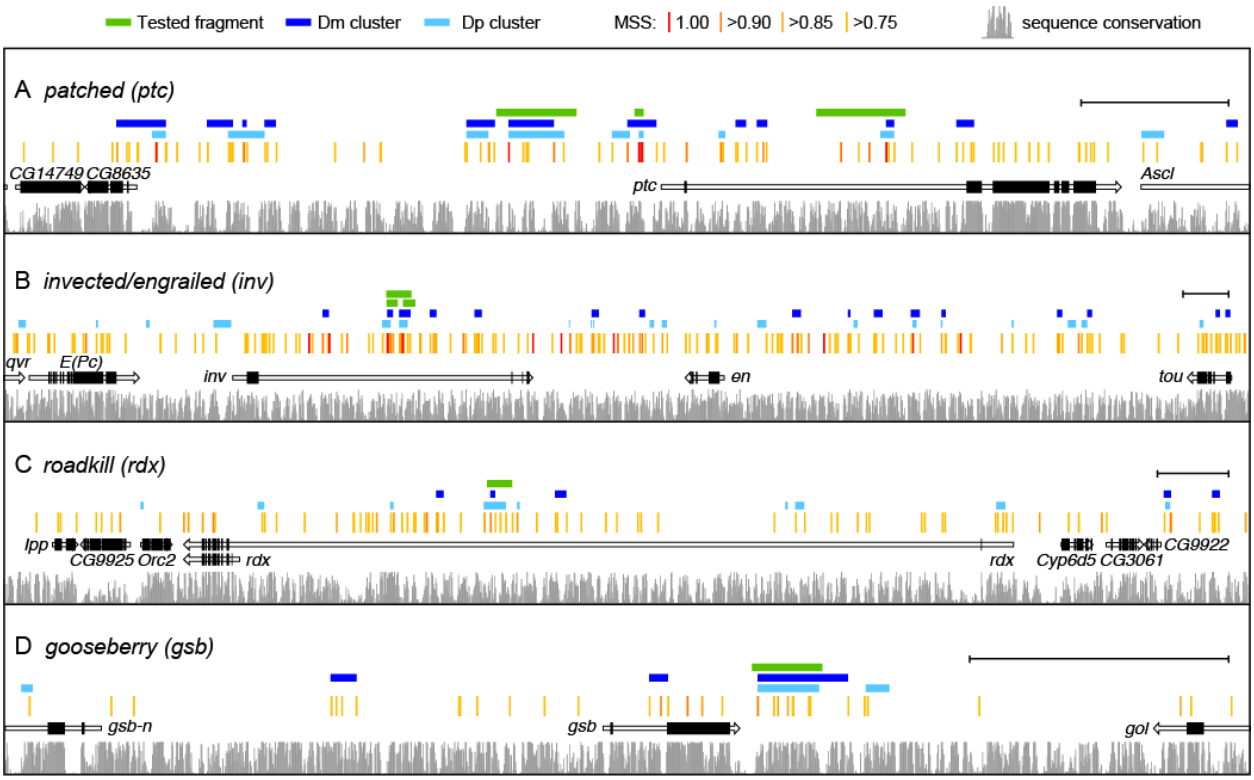
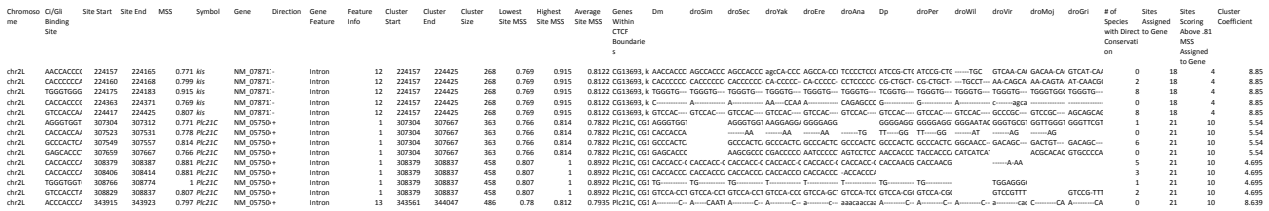


Figure 4.9 Mapping six Hh regulated enhancers in four genetic loci

(A-D) Genomic landscape of the *ptc*, *inv*, *rdx* and *gsb* loci with fragments tested marked by green bars. All predicted Ci/GLI binding sites are highlighted (red/orange tick marks, annotated according to MSS, as noted at top of Figure). The sequence conservation track (gray bars) marks conservation among the 12 sequenced *Drosophila* species, whereas the dark and light blue bars represent clusters of predicted Ci/GLI binding sites in Dm and Dp, respectively. Black brackets at right indicate 5Kb.



Chromosome	Ci/Gli Binding Site	Site Start	Site End	MSS	Symbol	Gene	Direction	Gene Feature	Feature Info	Cluster Start	Cluster End	Cluster Size	Lowest Site MSS	Highest Site MSS	Average Site MSS	Genes Within CTCF Boundaries	Dm	droSim	droSec	droYak	droEre	droAna	Dp	droPer	droWil	droVir	droMoj	droGri	# of Species with Direct to Conservation	Sites Assigned to Gene	Sites Scoring Above .81 MSS Assigned to Gene	Cluster Coefficient
chr2L	AACCACCCC	224157	224165	0.771	*kis*	NM_07871	-	Intron	12	224157	224425	268	0.769	0.915	0.8122	CG13693, k	AACCACCC	AGCCACCC	AGCCACCC	agcCA-CCC	AGCCA-CC	TCCCCTCC	ATCCG-CT	ATCCG-CT	------TGC	GTCAA-CA	GACAA-CA	GTCAT-CA	0	18	4	8.85
chr2L	CACCCCCCA	224160	224168	0.799	*kis*	NM_07871	-	Intron	12	224157	224425	268	0.769	0.915	0.8122	CG13693, k	CACCCCCC	CACCCCCC	CACCCCCC	CA-CCCCC-	CA-CCCCC-	CCTCCCCC	CG-CTGCT-	CG-CTGCT-	---TGCCT---	AA-CAGCA	AA-CAGTA	AT-CAACG	2	18	4	8.85
chr2L	TGGGTGGG	224175	224183	0.915	*kis*	NM_07871	-	Intron	12	224157	224425	268	0.769	0.915	0.8122	CG13693, k	TGGGTG---	TGGGTG---	TGGGTG---	TGGGTG---	TGGGTG---	TGGGTG---	TCGGTG---	TGGGTG---	TGGGTG---	TGGGTG---	TGGGTGG	TGGGTG---	8	18	4	8.85
chr2L	CACCACCCC	224363	224371	0.769	*kis*	NM_07871	-	Intron	12	224157	224425	268	0.769	0.915	0.8122	CG13693, k	C----------	A----------	A----------	AA----CCAA	A----------	CAGAGCCC	G----------	G----------	A----------	c------agcg	----------	----------	0	18	4	8.85
chr2L	GTCCACCAA	224417	224425	0.807	*kis*	NM_07871	-	Intron	12	224157	224425	268	0.769	0.915	0.8122	CG13693, k	GTCCAC----	GTCCAC----	GTCCAC----	GTCCAC----	GTCCAC----	GTCCAC---	GTCCAC---	GTCCAC----	GTCCAC----	GCCCGC---	GTCCGC---	AGCAGCAC	8	18	4	8.85
chr2L	AGGGTGGT	307304	307312	0.771	*Plc21C*	NM_05750	+	Intron	1	307304	307667	363	0.766	0.814	0.7822	Plc21C, CG1	AGGGTGG	AGGGTGG	AAGGAGG	GGGGAGG		GGGGAGG	GGGGAGG	GGGAATAC	GGGTGCG	GGTTGGG	GGGTTCGT	1	21	10	5.54
chr2L	CACACACAA	307523	307531	0.778	*Plc21C*	NM_05750	+	Intron	1	307304	307667	363	0.766	0.814	0.7822	Plc21C, CG1	CACCACCA	-------AA	-------AA	-------AA	--------TG	TT-----GG	TT-----GG	-------AT	-------AG	-------AG	-------AG	0	21	10	5.54
chr2L	GCCCACTCA	307549	307557	0.814	*Plc21C*	NM_05750	+	Intron	1	307304	307667	363	0.766	0.814	0.7822	Plc21C, CG1	GCCCACTC	GCCCACTC	GCCCACTC	GCCCACTC	GCCCACTC	GCCCACTC	GCCCACTC	GCCCACTC	GGCAACC-	GACAGC---	GACTGT---	GACAGC---	6	21	10	5.54
chr2L	GAGCACCC	307659	307667	0.766	*Plc21C*	NM_05750	+	Intron	1	307304	307667	363	0.766	0.814	0.7822	Plc21C, CG1	GAGCACCC	AAGCGCCC	CGACCCCC	AATCCCCC	AGTCCTCC	AACCACCC	TACCACCC	CATCATCA	ACGCACAC	GTGCCCCA		0	21	10	5.54
chr2L	CACCACCCA	308379	308387	0.881	*Plc21C*	NM_05750	+	Intron	1	308379	308837	458	0.807	1	0.8922	Plc21C, CG1	CACCACC-(	CACCACC-(	CACCACC-(	CACCACC-(	CACCACC-(	CACCACC-(	CACCAACG	CACCAACG	-------A-AA				5	21	10	4.695
chr2L	CACCACCCA	308406	308414	0.881	*Plc21C*	NM_05750	+	Intron	1	308379	308837	458	0.807	1	0.8922	Plc21C, CG1	CACCACCC	CACCACCC	CACCACCC	CACCACCC	CACCACCC	-ACCACCC							3	21	10	4.695
chr2L	TGGGTGGT	308766	308774	1	*Plc21C*	NM_05750	+	Intron	1	308379	308837	458	0.807	1	0.8922	Plc21C, CG1	TG----------	TG----------	TG----------	T----------	T----------	T----------	TG----------	TG----------	TGGAGGGI				1	21	10	4.695
chr2L	GTCCACCTA	308829	308837	0.807	*Plc21C*	NM_05750	+	Intron	1	308379	308837	458	0.807	1	0.8922	Plc21C, CG1	GTCCA-CC1	GTCCA-CC1	GTCCA-CC1	GTCCA-CCC	GTCCA-GC	GTCCA-TC(	GTCCA-CG(	GTCCA-CG(	GTCCGTTT	GTCCG-TT1			2	21	10	4.695
chr2L	ACCCACCCA	343915	343923	0.797	*Plc21C*	NM_05750	+	Intron	13	343561	344047	486	0.78	0.812	0.7935	Plc21C, CG1	A----------	C--	A-----CAATI	A----------	C--	A----------	C--	a----------	C--	aaacaccca	A----------	C--	A----------	C--	A----------	C--	a----------	cac	C----------	CA	A----------	CA	0	21	10	8.639

Table 4.1 Predicted clusters of Ci/GLI binding sites in the *Drosophila melanogaster* genome (using the dm3 build of the genome browser).

This is a snapshot of the full table available at http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0145225#sec021 as S4 Table. Columns A-Q are labeled accordingly in row 1. Columns R through AC represent sequence for each *Drosophila* species that corresponds to the multiple sequence alignment (9-mer at the position of the Ci/GLI site in Dm). Number of species that show 100% conservation is shown in column AD. The number of sites assigned to each locus is listed in column AE (boundaries between loci are considered as half of the distance between two neighboring loci). Column AF indicates the number of sites in the locus with MSS≥0.81

Annotated Gene	Genomic coordinates (dm3)	Number of Ci/Gli Sites	Average MSS	Hh Responsive in Chicken Neural Tube	Hh Responsive in Transgenic Fly
ptc$^{-0.6}$	chr2R:4536264-4536572	3	1	+	+
inv$^{+16.8}$	chr2R:7378801-7380000	4	0.941	+	-
Sox100B	chr3R:26894840-26896225	3	0.92	-	-
inv$^{+18.6}$	chr2R:7380576-7381900	4	0.903	+	+
beat-IV	chr3R:19385801-19387033	5	0.899	-	-
CG6475	chr3R:17227902-17229095	4	0.898	-	-
CG34139	chr3R:16067525-16068300	3	0.893	-	-
Plc21C	chr2L:308225-309200	4	0.892	-	-
CG4704	chr3R:18671231-18671930	3	0.891	-	-
Bi	chrX:4316001-4317440	4	0.886	-	-
HGTX	chr3L:14583895-14584670	4	0.886	-	-
Cpr100A	chr3R:26692110-26692580	3	0.886	+	-
Ets21C	chr2L:550010-551035	4	0.885	-	-
CG12541	chrX:6927600-6928375	5	0.884	-	-
Sp1	chrX:9613671-9614922	4	0.881	-	-
Hth	chr3R:6433650-6434996	5	0.879	-	-
Ko	chr3L:21072420-21073658	3	0.879	-	-
ptc$^{+5.3}$	chr2R:4542467-4545417	7	0.875	-	+
ptc$^{-2.8}$	chr2R:4531601-4534319	5	0.847	-	+
Rx	chr2R:16820211-16822050	5	0.845	-	-
Rdx	chr3R:9815295-9817061	3	0.838	+	+
Gsb	chr2R:20952400-20953750	7	0.834	-	+

Table 4.2 Summary of top predicted Hh-responsive enhancers in *Drosophila melanogaster*

The chicken neural tube (CNT) and transgenic fly (TF) assays together identified eight predicted regions as Hh-responsive. Both assays showed positive Hh activity for inv+18.6 and rdx as well as the previously identified *ptc* enhancer region. Five additional regions were positive in only one assay: CNT: Cpr100A and inv+16.8; TF: gsb and two *ptc* additional genomic regions (*ptc*+5.3 and *ptc*-2.8). All enhancer regions were verified by mutagenesis to be Ci/GLI binding site dependent.

Annotation	Genomic coordinates (dm3)
dpp +35	chr2L:2464301-2465300
wg -3.7	chr2L:7303415-7303860
wg +4.2	chr2L:7311428-7312755
Sema-1a	chr2L:8639890-8640550
ptc -6.4	chr2R:4530660-4531850
ptc -4.8	chr2R:4532251-4533000
PKA-R2	chr2R:5906490-5907830
kn +11	chr2R:10674125-10674700
kn -13	chr2R:10699301-10700165
Hs3st-A +4.3	chr2R:14609939-14611000
Hs3st-A +21	chr2R:14627251-14628100
dve -6.8	chr2R:18124601-18125550
dve +21	chr2R:18152960-18154300
dve +25	chr2R:18157301-18158200
dally	chr3L:8818181-8820095
PKA-C3	chr3L:15913901-15915270
boi	chrX:2350251-2351550
sgg	chrX:2538250-2539000

Table 4.3 Clusters containing Ci/GLI sites of low MSS tested in the chicken neural tube assay

DamID locations from Biehs et al. (Build dm3)

Chromosome	Start	End	Cluster Start	Cluster End	Symbol	Gene
chr2L	1455284	1456528	1455873	1456140	snRNA:U3:22A	NR_001600
chr2L	1927769	1928140	1927228	1928142	CG7337	NM_001103581
chr2L	7538466	7543325	7538294	7539165	Rapgap1	NM_001103644
chr2L	9111843	9113084	9111745	9112729	Or30a	NM_078796
chr2L	15937296	15944819	15937185	15937383	beat-Ib	NM_078855
chr2L	17431677	17434112	17433182	17433529	CG5043	NM_135994
chr2R	2655624	2656580	2656051	2656477	Epac	NM_001103732
chr2R	6418206	6419765	6418353	6418759	lola*	NM_080027
chr2R	6462021	6462381	6461408	6462025	psq	NM_165790
chr2R	11004693	11007129	11004180	11005090	chn	NM_206119
chr2R	11336766	11337435	11336714	11337271	CG34365	NM_001103852
chr2R	12283866	12284542	12284061	12284394	CG33960	NM_001038863
chr2R	12412600	12412659	12412235	12412734	Sema-2a	NM_166178
chr2R	12414984	12419228	12416993	12417312	Sema-2a	NM_166178
chr2R	12498703	12500537	12498903	12499084	CG5065	NM_137299
chr2R	12658866	12659557	12659107	12659386	unc-104	NM_166192
chr2R	12677166	12679045	12676636	12677619	CG5522	NM_166193
chr2R	13045282	13055248	13051217	13051370	CG10953	NM_137355
chr2R	13073470	13078925	13073362	13073564	CG10950	NM_137356
chr2R	13166767	13167736	13166630	13167109	mbl	NM_176211
chr2R	13391488	13396940	13394739	13395102	CG4844	NM_001032265
chr2R	15209157	15212514	15211205	15211546	Rgk1	NM_137567
chr2R	15681071	15682013	15680992	15681678	Obp56g	NM_137603
chr2R	15742583	15743824	15743081	15743609	Obp56i	NM_166376
chr2R	18226254	18226613	18226458	18226970	CG11206	NM_001103944
chr2R	18259854	18261720	18260459	18261287	Rtf1	NM_137821
chr2R	18431154	18433651	18432448	18432907	px*	NM_001103947
chr2R	20272172	20272571	20272357	20273224	mAcR-60C	NM_079120
chr3L	8052940	8054760	8052894	8053443	bip1	NM_139912
chr3L	11862617	11864769	11864451	11864952	CG5906	NM_140261
chr3L	12458113	12459080	12457913	12458121	eyg	NM_079318
chr3L	13940701	13941699	13941029	13941736	CG32137	NM_168564
chr3L	17339408	17344264	17341139	17341588	Mip	NM_140714
chr3L	20692805	20697982	20692956	20693766	kni*	NM_079463
chr3R	2329935	2332092	2329972	2330257	CG15186	NM_169145
chr3R	3976809	3982873	3978441	3978655	grn	NM_001144554
chr3R	4101609	4106766	4102638	4102804	ato*	NM_169213
chr3R	4401007	4404074	4403133	4403663	Cenp-C	NM_169228
chr3R	5682908	5683267	5682873	5683534	Teh1	NM_141702
chr3R	9669658	9672691	9670696	9671137	CG31495	NM_169559
chr3R	17217597	17220086	17219899	17220257	CG6475	NM_001104389
chr3R	17227813	17229687	17227902	17228674	CG6475	NM_001104389
chr3R	20530764	20536210	20530430	20531229	slo	NM_001170241
chr3R	20549088	20550942	20549943	20550067	tok	NM_170168
chr3R	21180551	21181210	21180942	21181629	Fur1	NM_170654
chrX	3720410	3720469	3719976	3720491	ec	NM_130722
chrX	4249010	4249973	4249936	4250271	norpA	NM_001169190
chrX	4256130	4258561	4256588	4257103	norpA	NM_080330
chrX	4537017	4537076	4537020	4537166	CG3062	NM_131949
chrX	5540520	5540876	5539974	5540628	CG42492	NM_001169205
chrX	6202614	6205080	6202296	6202630	CG3823	NM_132085
chrX	6297612	6298259	6296956	6297618	CG33669	NM_001031872
chrX	7309802	7314072	7309556	7310205	CG34337	NM_001103431
chrX	7400950	7404588	7400758	7401576	CG32719	NM_167128
chrX	15536311	15537530	15537132	15537533	CG8117	NM_132821

Table 4.4 Overlap between clusters predicted in this study and DamID protected sites

Asterisks indicate four sites that map to one of the 52 probable Ci target genes identified by Biehs et al.

Forward	Reverse	Matrix Similarity Score (MSS)
GACCACCCA	TGGGTGGTC	1
GACCCCCCA	TGGGGGGTC	0.918
GACCTCCCA	TGGGAGGTC	0.916
GGCCACCCA	TGGGTGGCC	0.915
GCCCACCCA	TGGGTGGGC	0.915
GACCGCCCA	TGGGCGGTC	0.914
GTCCACCCA	TGGGTGGAC	0.91
GACCACACA	TGTGTGGTC	0.907
GACCACGCA	TGCGTGGTC	0.899
GACCACTCA	TGAGTGGTC	0.899
GACCACCTA	TAGGTGGTC	0.897
GACCACCAA	TTGGTGGTC	0.897
GACCACCGA	TCGGTGGTC	0.895
GACCACCCC	GGGGTGGTC	0.889
GACCACCCT	AGGGTGGTC	0.889
GACCACCCG	CGGGTGGTC	0.888
AACCACCCA	TGGGTGGTT	0.882
TACCACCCA	TGGGTGGTA	0.882
CACCACCCA	TGGGTGGTG	0.881
GAACACCCA	TGGGTGTTC	0.879
GAGCACCCA	TGGGTGCTC	0.878
GATCACCCA	TGGGTGATC	0.878
GCCCCCCCA	TGGGGGGGC	0.833

Table 4.5 Predicted 9-mer GLI binding sites with a minimum level (≥0.75) Ci matrix similarity score

Full Table can be found at http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0145225#sec021 as Table S1.

Annotation	Forward primer	Reverse primer	Genomic coordinates (dm3)
ptc -0.6	CACCGGCGCGCCATGCATGCGCAGCCTGCCAC	GTACCGCGGTTTCTATTGTTATTCGCATG	chr2R:4536264-4536572
inv +16.8	CACCTGATATCTTAGGTTAGTAGTAT	AATCTAATTTTGCCCTGATATT	chr2R:7378801-7380000
Sox100B	CACCTAAGCTCGGGATATTTTGCC	AGCTTAGAGGTCCTGCATAG	chr3R:26894840-26896225
inv +18.6	CACCTATGTTATAAAATTTGTAATAT	TTTGGTTATACTGTCTAACAAA	chr2R:7380576-7381900
beat-IV	CACCGTTTTTTTGCATTTCACC	AAACTACACGGCTGCCCTG	chr3R:19385801-19387033
CG6475	CACCGACCACACAACAGACGC	TACTTGAGCACCCGATTGG	chr3R:17227902-17229095
CG34139	CACCCTTTCGTTTTATGTTAACG	TTGTTTTTTTTCTTTTCGCTGTGCG	chr3R:16067525-16068300
Plc21C	CACCTCGTTATGATGTGCCTTAAAAG	AAAATATTAACGCGAAATAGG	chr2L:308225-309200
CG4704	CACCGTCATATTAGGCTATTTC	CATTTTATTAGCCGAATGC	chr3R:18671231-18671930
bi	CACCGAGAGGGAGCGAGTGAGTAAG	TGAGGCAATCGATAAAATTAGC	chrX:4316001-4317440
HGTX	CACCTGCAGCCGCTTAATAATTCC	AGTGCCGTGCTTAACCCG	chr3L:14583895-14584670
Cpr100A	CACCATAATGCCAAAAGTTCTCTG	TGTCTTTTTGATTTTTCCAGTG	chr3R:26692110-26692580
Ets21C	CACCGTTTGTACCCTGTAAAGGG	ACTTAAACGGAGCCACATTTTTCTC	chr2L:550010-551035
CG12541	CACCCAGCAAGAAGCATACCAAAG	ACTATTAGCTACATTTTCTTCC	chrX:6927600-6928375
Sp1	CACCTTCGCCGTGTGTATGTATTAGC	TATCAACGGAAATTCATTAC	chrX:9613671-9614922
hth	CACCTAAAGCCGAAAGCCTAAAATAG	TTTGCTTGTATTTTCCGAAC	chr3R:6433650-6434996
ko	CACCAGGAGACAGGTGGTATGGTC	ATTCACAGTGTAATTTTACAGC	chr3L:21072420-21073658
ptc +5.3+A3	CACCGGCGCGCCGAAGTGCTTAACAAGTTAAC	GTACCGCGGCACGACAACCAATGAGATCG	chr2R:4542467-4545417
ptc -2.8	CACCGGCGCGCCTACGTACTCTTATTACTCCACTC	GTACCGCGGGCTATTGCATTTGTCATTGGC	chr2R:4531601-4534319
Rx	CACCACTCCTTCCCCGACTTAC	GTTTAAACCGAAAAACGTTTAATTTAATCTGG	chr2R:16820211-16822050
rdx	CACCTTTAGCCAGGTGTGGATTGTG	CCAGCGAAAGCAAACAGAGTAC	chr3R:9815295-9817061
gsb	CACCGCGAGTCAAACTCATTCCGTG	AAGTGTACGGTGAATTC	chr2R:20952400-20953750

Table 4.6 PCR primers used to amplify genomic DNA from the *D. melanogaster* genome (build dm3)

Chapter 5

Genomic Engineering of the Hedgehog Signaling Pathway

5.1 Abstract

Thus far I have examined how the Hedgehog signaling pathway regulates gene expression using mainly GFP reporter assays. These experiments provide an in vivo transcriptional readout of Hh target gene enhancers in several developmental contexts, but two caveats still apply. First, I have not demonstrated tissue-specific Ci occupancy in vivo at any of the *Drosophila* enhancers identified in chapters 2-4. Instead, I have used EMSA analysis to demonstrate that Ci recognizes a particular binding site in vitro, but this neglects other co-activators, repressors, or other TFs that might affect Ci binding to these sites in vivo. Second, I have not demonstrated the functional contribution of these enhancers to the development of the animal. The GFP expression suggests functional significance, but I have not directly demonstrated it. In order to address these questions, I have employed the powerful CRISPR/Cas9 approach to quickly and efficiently target genomic sequences in vivo, allowing me to more effectively examine enhancer function.

5.2 Introduction

Direct manipulation of the genome is the new gold standard for assessing the function of genes and *cis*-regulatory elements. Historically, researchers have manipulated the genome in many ways—for instance, by using chemical mutagens to

disrupt functionally relevant genes. This is how *hh*, *ptc*, *odd-skipped* and many other critical patterning genes were identified by Christiane Nusslein Volhard and Eric Weishchaus in a series of Nobel Prize-winning forward genetic screens (Nüsslein-Volhard & Wieschaus 1980; Adams & Sekelsky 2002). In flies, “jumping genes” called transposons have long been used as an alternative method for creating mutations (Spradling et al. 1999). Though these genomic manipulation strategies are effective for conducting large-scale screens, they can only produce mutations at random and thus are not useful for making precise, targeted alterations to the genome. A more recent technique, homologous recombination, makes targeted alterations to the genome possible, but it is inefficient and technically challenging (Gong & Golic 2003; Huang et al. 2009). Other genome editing technologies employ sequence-specific DNA-binding proteins, such as zinc finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs). These approaches have the drawbacks that they require the generation of a customized protein for each targeting event, that the process is cumbersome, and that they generate undesired off-target effects in complex genomes (Wood et al. 2011; Gaj et al. 2013).

More recently, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technique has exploded in popularity in the scientific community (Doudna & Charpentier 2014; Sternberg & Doudna 2015; Mohr et al. 2016). CRISPR was originally identified in 1987 as a means by which *E. coli* defend against invading viral DNA. The system was later adapted by Jennifer Doudna and Emmanuelle Charpentier to precisely and efficiently edit the genome of other organisms (Ishino et al. 1987; Jinek et al. 2012). CRISPR genome editing, which I employ in the work described in this chapter, relies on

the Cas9 enzyme, an endonuclease that cleaves double stranded DNA. In order for genomic manipulations to be inherited, they must occur in the germline. In *Drosophila*, this is achieved by driving Cas9 expression in germ cells using promoter of the germline-specific gene *vasa* (Sebo et al. 2013). When injected into embryos, guide RNAs (gRNAs) will recruit the Cas9 nuclease to a region of the genome that is complementary to the gRNA sequence, producing a double-stranded break precisely at that location. Once Cas9 cuts the DNA, homologous recombination will occur between the freshly cut genomic DNA and another plasmid that was co-injected with the gRNAs carrying a replacement cassette. This replacement cassette contains two 1kb homology arms to facilitate recombination and create a freshly engineered locus. This technique has become extremely popular for its precision and efficiency, and seems to be very specific, as recent reports have shown that there are minimal off target effects using this technique (Bassett et al., 2013; Kim et al., 2015; Kleinstiver et al., 2016). I suspect CRISPR/Cas9 engineering will be as commonly used as Polymerase Chain Reaction (PCR) in the lab within just a few years.

I have used CRISPR genome editing to alter two genes critical to Hh signaling: *ci* and *patched*. First, the *ci* locus was targeted because the field lacks an antibody sufficient to complete a successful ChIP-seq experiment. This major limitation has led researchers to employ less efficient workarounds, like the DamID technique to detect where Ci binds to genomic sequences (Vogel et al. 2007; Biehs et al. 2010). The results of that screen, however, predicted only 52 targets and these were not functionally validated by Ci binding site mutagenesis (see chapters 2 and 4 for more discussion). Therefore, I introduced a 3xFLAG epitope, which is recognized with high affinity and

specificity by a commercially available antibody, at the N-terminus of the *ci* coding sequence. To accomplish this, I examined the transcriptional landscape of *ci* to identify essential regulatory elements controlling c*i* expression. Published work by other groups have shown that the regulatory elements controlling *ci* in both the embryo and wing imaginal discs are located upstream of the gene (Schwartz et al. 1995)**;** highlighted in yellow in Figure 5.1), so when I was forced to remove roughly 3kb of the large first intron because of highly repetitive sequence that was not specific enough to target with gRNA, I was not concerned with losing regulatory information (Figure 5.1).

The second genomic manipulation I made targeted the *patched* locus so that I could more directly assess the function of the promoter and promoter-proximal regulatory elements discussed in Chapters 2 and 3. For this, I took advantage of site-specific recombination to accurately and efficiently recombine sequences into the locus using already established technologies. Both of these projects are ongoing, but the CRISPR editing steps have been successful. Validation experiments and future directions are discussed below.

5.3 Results

5.3.1 Generation of a FLAG-tagged Ci allele

To insert a 3xFLAG epitope at the N-terminus of *ci*, I injected a replacement cassette and two gRNAs into stage 2-3 *Drosophila* embryos expressing Cas9 in germ cells (Figure 5.2A). These constructs were all transfected as circular DNA plasmids, with the gRNAs under the control of a U6 promoter (Gratz et al. 2014). Out of 600 injected embryos, 97 survived. I outcrossed the survivors to wild type flies; of these crosses, 46 were fertile. Using a fluorescence dissecting stereomicroscope, I detected

fluorescent red eyes in at least one progeny in 6 crosses, indicating successful transformation. I then crossed these transformants into a line of flies that expresses Cre recombinase under the control of a constitutively active promoter to excise the dsRed cassette, leaving a single loxP scar in the intron of *ci* (Figure 5.2B-C). While I was then able to screen for the loss of red eyes, I still needed to genotype these flies to confirm proper recombination had taken place, so I extracted genomic DNA from a single wing of these flies for genotyping via PCR. I designed a forward primer (P1) upstream of homology region 1 and a reverse primer at the 3' end of the intron (Figure 5.2C,D), which produced a ~1.6kb band, as expected. I then designed internal primers surrounding the 3xFLAG tag (P3 and P4), to easily screen wild type and FLAG tagged alleles (Figure 5.2C,E). All genotyping reactions yielded correct amplicon sizes, which were then sequenced to confirm successful targeting.

The first clue that *ci* was still functional after this engineering came when I saw that flies homozygous for the ci^{FLAG} allele were viable and phenotypically normal, as *ci* mutants have defects in embryonic, larval and pupal development (Méthot & Basler 1999). I next sought to confirm that Ci protein was expressed in the correct cells during development. I accomplished this by comparing endogenous Ci antibody staining with anti-FLAG staining in the same samples. At time points in embryonic development when Ci is expressed, FLAG expression co-localized with that of Ci (Figure 5.3 A-C). Furthermore, I observed co-expression in larval tissues including the wing, leg and haltere imaginal discs (Figure 5.3D). Together, these results indicate that this Ci^{FLAG} allele successfully recapitulates endogenous Ci expression and will be useful for further downstream studies, as explained in the discussion section below.

5.3.2 An efficient strategy to target *patched*

In order to generate an easily targetable allele of *ptc*, I replaced the first exon of *ptc* and approximately 1kb of flanking sequence in either direction with an attP recombination site using CRISPR/Cas9 replacement. For *ptc*, I chose to incorporate the attP site-specific recombination system because of its ability to swap sequences with extremely high efficiency (Thorpe & Smith 1998; Thorpe et al. 2000). By replacing the first exon of *ptc* with an attP recombination site, I created a targetable heterozygous *ptc* knockout (Figure 5.3A, $ptc^{attP\text{-}dsRed/+}$). These flies still have one copy of wild type *ptc,* allowing them to survive, but can also be targeted by site specific recombination (Bischof et al. 2007). Using this strategy, recombination will be induced in the germ cells allowing us to quickly and efficiently make changes to the *ptc* locus (Figure 5.3B,C). Both experiments planned in Figure 5 have significant implications for better understanding the regulation of *ptc*. First, we will knock RFP into the first exon of *ptc* to generate a fluorescently marked line of flies, allowing us to easily assess *ptc* expression (Figure 5.4B). Second, we plan to re-introduce the wild type first exon of *patched*, but knock out the 3x GLI binding sites in the promoter-proximal *patched* enhancer to assess the functional contribution of these sites to *ptc* expression (Figure 5.4C). In both cases, the plasmid backbones will be removed by crossing flies into a Cre-expressing background (Figure 5.4B,C, bottom).

I have successfully targeted *ptc* using CRISPR/Cas9 to insert the attP recombination site and have crossed this allele into the PhiC31 background for site-specific recombination (Figure 5.4A). I confirmed this with the expression of dsRed in these flies as well as extensive PCR genotyping (Figure 5.5). The expression of dsRed confirmed that my construct was integrated into the genome. Primers P1, P2, P7 and

P8 were designed to anneal outside of the homology regions in order to confirm integration at the *ptc* locus. Internal primers (P3-P6) allowed smaller amplicons to facilitate efficient PCR. Genomic DNA was extracted and PCR completed using the primers in Figure 5.5B. While the expected products were generated, there were also smaller, additional bands in most of the reactions, most likely suggesting that PCR conditions were not optimized for each primer set, allowing for a more permissive environment in which the primers could recognize other regions of the genome (Figure 5.5C).

5.4 Discussion

In this chapter, I have begun to address three major caveats to the first four chapters of my dissertation. The first is the lack of a high-quality Ci antibody that would allow us to complete critical in vivo analysis of Ci occupancy in the genome. Ci-FLAG addresses this major issue. The second and third critiques deal with the inherent limitation of GFP reporter constructs: examining each enhancer in a single, isolated reporter construct excludes the effects of other *cis*-regulatory elements on gene expression and by using GFP as a proxy for *ptc,* I am unable to see precisely how patterning contributes to developmental processes. I would like to know how these enhancers work together in their native genomic context to regulate *patched* expression. These last two problems will be addressed by ptc^{attP}, which will allow highly efficient recombination into the *ptc* locus using the PhiC31 integration system. By examining sequences in their genomic contexts, I can address shortcomings of prior transgenic analyses.

The first experiment planned using these freshly generated alleles is to use Ci^{FLAG} to better understand the tissue-specific binding properties of Ci. By isolating tissues from different stages of development and then subjecting them to ChIP-seq experiments using the anti-FLAG antibody (Vokes et al. 2007), we can address this pressing issue. I suspect that many regions in the *ptc* locus will be constitutively bound by Ci, but that some of these sites will also change depending on the tissue examined, similar to the GLI2 binding in different mouse tissues and contexts (Figure 2.17) and as predicted by the low resolution DamID results (Biehs et al. 2010). We can also make specific predictions based on the reporter assays conducted in chapters 2-4. For example, enhancer DB responds to Hh signaling in the wing imaginal discs, but not the embryo. So, we would expect Ci to occupy enhancer DB in the larval context, but not in the embryonic context. Then, by doing bioinformatics analysis on the peaks generated by ChIP-seq, we can begin to characterize the affinity of Ci binding sites in each Ci-occupied region to learn more about how Ci recognizes DNA. This analysis will also allow us to scan predicted regions for additional transcription factor inputs regulating Hh signaling in these contexts by searching for non-Ci binding sites using the tools already discussed in Chapter 3 (Figure 3.1).

Even more broadly, by doing a global analysis of Ci occupancy in the genome, we can generate a more complete list of Hh target genes to better understand the mechanisms controlling patterning, morphogenesis and adult cell maintenance, depending on the context. We could then couple those ChIP-seq data to RNA-seq data to create lists of up- and down-regulated genes under conditions of increased or

decreased Hh signaling, using available genetic tools such as a temperature-sensitive allele of Hh (Ingham 1993**;** Figure 2.7).

In addition to global analysis of Ci binding preferences, we will also perform more targeted experiments to identify sequence-specific binding preferences of Ci within a single enhancer. For example, does the in vitro affinity of Ci for a particular motif correlate with in vivo affinity? Previous work has demonstrated that the highest-affinity Ci binding site in vitro is GACCACCCA in luciferase assays and EMSA competition analyses (Hallikas et al. 2006; Parker et al. 2011). These experiments, however, have many caveats. First, the EMSA analysis is performed with the DNA binding domain of Ci alone; any other regions of the TF that may be important for binding are not included in the analysis. Second, no cofactors of Ci, which might impact its binding preferences, are present in this analysis. Third, both EMSAs and luciferase assays use a small DNA probe consisting mainly of the predicted Ci binding site, which may obscure the importance of flanking DNA sequences for Ci binding. All of these issues can be addressed by using already generated synthetic enhancers constructs (Ramos & Barolo 2013; Figure 5.6). These enhancer-reporter constructs all contain a 3x multimer of a particular high- or low-affinity Ci binding site (Figure 5.6, left) and are integrated as single copy insertions into the same genomic position, so they experience the same position effects once they are integrated into the genome. To examine how well Ci binds these different GLI binding sites, we will cross our Ci-FLAG allele with these constructs and complete ChIP-qPCR using identical primers to yield an extremely well controlled analysis of Ci-binding in an in vivo context. Interestingly, the data presented in Figure 5.6 suggest that, at least in the embryo, predicted lower affinity Ci binding sites

generate a stronger response to Hh signaling (Figure 5.6, right). With Ci^{FLAG}, we can now address how well Ci binds these sites in vivo with all of its appropriate cofactors. I suspect that we will see much stronger preference for Ci binding to lower affinity sites in the embryo, as compared to the larval tissues, because of the altered ratio of Ci activator:repressor, as discussed in Chapter 2.

The second and third major issues to address relate to the design of our enhancer reporters. Because we are removing the enhancers from their native context in the *ptc* locus, they are no longer able to interact with other regulatory elements active in that tissue. While we do not know if this is happening, the evidence suggests there are multiple enhancers active in a single tissue at any particular time and raises this possibility (Figure 2.5, for example). In addition to this, we are also not addressing the functional contribution of these enhancers to *ptc* expression directly. For instance, if we removed any one of these enhancers in the context of all *ptc* regulatory elements, would *ptc* expression be changed even though there is potentially a large amount of redundancy built in to the locus? Furthermore, would there be a change in the morphology of that tissue because of disrupted signaling? With the creation of the ptc^{attP} allele, we can address both of these shortcomings (Figure 5.4C, for example).

To address the issue of enhancer redundancy, I plan to knock in an RFP allele into the *ptc* locus. If we then use CRISPR/Cas9 to begin deleting enhancers in the ptc^{RFP} allele (proposed in Figure 5.4B), we should be able to observe even subtle changes in expression. This CRISPR/Cas9 strategy should be highly efficient since we just need to create simple deletions, which will be achieved by using two gRNAs and Cas9 to create a Non-Homologous End Joining (NHEJ) deletion (Gratz et al. 2014). If

there are major effects on RFP expression, we will delete those enhancers in a wild type *ptc* locus using the same CRISPR/Cas9 NHEJ strategy, but this time will be able to see any change in phenotype associated with the mutation. I predict that these experiments will demonstrate individual enhancers, while sufficient to respond to Hh signaling, are not all required for the Hh response individually. Rather, since there are so many enhancers in the locus that are sufficient to respond to Hh signaling, we will need to remove multiple enhancers at once to see an effect. This is similar to the idea of redundant and shadow enhancers (Frankel et al. 2010; Perry et al. 2010; Barolo 2011a; Wunderlich et al. 2015; Staller et al. 2015).

We can also use this ptc^{attP} allele to examine the effect of removing the 3x Ci binding site cluster that responds to Hh signaling in the wing, but not in the embryo (Figure 5.4C, 2.1D-E) simply by creating a destination vector (pHD_ptc3xGBSko_destination, Figure 5.4C) using basic molecular biology techniques. Because we know that those 3 Ci binding sites are not sufficient to respond in the embryo, it suggests that removing those sites would not affect embryogenesis, since those sites seem to be important for larval development, a time point after embryogenesis. So, we might expect to see flies with morphologically deficient wings, but otherwise normal appendages. It would be interesting, however, if these flies are embryonic lethal, suggesting that there are signals being integrated through those binding sites, but they need input from other *cis*-regulatory sequences in the locus that are active in the embryo, such as enhancers 1EH, VT, or LK. There has been previous work published demonstrating that a *LacZ* reporter transgene carrying -12.5kb sequence from the *ptc* TSS recapitulates the embryonic Hh response (Forbes et al.

1993), but a version of this without those 3x Ci binding sites was never made. The proposed experiments here would be crucial to understanding how *ptc* is regulated using multiple enhancers in the context of the entire locus, which will add to our current knowledge of *ptc* regulation and how this regulation affects Hh signaling in different developmental contexts.

The experiments described in this section are just a fraction of the work that could be done with these alleles. For example, instead of destroying the 3x Ci binding sites as depicted in Figure 5.4C, we could just as easily manipulate the binding sites within the *ptc* promoter, to examine how this Polycomb Response Element (PRE, described in Chapter 2) affects embryonic development. For this experiment, depending on how we manipulate the promoter sequence, I would expect disrupted development during embryogenesis, since we see PRE effects by then already in our GFP reporter assays (Figures 2.11-2.14). The generation of these alleles opens the door to an infinite number of experiments to functionally assess the role of *cis*-regulatory elements in the regulation of *ptc* and the corresponding effects of Hh signaling on the development of the organism.

5.5 Materials and Methods

DNA sequence alignments

Sequences, DNAse hypersensitivity tracks, and multi-species alignments were obtained from the UCSC Genome Browser (genome.ucsc.edu).

DNA cloning

Homology directed repair cassettes were created by PCR amplification of each ~1kb homology region from Vasa-Cas9 genomic DNA and direct cloning into the pHD_dsRed_attP (Gratz et al., 2014) cassette via Aar1 (5' homology arm) or Sap1 (3' homology arm) restriction sites. Guide RNA were designed using Fly CRISPR Target finder (http://tools.flycrispr.molbio.wisc.edu/targetFinder/) using their highest stringency filters allowing 1 or less off target effects for each gRNA. These gRNAs were ordered as complimentary oligos and annealed by standard protocols for cloning into pU6-Bbs1-chiRNA vector (Gratz et al., 2013). Primers can be found in Table 5.1.

Genotyping

Genomic DNA was isolated from either whole flies or single wings to keep flies alive for additional crosses (Carvalho et al. 2009). This DNA was subjected to PCR as previously described in Chapter 2, and examined on a 1-2% agarose gel. All amplicons were sequenced at the University of Michigan DNA Sequencing Core Facility. Primers can be found in Figure 5.7.

***Drosophila* transgenesis and stocks**

Injection of gRNA and Homology Directed Repair cassette was performed as previously described (Gratz et al. 2014) into the Bloomington 55821 strain (y[1] M{vas-Cas9.RFP-}ZH2A w[1118]/FM7a, P{w[+mC]=Tb[1]}FM7-A). The dsRed cassette was removed by crossing into the Cre-expression Bloomington Stock 766 (y[1] w[67c23] P{y[+mDint2]=Crey}1b;noc[Sco]/CyO). In the *ptc*-attP lines that were put in the background of PhiC31, this was done using Bloomington stock 24749 (y[1] M{vas-

int.Dm}ZH-2A w[*]; M{3xP3-RFP.attP}ZH-86Fb). All synthetic enhancer reporters were made as previously described (Ramos & Barolo 2013).

Immunohistochemistry and confocal microscopy

Drosophila embryos and third-instar imaginal discs were fixed and stained using standard methods (D. S. Parker et al. 2011; Ramos & Barolo 2013; White et al. 2012). Primary antibodies used included rat anti-Ci (Developmental Studies Hybridoma Bank 2A1, 1:10), mouse anti-FLAG (Sigma F1804, 1:50), rabbit anti-EGFP (Invitrogen 1:100), and mouse anti-Engrailed (Developmental Studies Hybridoma Bank 4D9, 1:50). Secondary Antibodies included goat anti-rat 568 (Invitrogen A11077, 1:2000), goat anti-rabbit 488 (Invitrogen A11008) and goat anti-mouse 488 (Invitrogen A11001, 1:2000), Antibodies obtained from the Developmental Studies Hybridoma Bank were developed under the auspices of the National Institute of Child Health and Human Development and maintained by the Department of Biological Sciences, The University of Iowa (Iowa City, IA). Confocal images were captured on an Olympus FluoView 500 Laser Scanning Confocal Microscope mounted on an Olympus IX-71 inverted microscope, and on a Nikon A1 confocal microscope. Samples to be directly compared were fixed, prepared, and imaged under identical confocal microscopy conditions and settings.

5.6 Acknowledgements

I would like to thank Scott Barolo and Ben Allen for extremely important and helpful discussions about how to design both the 3x FLAG tag for *ci* and the attP recombination

designed at the *patched* locus. In addition, I am grateful for Andrea Ramos who contributed the synthetic enhancer data discussed in Figure 5.6.

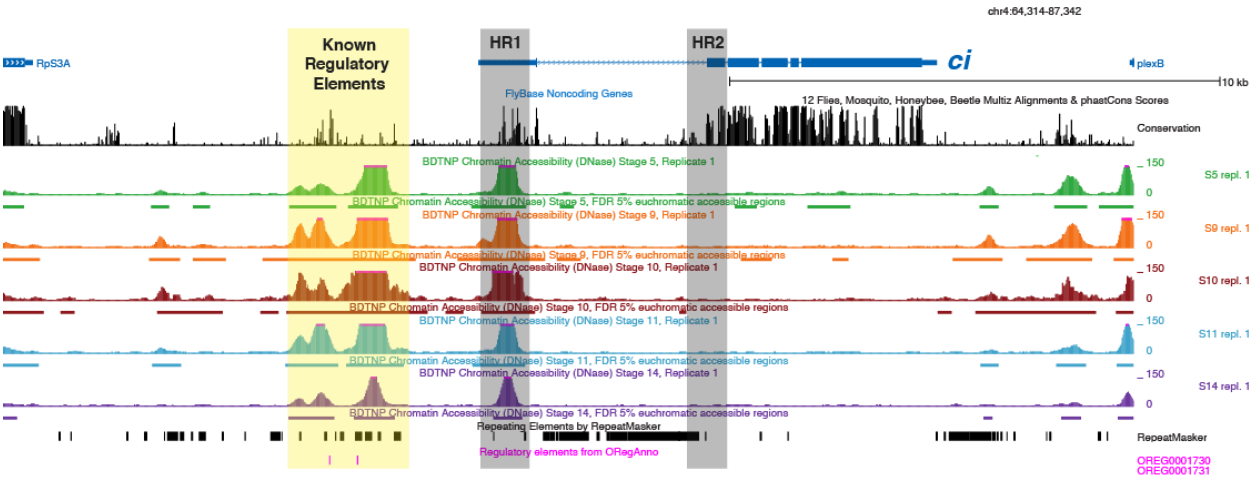


Figure 5.1 Genomic landscape of *ci* annotated with data from the UCSC genome browser

Known regulatory elements are highlighted in yellow upstream of the *ci* coding region (see bottom for UCSC identification of published enhancers in pink). Gray shading indicates homology regions used in CRISPR experiments. Conservation among *Drosophilids* is shown as black bars underneath *ci* gene map. Green, orange, red, blue, and purple tracks represent DNAse hypersensitivity, horizontal bars across peaks indicate significant peaks. Black horizontal bars indicating repetitive DNA sequences are shown (note highly repetitive, weakly conserved DNA within first *ci* intron).

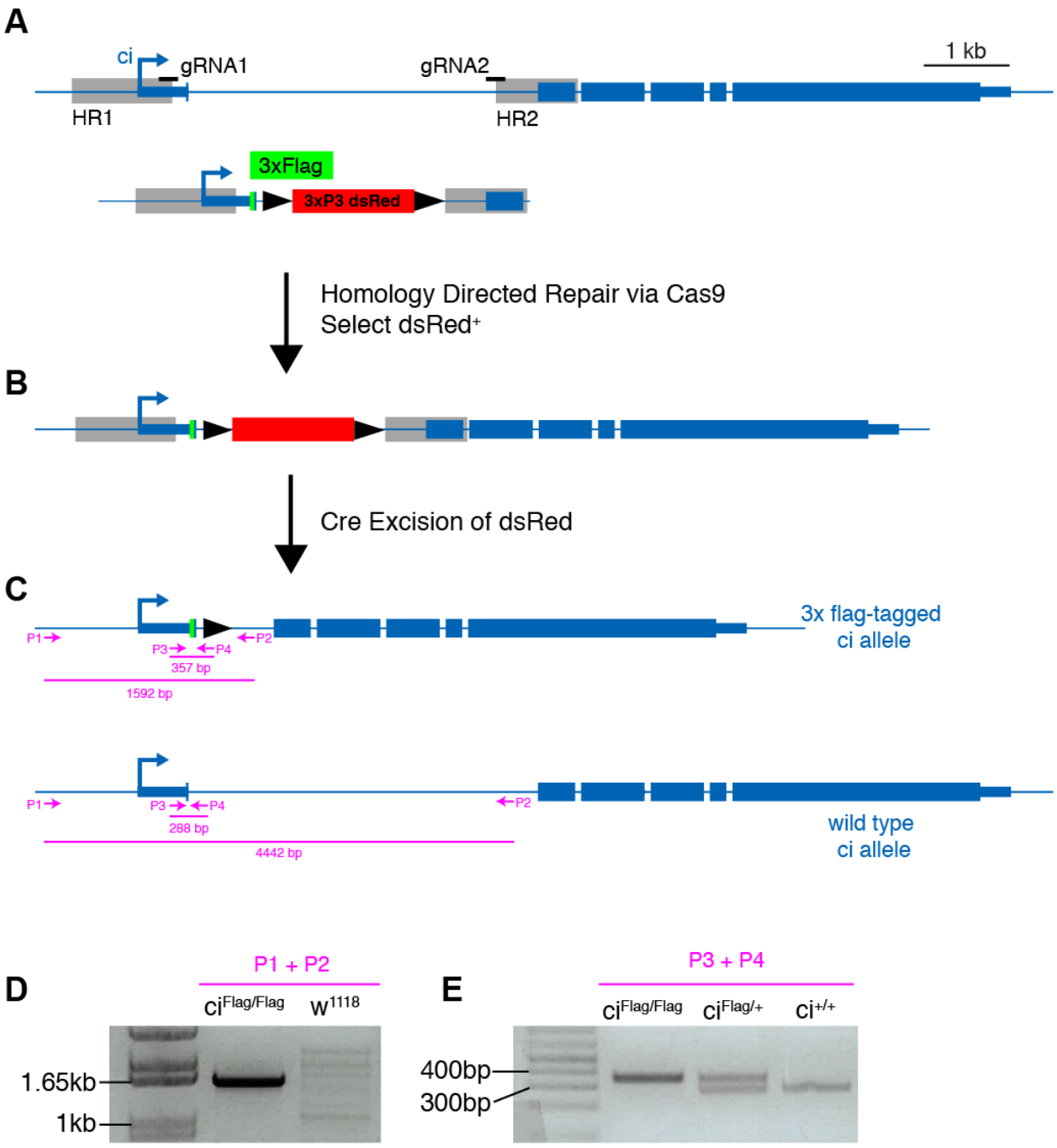


Figure 5.2 CRISPR/Cas9 strategy to tag *ci* with FLAG

(A) Gene map of *ci*. Homology regions 1 and 2 are marked in gray. gRNAs are also shown. The homology cassette used to repair the locus after Cas9 cuts contains a 3xFLAG inserted at the 5'end of the *ci* coding region (ATG added to 3xFLAG, the wild type ATG for Ci was preserved) and a floxed 3xP3-dsRed cassette. The 3xP3 is an eye specific promoter that will mark transgenic flies for easy screening. (B) Targeted allele of *ci* as marked by 3xP3-dsRed. (C) Genotyping strategy. Primers shown in pink. (D-E)

Genotyping results. W^{1118} flies are used as a control in all genotyping reactions. The primers correspond to C.

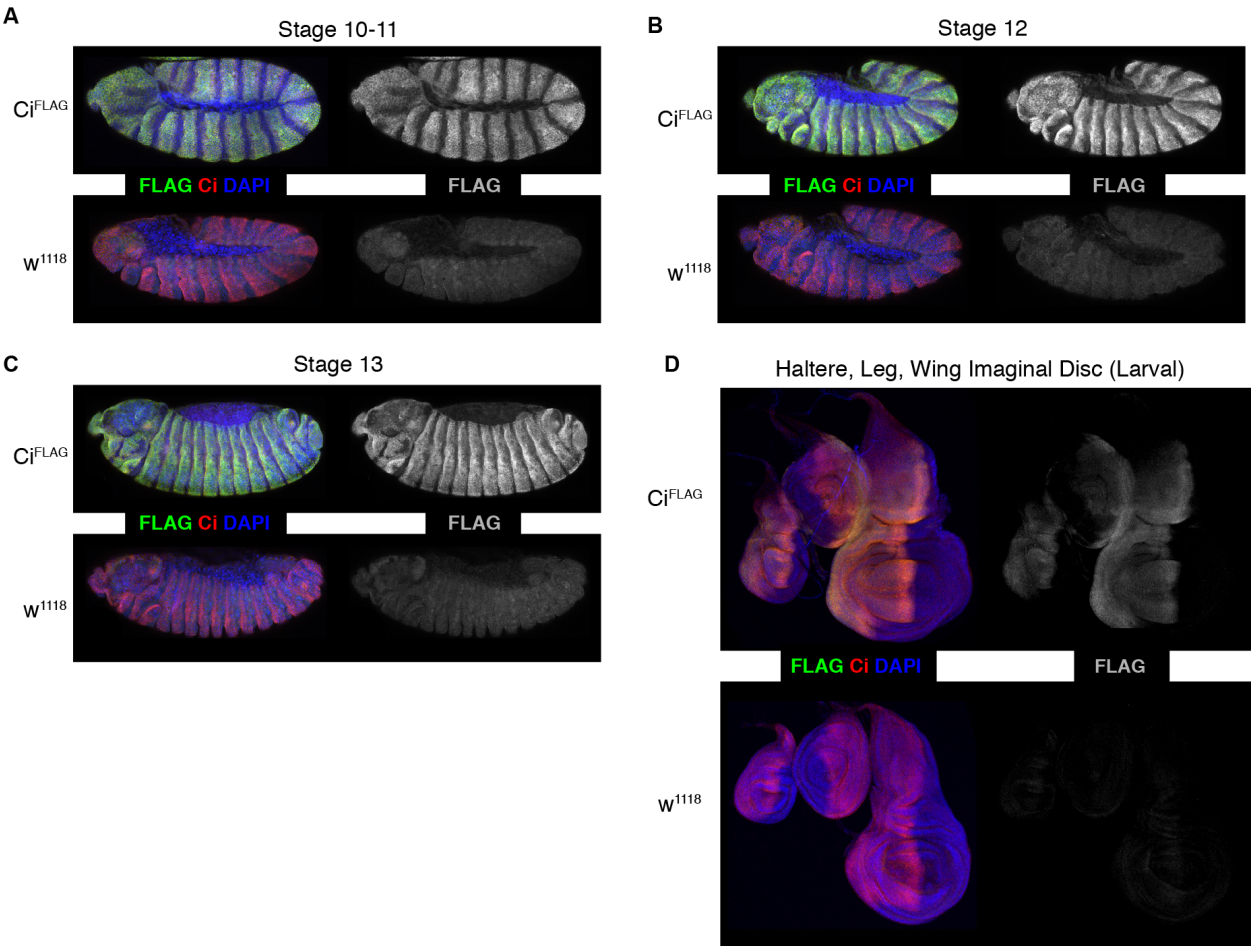


Figure 5.3 - Ci^{FLAG} is co-expressed with Ci in embryos and larval tissues

(A-C) Similarly staged transgenic and wild type embryos stained for FLAG and Ci. (D) Larval imaginal discs stained for FLAG and Ci. DAPI is blue throughout.

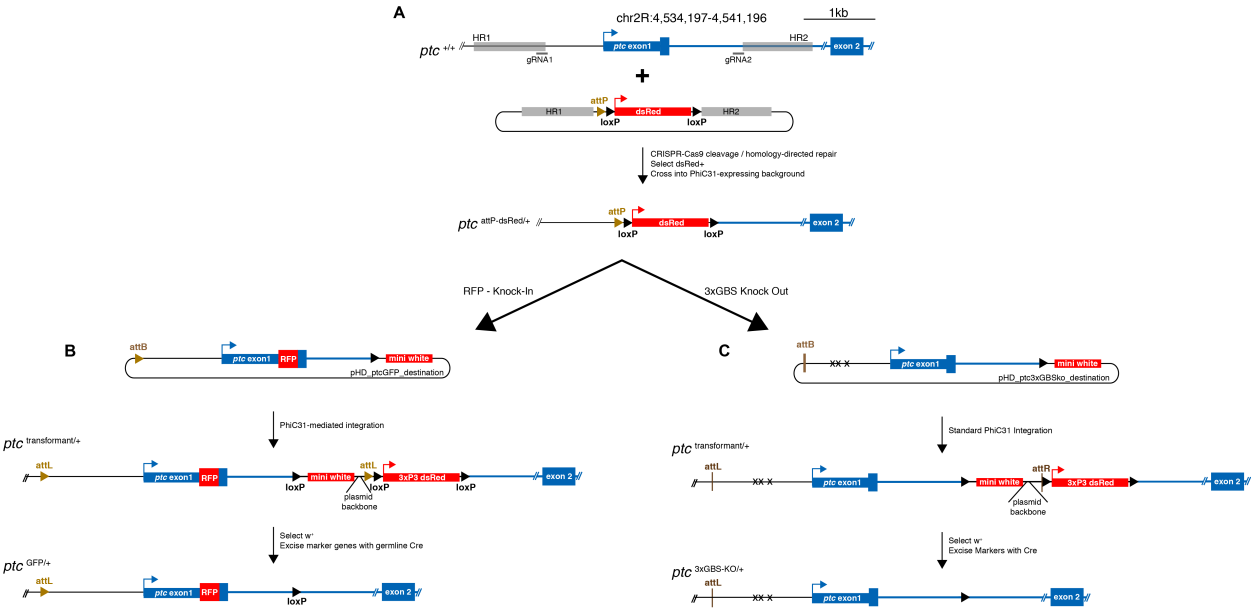


Figure 5.4 CRISPR/Cas9 strategy to target the *ptc* locus

(A) The *ptc* gene map is shown and is focused on the first exon and about 3kb flanking sequence in either direction, at top. The homology directed repair cassette is shown immediately underneath the *ptc* locus and contains a floxed 3xP3-dsRed cassette (same as in Figure 5.2), and an attP recombination site to replace the first exon of *ptc.* (B) Strategy to knock in RFP into the *ptc* locus using site-specific recombination (Bischof et al., 2007). (C) Strategy to knock out three consensus, high affinity GLI binding sites using a similar strategy as in B.

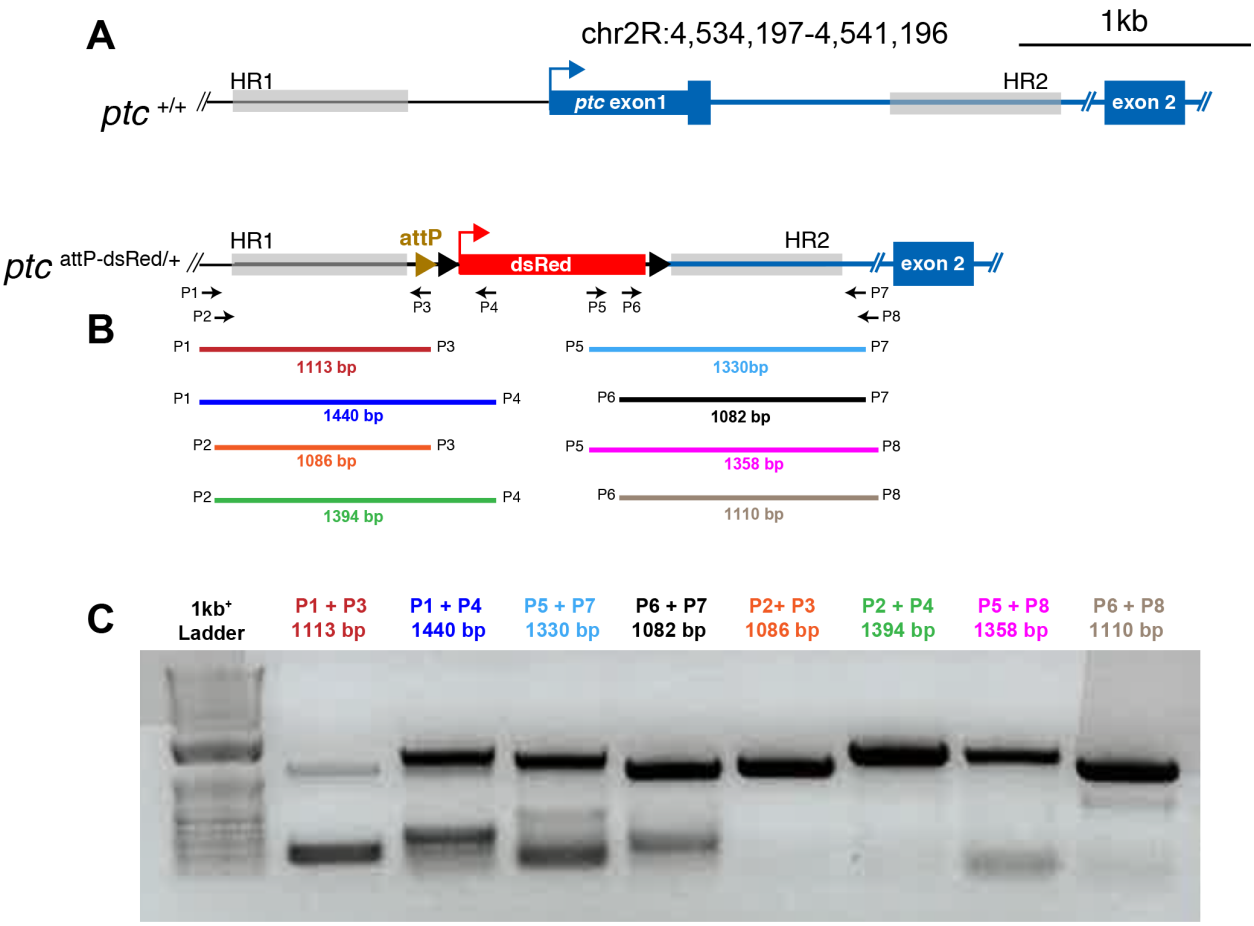


Figure 5.5 Genotyping results suggest *ptc* locus was successfully targeted

(A) A wild type and *ptc-attP* knockin allele are depicted. (B) Primers used for genotyping are shown as black arrows. Colored lines correspond to expected sizes of products. (C) Genotyping results, template DNA is from whole genomic preps of flies containing the ptc^{attP} allele.

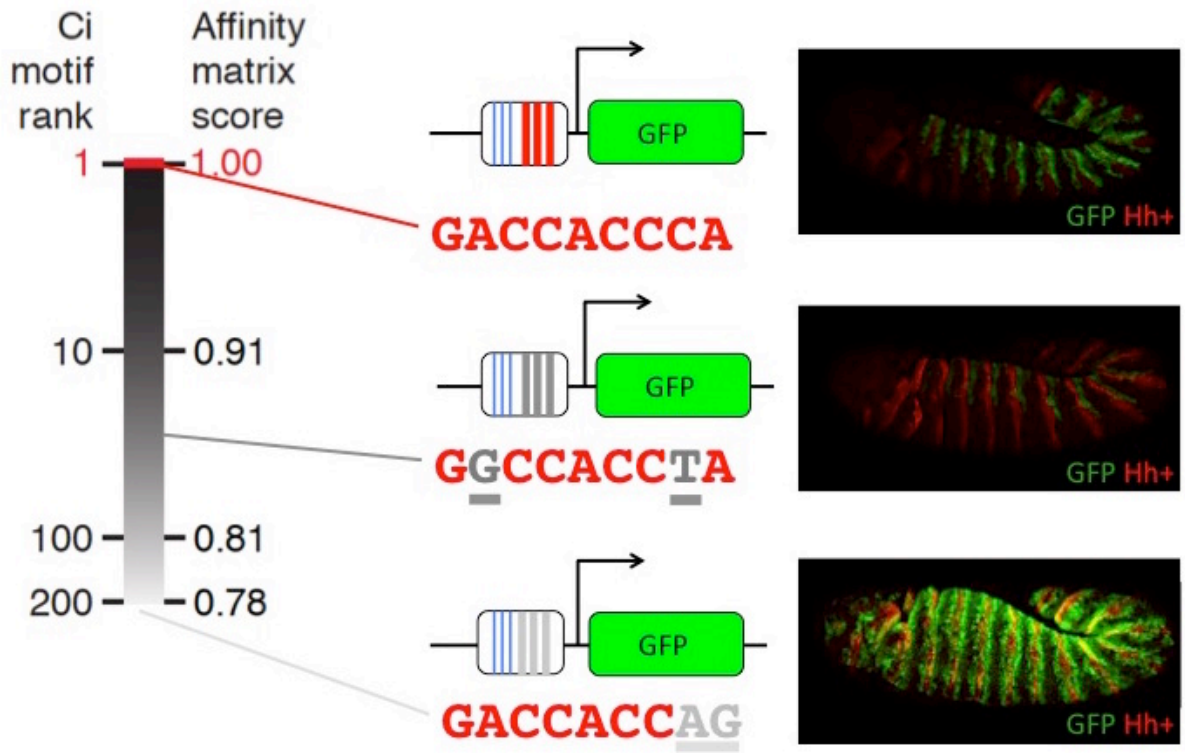


Figure 5.6 Low affinity GLI binding site enhancers respond strongly to Hh signaling

The grayscale chart on the left indicates Ci binding site quality (see Chapter 2 and 4 for more information). Synthetic enhancers each contained 3x binding sites for the broad transcriptional activator, Grainyhead, and 3x GLI binding sites of varying affinity, as predicted by shade of gray. Each reporter uses the hsp70 minimal promoter to drive GFP expression in transgenic animals. On the right, stage 12 embryos showing expression of each construct were examined.

Homology Regions Oligos	**Sequence**
ci_HR1_F	gatcCACCTGCgcat**TCGC**CGATGCATTAAGTACACTGA
ci_HR1_R	gatgCACCTGCgcat**ACTT**TTGATCGGCAACTTAAACAA
ci_HR2_F	gatcGCTCTTCa**TAT**caacttgtatctgatagtcg
ci_HR2_R	gatgGCTCTTCa**GAC**ccatacttttaaggttac
ptc_HR1_F	gatcCACCTGCgcat**TCGC**AATGCGCCACACGCAGAGTG
ptc_HR1_R	gatgCACCTGCgcat**CTAC**GAGCTTTCGTCTCAGTGTGT
ptc_HR2_F	gatcGCTCTTCa**TAT**CTTTGGCAGCCGCAGCAAA
ptc_HR2_F	gatgGCTCTTCa**GAC**CTTGAGGCGGTCCACACAG

Notes: homology to ci or ptc DNA
Aar1
filler/spacer
FOR CLONING
SapI
PAM sequence (from cgg to ctt in Ci_HR3'_2Fmutpam and from tgg to tta in Ci_HR3'_3_Fmutpam)

gRNA Complimentary Oligos	
ci_gRNA1_F	CTTCGTTTAAGTTGCCGATCAAAAG
ci_gRNA1_R	AAACCTTTTGATCGGCAACTTAAAC
ci_gRNA2_F	CTTCGATTCACACTGGTATTAACAA
ci_gRNA2_R	AAACTTGTTAATACCAGTGTGAATC
ptc_gRNA1_F	CTTCGCACTGAGACGAAAGCTCCAT
ptc_gRNA1_R	AAACATGGAGCTTTCGTCTCAGTGC
ptc_gRNA2_F	CTTCGAGGCAAAAAGCAGAAAGCTT
ptc_gRNA2_R	AAACAAGCTTTCTGCTTTTTGCCTC

Sequencing Primers	
ci_P1	tgttggattgttagattacc
ci_P2	ctagtgacgcctacactgcta
ci_P3	GAAGCCGAGCGCATAAGCTCAC
ci_P4	GAAAATATGTAGGTAACGCG
ptc_P1	CTCGAGAGCAAGCCAAAGATG
ptc_P2	GCTAAAATATGCATGTTGG
ptc_P3	tgagagaactcaaaggttac
ptc_P4	CTCCTTGATGACGTCCTCGGA
ptc_P5	CATCGTGGAGCAGTACGAGC
ptc_P6	CACTGCATTCTAGTTGTGGT
ptc_P7	GCATCAGCGAAGCAGCGAGA
ptc_P8	CAATGATGAAGACATCAAGC

Table 5.1 Primers used for CRISPR/Cas9 genomic engineering of *ptc* and *ci*

All primers are written from 5’ to 3’ orientation. F refers to forward primer, R refers to reverse primer. Additional binding site to facilitate cloning are depicted in colors annotated within the table.

Chapter 6

Conclusion

6.1 Summary of Findings

Through an in-depth analysis of Hh dependent enhancers in flies and mice, I have significantly contributed to the fields of cellular, developmental, and molecular biology. I began by demonstrating that *patched*, a well studied Hh target gene and a constitutive target of Hh signaling, surprisingly requires a battery of context- and stage-specific enhancers to properly respond to signaling. By not relying on a single, master regulatory element, *patched* affords itself the ability to adapt to new situations and this helps to explain, at the *cis*-regulatory level, one way that this highly conserved cell signaling pathway is able to pattern such disparate tissues, like wings in flies and the neural tube in humans.

I next showed that nearly all of these novel *ptc* enhancers are regulated by GLI through low affinity binding sites, and that the quality of GLI binding site is a key factor used for interpreting tissue specific responses of these enhancers – while low affinity sites might be critical for activation in the context of the embryonic ectoderm, they may not be preferred in the wing imaginal disc. Building on this theme of tissue specificity, I went on to identify novel regulatory inputs into these enhancers such as dGATAe and RFX, highlighting that there is much more to learn about tissue specific gene regulation.

I then applied this information about how GLI regulates *ptc* enhancers to the development of an algorithm that scanned the entire *Drosophila* genome for novel enhancers for other Hh target genes, besides *ptc,* and identified 8 additional GLI-responsive enhancers. While this computational method of enhancer prediction was useful for finding some new enhancers, it still produced many false positives, mainly because many sequences might look like binding sites at a sequence level, but not all of them are functional.

Lingering questions about enhancer identification, lack of in vivo analysis of Ci binding at the *ptc* locus, and limited information about how enhancers might cooperate to regulate gene expression led me to employ the powerful genome editing technique of CRISPR/Cas9 to address these questions at the actual genomic loci in question. To these ends, I generated a FLAG tagged allele of Ci to do tissue specific ChIP analyses to better understand how Ci binds to enhancers and regulates target gene expression. These experiments will yield valuable insights into how the quality of Ci binding sites influences tissue specific gene responses. To better address the question of enhancer function, I generated an easily targetable allele of *ptc* to learn more about how the more than 30 newly identified enhancers contribute to morphogenetic changes during development and to cellular homeostasis into adulthood to prevent misregulation of cellular processes that lead to disease.

6.2 Implications

My in depth dissection of *ptc* regulation has contributed a great deal of insight into how genes are controlled at the *cis*-regulatory level. While not all genes are regulated with such a complex assortment of enhancers like *ptc*, these results highlight

that we should still consider the entire locus when asking how an enhancer contributes to gene expression. This is especially true in cases where genes are expressed in more than one tissue or timepoint throughout development, since there might be more than one enhancer controlling context specific expression. We have known that multiple regulatory elements control precise patterns of gene expression for many years and their placement in the gene locus is critically important; see the story about the Locus Control Region of the human *β-globin* gene (Tanimoto et al., 1999; Tolhuis et al., 2002). Enhancers most likely function by looping to their promoter in order to act at a distance to the gene they regulate, but they can also interact with other enhancers active at the same developmental timepoints (Ghavi-Helm et al., 2014; Whalen et al., 2016). We do not know exactly how these processes occur, but are able to visualize such events using the Chromatin Conformation Capture technique, which has led to the widespread ability to recognize Topological Activation Domains, or TADs, which provide insight into the 3D architecture of the genome and strongly suggest that this architecture plays a role in regulating gene expression (Dixon et al., 2012). The idea of non-coding regions of the genome being considered junk DNA has been thoroughly debunked, and gene regulation experts are constantly identifying new ways that DNA is able to code for key information to control gene expression. In order to fully understand these mechanisms, we must look at gene regulation from many angles: the bird's eye view of genome architecture is just as important as the nucleotides making up each enhancer.

My analysis of *ptc* and other Hh target genes has also made it clear that there is no all encompassing enhancer identification method able to accurately predict a complete list of enhancers. There are caveats and shortcomings to each enhancer-

detection method discussed, from a lack of functional data from ChIP-seq binding regions, to the misidentification of binding sites using purely sequence based analysis. My work suggests that the best way to identify enhancers is by using a combination of approaches. There is a wealth of genomic transcription factor binding data and enhancer signatures available, and this can easily be mistaken for a comprehensive list of all known enhancers. These immense datasets should be the starting point to identifying a regulatory element, not the end. Predicted enhancers should be examined both as stably integrated transgenic reporters and in their native contexts, by deletion using genomic engineering techniques, as it is most likely not the only enhancer regulating gene expression.

One specific implication of my characterization of the *ptc* locus deals with the non-Hh regulated expression of *ptc*. In order for Hh signaling to remain off, Ptc must inhibit Smo, even in the absence of Hh ligand (Figure 1.1). The only known mechanism by which *ptc* is activated, however, is by Hh signaling. As already extensively discussed in this work, this negative feedback mechanism is essential for maintaining proper levels of signaling across timepoints and tissues in all Hh-regulated tissues. By examining the locus in such detail, I was able to locate a promoter proximal region, termed *ptc*B, that did not contain predicted Ci binding sites and was distinct from the promoter, which was identified by its core promoter elements like the initiator sequence and downstream promoter element (Figure 2.11). *ptc*B was able to drive low-level expression when examined on the core *ptc* promoter (Figure 2.11), which might account for the non-GLI regulated expression of *ptc* necessary to keep the pathway off in the absence of Hh.

A broader interpretation of these data offers several evolutionary implications as well. While I spent much time and effort demonstrating that the *patched* loci in mice and flies are regulated similarly, it is extremely important to note that the actual enhancers for each gene are not conserved by sequence. This was demonstrated on an experimental level when Peak2, the mouse *Ptch1* enhancer, was examined in flies and did not recapitulate a Hh response in any imaginal discs (Chapter 3). Furthermore, when I searched vertebrate genomes for similar sequences using my *Drosophila* enhancers as templates, there were no matches, demonstrating that these enhancers are not conserved by sequence or position. Completing this same screen using the mouse DNA sequences as templates against the fly genome similarly produced no matches. So, while the sequences in these enhancers themselves are not strictly conserved by sequence, I was able to demonstrate that overall *cis*-regulatory logic of *patched* regulation is conserved; in both animals, the constitutive response of *patched* to Hh signaling is regulated by a battery of context specific enhancers rather than a single, master regulatory element. The enhancers themselves are probably not explicitly conserved since each organism has different requirements for gene expression – from different TF inputs into the enhancers to different morphological outcomes: flies need wings and humans need hands. In addition to this insight, my data bring up an interesting idea about how to think about TFs recognizing their cognate binding sites. While we know that binding sites can vary quite a bit in sequence, this is why they are represented as positional weight matrices, perhaps TFs do not care as much about the exact sequence they recognize as they do about the overall binding site affinity. For example, GACCACCCA is considered a consensus GLI binding site in all genomes

based on exhaustive EMSA and luciferase analysis (Chapter 2). Once nucleotide substitutions are made in this sequence, the affinity is lowered, and we consider the new sequence a low affinity binding site. Perhaps instead of thinking of these binding sites individually, we should be thinking about them as “affinity groups”, meaning they are not just important because of the nucleotide sequence, but for the affinity by which a TF can bind. Many different arrangements of nucleotides could achieve any particular affinity, rendering the exact sequences less important. This logic allows for greater evolutionary discretion – binding sites become more malleable and far less rigid. So, any mutation to a binding site would be less detrimental to changes in gene expression, especially when multiple enhancers regulate gene expression in a particular context. Since evolution usually cannot change the coding sequence of a gene without creating a drastic, and often detrimental, effect on the organism, fine scale changes made at the level of *cis*-regulatory elements can be an important tool for explaining how changes to gene expression occur over time.

6.3 Future Directions

While I have made several contributions to understanding gene regulation during my graduate career, there are many experiments and future projects that are needed to address these implications and to complete a more thorough analysis of *ptc* regulation during development and into adulthood.

I have mainly focused on two main tissues in my analysis of *ptc* regulation: the embryonic ectoderm and the wing imaginal disc. Perhaps the lowest hanging fruit left out of this dissertation is the characterization of how *ptc* responds to Hh signaling in additional developmental and adult contexts. I demonstrated that low affinity Ci binding

sites are critical in both wings and the embryonic ectoderm, and that the affinity of these sites helps tune signaling. In my initial screen of *ptc* enhancers, I identified elements that were active in nearly every single developmental context that relies on Hh signaling (Figure 2.3) including adult stem-cell dependent contexts like the testis and ovary (Michel et al., 2012; Sahai-Hernandez and Nystul, 2013). In the testis, Hh is secreted from hub cells, considered the signaling center of this tissue, and received by nearby somatic cyst cells, where *ptc* is known to be active (Michel et al., 2012). Interestingly, the Ptc antibody stain detects expression in the hub, where Hh is expressed as well. Our analysis confirmed that expression driven by the GC enhancer is also found in both the somatic cells and the hub itself (Figure 2.3G). This is unique from most other Hh responsive tissues like the wing or embryonic ectoderm because normally *ptc* is most active in cells responding to the ligand, not in cells secreting it. It would be interesting to more fully investigate this expression pattern and see if that unexpected expression is dependant on Ci binding sites or other factors. Similar to the testis, the adult ovary contains populations of stem cells. Hh signaling is essential for maintaining both ovarian germ and follicle stem cells (FSCs) (Sahai-Hernandez and Nystul, 2013; Zhang and Kalderon, 2001). In relation to my analysis, FSCs are found in the region where enhancer 1AC is most strongly activate, making it a prime candidate for an FSC enhancer (Figure 2.3H). Scrutinizing this additional adult context of the ovary would be very important to help us better understand how *ptc* responds to Hh signaling to learn more about how Hh is maintaining these distinct populations of adult stem cells. Furthermore, a functional analysis of these enhancers at the endogenous *ptc* locus would be extremely informative to understanding the *ptc* response, Hh regulation, and

role in maintaining these stem cell dependent contexts. These contexts might also provide new, non-canonical sites of Hh signaling in *Drosophila.*

Another direction this project should take is derived from results presented in Chapter 3, which suggest that Hh signaling might also regulate salivary gland development in flies and this regulation might be augmented by the ciliary factor RFX. There is not much known about Hh signaling in the *Drosophila* salivary gland, so the first steps would be to characterize expression of all the pathway components, like Hh, Ptc, Smo, Cos2, Fused and Ci to show that the required components are present. There are either antibodies or *LacZ* knock-in alleles for all of these key proteins, so detection should be possible. Next, we should be able to augment signaling to demonstrate a change in signaling or development. We could achieve this by removing any Hh signaling components genetically and measuring expression of a target gene, like *ptc,* and also examine the phenotypic consequences of disrupting signaling. If we could establish that Hh signaling is important for salivary gland development, then we can ask if this tissue is ciliated and how this might also be important for development. Most cell types in *Drosophila* are not ciliated, although there is one notable exception: the olfactory sensory neurons, which require cilia for normal development (Vieillard et al., 2015) and references therein). Interestingly, Hh signaling occurs in these cell types and many of the components of the pathway are processed in cilia suggesting a possible non-canonical path for Hh signaling in the fly (Kuzhandaivel et al., 2014). To then apply this to my work, we would need to demonstrate that *Drosophila* salivary glands are ciliated by using any number of cilia markers available to the fly community such as Arl13b::GFP or CBY::GFP, both GFP tagged versions of ciliary proteins expressed in

the proximal ciliary segment and transition zone, respectively (Enjolras et al., 2012). Using these lines to examine expression in salivary glands would help us determine if cilia are present in this tissue. We could then merge these two experiments to see if Hh pathway components are being processed in salivary gland cilia, and begin asking how this potential cilia-mediated Hh signaling contributes to salivary gland development and function. It is also of note to mention that mammalian salivary glands are ciliated (Bernfield et al., 1972) and also require SHH signaling for branching morphogenesis during development (Jaskoll et al., 2004). This suggests that *Drosophila* salivary glands might be another evolutionary link between signaling in flies, mice and humans, since cilia play a significant role in Hh signaling in most mammalian contexts. It is also a potential opportunity to characterize another non-canonical pathway for Hh signaling in *Drosophila*.

One of the most significant (and I believe most important) future directions of my work is to explore the function of all of these newly identified enhancers in their native genomic context. This is absolutely critical to understanding how Hh signaling contributes to development and adult tissue maintenance. To address these essential questions, I have adapted the CRISPR/Cas9 genome editing technology for use in the Barolo lab and started several critical experiments using the technique (see Chapter 5 for a longer discussion). The first experiment in progress is tagging Ci with a 3xFLAG epitope to better understand where and how Ci binds to enhancers in different developmental contexts. Nearly all of what is known about how Ci binds to DNA has been extrapolated from luciferase assays and EMSA analyses. While these provide critical information, they do not tell us the whole story of how Ci regulates Hh signaling

and need to be complemented with ChIP-seq analyses to give us more information about co-factors and binding conditions in Ci's native genomic context in flies (Chapter 5). I also have enlisted the expertise of the Kassis Lab at the NIH for a collaboration using this allele for tissue specific ChIP-seq analysis. The second experiment that I have started is the generation of the ptc^{attP} allele, which contains an attP recombination site in place of exon1 of *ptc*. This allele will allow us to address several lingering questions about how *ptc* is regulated by the enhancers I have characterized in this dissertation and provide insights into how these enhancers work together to contribute to morphogenesis of different tissues (Figure 5.4). By manipulating the *cis*-regulatory elements that control *ptc* regulation at the native locus, we will be able to see how enhancers are working together to regulate expression, how they interact with the promoter, and be able to better address the role of Polycomb and Trithorax regulation at this locus. Most importantly, we can finally address the functional contribution these elements make to the development of the organism.

While there is still much to learn about how the Hh pathway regulates its target genes, this in-depth analysis of gene regulation has uncovered a highly evolutionarily conserved mechanism by which GLI is able to transduce its signal in many different tissues and cell types using an array of highly context specific enhancers. The future directions discussed here, and throughout the rest of my thesis, will significantly contribute to our understanding of how cells communicate via transcriptional cues to create a healthy adult using the information contained within a single genome.

References

Adkins, N. L., Hagerman, T. A. and Georgel, P. (2006). GAGA protein: a multi-faceted transcription factor. *Biochem. Cell Biol.* **84**, 559–567.

Agren, M., Kogerman, P., Kleman, M. I., Wessling, M. and Toftgård, R. (2004). Expression of the PTCH1 tumor suppressor gene is regulated by alternative promoters and a single functional Gli-binding site. *Gene* **330**, 101–114.

Alexandre, C., Jacinto, A. and Ingham, P. W. (1996). Transcriptional activation of hedgehog target genes in Drosophila is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes & Development* **10**, 2003–2013.

Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M. and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077.

Ashique, A. M., Choe, Y., Karlen, M., May, S. R., Phamluong, K., Solloway, M. J., Ericson, J. and Peterson, A. S. (2009). The Rfx4 transcription factor modulates Shh signaling by regional control of ciliogenesis. *Sci Signal* **2**, ra70.

Avery, O. T., Macleod, C. M. and McCarty, M. (1944). STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *J. Exp. Med.* **79**, 137–158.

Aza-Blanc, P., Ramírez-Weber, F. A., Laget, M. P., Schwartz, C. and Kornberg, T. B. (1997). Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* **89**, 1043–1053.

Banerji, J., Rusconi, S. and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308.

Barakat, M. T., Humke, E. W. and Scott, M. P. (2010). Learning from Jekyll to control Hyde: Hedgehog signaling in development and cancer. *Trends Mol Med* **16**, 337–348.

Barolo, S. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes & Development* **16**, 1167–1181.

Barolo, S. (2006). Transgenic Wnt/TCF pathway reporters: all you need is Lef? *Oncogene* **25**, 7505–7511.

Barolo, S. (2011a). Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays*.

Barolo, S. (2011b). Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* **34**, 135–141.

Barolo, S., Carver, L. A. and Posakony, J. W. (2000). GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. *BioTechniques* **29**, 726, 728, 730, 732.

Basler, K. and Struhl, G. (1994). Compartment boundaries and the control of Drosophila limb pattern by hedgehog protein. *Nature* **368**, 208–214.

Bassett, A. R., Tibbit, C., Ponting, C. P. and Liu, J.-L. (2013). Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell Rep* **4**, 220–228.

Bellusci, S., Furuta, Y., Rush, M. G., Henderson, R., Winnier, G. and Hogan, B. L. (1997). Involvement of Sonic hedgehog (Shh) in mouse embryonic lung growth and morphogenesis. *Development* **124**, 53–63.

Benoist, C. and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. *Nature* **290**, 304–310.

Bergsland, M., Ramsköld, D., Zaouter, C., Klum, S., Sandberg, R. and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. *Genes & Development* **25**, 2453–2464.

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

Berman, B. P., Pfeiffer, B. D., Laverty, T. R., Salzberg, S. L., Rubin, G. M., Eisen, M. B. and Celniker, S. E. (2004). Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in Drosophila melanogaster and Drosophila pseudoobscura. *Genome Biology* **5**, R61.

Bernfield, M. R., Banerjee, S. D. and Cohn, R. H. (1972). Dependence of salivary epithelial morphology and branching morphogenesis upon acid mucopolysaccharide-protein (proteoglycan) at the epithelial surface. *J. Cell Biol.* **52**, 674–689.

Biehs, B., Kechris, K., Liu, S. and Kornberg, T. B. (2010). Hedgehog targets in the Drosophila embryo and the mechanisms that generate tissue-specific outputs of Hedgehog signaling. *Development* **137**, 3887–3898.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F. and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases.

Proc. Natl. Acad. Sci. U.S.A. **104**, 3312–3317.

Blair, S. S. and Ralston, A. (1997). Smoothened-mediated Hedgehog signalling is required for the maintenance of the anterior-posterior lineage restriction in the developing wing of Drosophila. *Development* **124**, 4053–4063.

Blanch, M., Piñeyro, D. and Bernués, J. (2015). New insights for Drosophila GAGA factor in larvae. *R Soc Open Sci* **2**, 150011.

Blanchette, M., Kent, W. J., Riemer, C., Elnitski, L., Smit, A. F. A., Roskin, K. M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E. D., et al. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715.

Blanco, J., Seimiya, M., Pauli, T., Reichert, H. and Gehring, W. J. (2009). Wingless and Hedgehog signaling pathways regulate orthodenticle and eyes absent during ocelli development in Drosophila. *Dev. Biol.* **329**, 104–115.

Bouchard, M., St-Amand, J. and Côté, S. (2000). Combinatorial activity of pair-rule proteins on the Drosophila gooseberry early enhancer. *Dev. Biol.* **222**, 135–146.

Bowman, S. K., Deaton, A. M., Domingues, H., Wang, P. I., Sadreyev, R. I., Kingston, R. E. and Bender, W. (2014). H3K27 modifications define segmental regulatory domains in the Drosophila bithorax complex. *Elife* **3**, e02833.

Boy, A. L., Zhai, Z., Habring-Müller, A., Kussler-Schneider, Y., Kaspar, P. and Lohmann, I. (2010). Vectors for efficient and high-throughput construction of fluorescent drosophila reporters using the PhiC31 site-specific integration system. *Genesis* **48**, 452–456.

Briscoe, J. and Thérond, P. P. (2013). The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* **14**, 416–429.

Brown, J. L. and Kassis, J. A. (2013). Architectural and functional diversity of polycomb group response elements in Drosophila. *Genetics* **195**, 407–419.

Buttitta, L., Mo, R., Hui, C.-C. and Fan, C.-M. (2003). Interplays of Gli2 and Gli3 and their requirement in mediating Shh-dependent sclerotome induction. *Development* **130**, 6233–6243.

Cachero, S., Simpson, T. I., Lage, zur, P. I., Ma, L., Newton, F. G., Holohan, E. E., Armstrong, J. D. and Jarman, A. P. (2011). The gene regulatory cascade linking proneural specification with differentiation in Drosophila sensory neurons. *PLoS Biol* **9**, e1000568.

Cadigan, K. M., Grossniklaus, U. and Gehring, W. J. (1994). Localized expression of sloppy paired protein maintains the polarity of Drosophila parasegments. *Genes & Development* **8**, 899–913.

Calleja, M., Herranz, H., Estella, C., Casal, J., Lawrence, P., Simpson, P. and Morata, G. (2000). Generation of medial and lateral dorsal body domains by the pannier gene of Drosophila. *Development* **127**, 3971–3980.

Carpenter, B. S., Barry, R. L., Verhey, K. J. and Allen, B. L. (2015). The heterotrimeric kinesin-2 complex interacts with and regulates GLI protein function. *Journal of Cell Science* **128**, 1034–1050.

Celniker, S. E., Dillon, L. A. L., Gerstein, M. B., Gunsalus, K. C., Henikoff, S., Karpen, G. H., Kellis, M., Lai, E. C., Lieb, J. D., MacAlpine, D. M., et al. (2009). Unlocking the secrets of the genome. *Nature* **459**, 927–930.

Celniker, S. E., Wheeler, D. A., Kronmiller, B., Carlson, J. W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S. P., Frise, E., et al. (2002). Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence. *Genome Biology* **3**, RESEARCH0079.

Chen, C.-H., Kessler, von, D. P., Park, W., Wang, B., Ma, Y. and Beachy, P. A. (1999). Nuclear Trafficking of Cubitus interruptus in the Transcriptional Regulation of Hedgehog Target Gene Expression. *Cell* **98**, 305–316.

Chen, Y. and Struhl, G. (1996). Dual roles for patched in sequestering and transducing Hedgehog. *Cell* **87**, 553–563.

Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H. and Beachy, P. A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**, 407–413.

Choksi, S. P., Lauter, G., Swoboda, P. and Roy, S. (2014). Switching on cilia: transcriptional networks regulating ciliogenesis. *Development* **141**, 1427–1441.

Christensen, R. G., Enuameh, M. S., Noyes, M. B., Brodsky, M. H., Wolfe, S. A. and Stormo, G. D. (2012). Recognition models to predict DNA-binding specificities of homeodomain proteins. *Bioinformatics* **28**, i84–9.

Cobourne, M. T., Xavier, G. M., Depew, M., Hagan, L., Sealby, J., Webster, Z. and Sharpe, P. T. (2009). Sonic hedgehog signalling inhibits palatogenesis and arrests tooth development in a mouse model of the nevoid basal cell carcinoma syndrome. *Dev. Biol.* **331**, 38–49.

Coward, E. (1999). Shufflet: shuffling sequences while conserving the k-let counts. *Bioinformatics* **15**, 1058–1059.

Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., et al. (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191–203.

Dahm, R. (2005). *Friedrich Miescher and the discovery of DNA.*

Dai, P., Akimaru, H., Tanaka, Y., Maekawa, T., Nakafuku, M. and Ishii, S. (1999). Sonic Hedgehog-induced activation of the Gli1 promoter is mediated by GLI3. *J. Biol. Chem.* **274**, 8143–8152.

de Sauvage, F. (2007). The Hh signaling pathway in cancer. *Bull. Mem. Acad. R. Med. Belg.* **162**, 219–223.

Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002). Capturing chromosome conformation. *Science* **295**, 1306–1311.

Dessaud, E., McMahon, A. P. and Briscoe, J. (2008). Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development* **135**, 2489–2503.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380.

Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development* **134**, 2549–2560.

Dubruille, R. (2002). Drosophila Regulatory factor X is necessary for ciliated sensory neuron differentiation. *Development* **129**, 5487–5498.

Eaton, S. and Kornberg, T. B. (1990). Repression of ci-D in posterior compartments of Drosophila by engrailed. *Genes & Development* **4**, 1068–1077.

Elgar, G. and Vavouri, T. (2008). Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet.* **24**, 344–352.

Emanuel, P. A. and Gilmour, D. S. (1993). Transcription factor TFIID recognizes DNA sequences downstream of the TATA element in the Hsp70 heat shock gene. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8449–8453.

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

Enjolras, C., Thomas, J., Chhin, B., Cortier, E., Duteyrat, J.-L., Soulavie, F., Kernan, M. J., Laurençon, A. and Durand, B. (2012). Drosophila chibby is required for basal body formation and ciliogenesis but not for Wg signaling. *J. Cell Biol.* **197**, 313–325.

Epstein, E. H. (2008). Basal cell carcinomas: attack of the hedgehog. *Nat Rev Cancer* **8**, 743–754.

Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S. and Levine, M. S. (2015). Suboptimization of developmental enhancers. *Science* **350**, 325–328.

Fietz, M. J., Jacinto, A., Taylor, A. M., Alexandre, C. and Ingham, P. W. (1995). Secretion of the amino-terminal fragment of the hedgehog protein is necessary and sufficient for hedgehog signalling in Drosophila. *Curr. Biol.* **5**, 643–650.

Fitch, W. M. (1983). Random sequences. *Journal of Molecular Biology* **163**, 171–176.

Flores, G. V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M. and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. *Cell* **103**, 75–85.

Forbes, A. J., Nakano, Y., Taylor, A. M. and Ingham, P. W. (1993). Genetic analysis of hedgehog signalling in the Drosophila embryo. *Dev. Suppl.* 115–124.

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

Fromm, M. and Berg, P. (1983). Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. *Mol. Cell. Biol.* **3**, 991–999.

Fu, W. and Noll, M. (1997). The Pax2 homolog sparkling is required for development of cone and pigment cells in the Drosophila eye. *Genes & Development* **11**, 2066–2078.

Fu, W., Duan, H., Frei, E. and Noll, M. (1998). shaven and sparkling are mutations in separate enhancers of the Drosophila Pax2 homolog. *Development* **125**, 2943–2950.

Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., Goldman, M., Barber, G. P., Clawson, H., Coelho, A., et al. (2011). The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876–82.

Galas, D. J. and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5**, 3157–3170.

Gebelein, B., McKay, D. J. and Mann, R. S. (2004). Direct integration of Hox and segmentation gene inputs during Drosophila development. *Nature* **431**, 653–659.

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W. and Furlong, E. E. M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100.

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. and Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877–885.

Goddeeris, M. M., Rho, S., Petiet, A., Davenport, C. L., Johnson, G. A., Meyers, E.

N. and Klingensmith, J. (2008). Intracardiac septation requires hedgehog-dependent cellular contributions from outside the heart. *Development* **135**, 1887–1895.

Goodrich, L. V., Johnson, R. L., Milenkovic, L., McMahon, J. A. and Scott, M. P. (1996). Conservation of the hedgehog/patched signaling pathway from flies to mice: induction of a mouse patched gene by Hedgehog. *Genes & Development* **10**, 301–312.

Gotea, V., Visel, A., Westlund, J. M., Nobrega, M. A., Pennacchio, L. A. and Ovcharenko, I. (2010). Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res.* **20**, 565–577.

Gurdziel, K., Lorberbaum, D. S., Udager, A. M., Song, J. Y., Richards, N., Parker, D. S., Johnson, L. A., Allen, B. L., Barolo, S. and Gumucio, D. L. (2015). Identification and Validation of Novel Hedgehog-Responsive Enhancers Predicted by Computational Analysis of Ci/Gli Binding Site Density. *PLoS ONE* **10**, e0145225.

Hahn, H., Wicking, C., Zaphiropoulous, P. G., Gailani, M. R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Unden, A. B., Gillies, S., et al. (1996). Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell* **85**, 841–851.

Halfon, M. S., Zhu, Q., Brennan, E. R. and Zhou, Y. (2011). Erroneous attribution of relevant transcription factor binding sites despite successful prediction of cis-regulatory modules. *BMC Genomics 2008 9:1* **12**, 578.

Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E. and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* **124**, 47–59.

Hammonds, A. S., Bristow, C. A., Fisher, W. W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E. and Celniker, S. E. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development. *Genome Biology* **14**, R140.

Hendrix, D. A., Hong, J.-W., Zeitlinger, J., Rokhsar, D. S. and Levine, M. S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. *Proceedings of the National Academy of Sciences* **105**, 7762–7767.

Herranz, H. and Morata, G. (2001). The functions of pannier during Drosophila embryogenesis. *Development* **128**, 4837–4846.

Hersh, B. M. and Carroll, S. B. (2005). Direct regulation of knot gene expression by Ultrabithorax and the evolution of cis-regulatory elements in Drosophila. *Development* **132**, 1567–1577.

Hidalgo, A. and Ingham, P. (1990). Cell patterning in the Drosophila segment: spatial

regulation of the segment polarity gene patched. *Development* **110**, 291–301.

Hildebrandt, F., Benzing, T. and Katsanis, N. (2011). Ciliopathies. *N. Engl. J. Med.* **364**, 1533–1543.

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A. and Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947.

Hoffmann, A. D., Peterson, M. A., Friedland-Little, J. M., Anderson, S. A. and Moskowitz, I. P. (2009). sonic hedgehog is required in pulmonary endoderm for atrial septation. *Development* **136**, 1761–1770.

Hoffmann, A. D., Yang, X. H., Burnicka-Turek, O., Bosman, J. D., Ren, X., Steimle, J. D., Vokes, S. A., McMahon, A. P., Kalinichenko, V. V. and Moskowitz, I. P. (2014). Foxf genes integrate tbx5 and hedgehog pathways in the second heart field for cardiac septation. *PLoS Genet.* **10**, e1004604.

Holohan, E. E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S. and White, R. (2007). CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. *PLoS Genet.* **3**, e112.

Hooper, J. E. and Scott, M. P. (1989). The Drosophila patched gene encodes a putative membrane protein required for segmental patterning. *Cell* **59**, 751–765.

Hui, C.-C. and Angers, S. (2011). Gli proteins in development and disease. *Annu. Rev. Cell Dev. Biol.* **27**, 513–537.

Infante, P., Mori, M., Alfonsi, R., Ghirga, F., Aiello, F., Toscano, S., Ingallina, C., Siler, M., Cucchi, D., Po, A., et al. (2015). Gli1/DNA interaction is a druggable target for Hedgehog-dependent tumors. *The EMBO Journal* **34**, 200–217.

Ingham, P. W. (2016). Drosophila Segment Polarity Mutants and the Rediscovery of the Hedgehog Pathway Genes. In *sciencedirect.com*, pp. 477–488. Elsevier.

Ingham, P. W. and McMahon, A. P. (2001). Hedgehog signaling in animal development: paradigms and principles. *Genes & Development* **15**, 3059–3087.

Ingham, P. W., Nakano, Y. and Seger, C. (2011). Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat Rev Genet* **12**, 393–406.

Ingham, P. W., Taylor, A. M. and Nakano, Y. (1991). Role of the Drosophila patched gene in positional signalling. *Nature* **353**, 184–187.

Jacinto, A., Woolner, S. and Martin, P. (2002). Dynamic analysis of dorsal closure in Drosophila: from genetics to cell biology. *Dev. Cell* **3**, 9–19.

Jaskoll, T., Leo, T., Witcher, D., Ormestad, M., Astorga, J., Bringas, P., Carlsson,

P. and Melnick, M. (2004). Sonic hedgehog signaling plays an essential role during embryonic salivary gland epithelial branching morphogenesis. *Dev. Dyn.* **229**, 722–732.

Jenett, A., Rubin, G. M., Ngo, T.-T. B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B. D., Cavallaro, A., Hall, D., Jeter, J., et al. (2012). A GAL4-driver line resource for Drosophila neurobiology. *Cell Rep* **2**, 991–1001.

Jeong, J. and McMahon, A. P. (2005). Growth and pattern of the mammalian neural tube are governed by partially overlapping feedback activities of the hedgehog antagonists patched 1 and Hhip1. *Development* **132**, 143–154.

Jiang, J. and Hui, C.-C. (2008). Hedgehog signaling in development and cancer. *Dev. Cell* **15**, 801–812.

Johnson, D. S., Mortazavi, A., Myers, R. M. and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502.

Johnson, L. A., Zhao, Y., Golden, K. and Barolo, S. (2008). Reverse-engineering a transcriptional enhancer: a case study in Drosophila. *Tissue Eng Part A* **14**, 1549–1559.

Johnson, R. L., Rothman, A. L., Xie, J., Goodrich, L. V., Bare, J. W., Bonifas, J. M., Quinn, A. G., Myers, R. M., Cox, D. R., Epstein, E. H., et al. (1996). Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science* **272**, 1668–1671.

Joiner, A. M., Green, W. W., McIntyre, J. C., Allen, B. L., Schwob, J. E. and Martens, J. R. (2015). Primary Cilia on Horizontal Basal Cells Regulate Regeneration of the Olfactory Epithelium. *J. Neurosci.* **35**, 13761–13772.

Jory, A., Estella, C., Giorgianni, M. W., Slattery, M., Laverty, T. R., Rubin, G. M. and Mann, R. S. (2012). A survey of 6,300 genomic fragments for cis-regulatory activity in the imaginal discs of Drosophila melanogaster. *Cell Rep* **2**, 1014–1024.

Kassis, J. A. and Brown, J. L. (2013). Polycomb Group Response Elements in Drosophila and Vertebrates. In *Advances in Genetics*, pp. 83–118. Elsevier.

Keightley, P. D., Trivedi, U., Thomson, M., Oliver, F., Kumar, S. and Blaxter, M. L. (2009). Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. *Genome Res.* **19**, 1195–1201.

Kent, D. (2006). Roadkill attenuates Hedgehog responses through degradation of Cubitus interruptus. *Development* **133**, 2001–2010.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H. R., Hwang, J., Kim, J.-I. and Kim, J.-S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–43– 1 p following 243.

Kinzler, K. W. and Vogelstein, B. (1990). The GLI gene encodes a nuclear protein which binds specific sequences in the human genome. *Mol. Cell. Biol.* **10**, 634–642.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z. and Joung, J. K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495.

Kuzhandaivel, A., Schultz, S. W., Alkhori, L. and Alenius, M. (2014). Cilia-mediated hedgehog signaling in Drosophila. *Cell Rep* **7**, 672–680.

Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., Dickson, B. J. and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91–95.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Lang, M., Hadzhiev, Y., Siegel, N., Amemiya, C. T., Parada, C., Strähle, U., Becker, M.-B., Müller, F. and Meyer, A. (2010). Conservation of shh cis-regulatory architecture of the coelacanth is consistent with its ancestral phylogenetic position. *Evodevo* **1**, 11.

Lee, E. Y., Ji, H., Ouyang, Z., Zhou, B., Ma, W., Vokes, S. A., McMahon, A. P., Wong, W. H. and Scott, M. P. (2010). Hedgehog pathway-regulated gene networks in cerebellum development and tumorigenesis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9736–9741.

Lei, Q., Jeong, Y., Misra, K., Li, S., Zelman, A. K., Epstein, D. J. and Matise, M. P. (2006). Wnt signaling inhibitors regulate the transcriptional response to morphogenetic Shh-Gli signaling in the neural tube. *Dev. Cell* **11**, 325–337.

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E. and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735.

Lettice, L. A., Hill, A. E., Devenney, P. S. and Hill, R. E. (2008). Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum. Mol. Genet.* **17**, 978–985.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–63.

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

Li, X. and Noll, M. (1993). Role of the gooseberry gene in Drosophila embryos:

maintenance of wingless expression by a wingless--gooseberry autoregulatory loop. *The EMBO Journal* **12**, 4499–4509.

Lifanov, A. P., Makeev, V. J., Nazina, A. G. and Papatsenko, D. A. (2003). Homotypic regulatory clusters in Drosophila. *Genome Res.* **13**, 579–588.

Lipinski, R. J., Song, C., Sulik, K. K., Everson, J. L., Gipp, J. J., Yan, D., Bushman, W. and Rowland, I. J. (2010). Cleft lip and palate results from Hedgehog signaling antagonism in the mouse: Phenotypic characterization and clinical implications. *Birth Defects Res. Part A Clin. Mol. Teratol.* **88**, 232–240.

Lopez-Rios, J., Duchesne, A., Speziale, D., Andrey, G., Peterson, K. A., Germann, P., Unal, E., Liu, J., Floriot, S., Barbey, S., et al. (2014). Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. *Nature* **511**, 46–51.

Lorberbaum, D. S. and Barolo, S. (2015). Enhancers: holding out for the right promoter. *Curr. Biol.* **25**, R290–3.

Lorberbaum, D. S., Ramos, A. I., Peterson, K. A., Carpenter, B. S., Parker, D. S., De, S., Hillers, L. E., Blake, V. M., Nishi, Y., McFarlane, M. R., et al. (2016). An ancient yet flexible cis-regulatory architecture allows localized Hedgehog tuning by patched/Ptch1. *Elife* **5**.

Mann, R. S., Lelli, K. M. and Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. *Curr. Top. Dev. Biol.* **88**, 63–101.

Markstein, M., Markstein, P., Markstein, V. and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 763–768.

Martin, P. and Parkhurst, S. M. (2004). Parallels between tissue repair and embryo morphogenesis. *Development* **131**, 3021–3034.

Marx, V. (2013). Finding the right antibody for the job. *Nat. Methods* **10**, 703–707.

Mathelier, A., Fornes, O., Arenillas, D. J., Chen, C.-Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44**, D110–5.

McKay, D. J. and Lieb, J. D. (2013). A common set of DNA regulatory elements shapes Drosophila appendages. *Dev. Cell* **27**, 306–318.

McMahon, A. P., Ingham, P. W. and Tabin, C. J. (2003). Developmental roles and clinical significance of hedgehog signaling. *Curr. Top. Dev. Biol.* **53**, 1–114.

McQuilton, P., St Pierre, S. E., Thurmond, J.FlyBase Consortium (2012). FlyBase 101--the basics of navigating FlyBase. *Nucleic Acids Res.* **40**, D706–14.

Megason, S. G. and McMahon, A. P. (2002). A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. *Development* **129**, 2087–2098.

Metzakopian, E., Lin, W., Salmon-Divon, M., Dvinge, H., Andersson, E., Ericson, J., Perlmann, T., Whitsett, J. A., Bertone, P. and Ang, S.-L. (2012). Genome-wide characterization of Foxa2 targets reveals upregulation of floor plate genes and repression of ventrolateral genes in midbrain dopaminergic progenitors. *Development* **139**, 2625–2634.

Méthot, N. and Basler, K. (2001). An absolute requirement for Cubitus interruptus in Hedgehog signaling. *Development* **128**, 733–742.

Michel, M., Kupinski, A. P., Raabe, I. and Bökel, C. (2012). Hh signalling is essential for somatic stem cell maintenance in the Drosophila testis niche. *Development*.

Milenkovic, L., Goodrich, L. V., Higgins, K. M. and Scott, M. P. (1999). Mouse patched1 controls body size determination and limb patterning. *Development* **126**, 4431–4440.

Millard, T. H. and Martin, P. (2008). Dynamic analysis of filopodial interactions during the zippering phase of Drosophila dorsal closure. *Development* **135**, 621–626.

Mohd-Sarip, A., Cléard, F., Mishra, R. K., Karch, F. and Verrijzer, C. P. (2005). Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes & Development* **19**, 1755–1760.

Mohler, J. (1988). Requirements for hedgehog, a segmental polarity gene, in patterning larval and adult cuticle of Drosophila. *Genetics* **120**, 1061–1072.

Motoyama, J., Milenkovic, L., Iwama, M., Shikata, Y., Scott, M. P. and Hui, C.-C. (2003). Differential requirement for Gli2 and Gli3 in ventral neural cell fate specification. *Dev. Biol.* **259**, 150–161.

Motzny, C. K. and Holmgren, R. (1995). The Drosophila cubitus interruptus protein and its role in the wingless and hedgehog signal transduction pathways. *Mech. Dev.* **52**, 137–150.

Mulinari, S. and Häcker, U. (2009). Hedgehog, but not Odd skipped, induces segmental grooves in the Drosophila epidermis. *Development* **136**, 3875–3880.

Murdoch, J. N. and Copp, A. J. (2010). The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. *Birth Defects Res. Part A Clin. Mol. Teratol.* **88**, 633–652.

Müller, B. and Basler, K. (2000). The repressor and activator forms of Cubitus interruptus control Hedgehog target genes through common generic gli-binding sites. *Development* **127**, 2999–3007.

Nakano, Y., Guerrero, I., Hidalgo, A., Taylor, A., Whittle, J. R. and Ingham, P. W. (1989). A protein with several possible membrane-spanning domains encoded by the Drosophila segment polarity gene patched. *Nature* **341**, 508–513.

Ngan, E. S.-W., Kim, K.-H. and Hui, C.-C. (2013). Sonic Hedgehog Signaling and VACTERL Association. *Mol Syndromol* **4**, 32–45.

Nishi, Y., Zhang, X., Jeong, J., Peterson, K. A., Vedenko, A., Bulyk, M. L., Hide, W. A. and McMahon, A. P. (2015). A direct fate exclusion mechanism by Sonic hedgehog-regulated transcriptional repressors. *Development* **142**, 3286–3293.

Niswander, L., Jeffrey, S., Martin, G. R. and Tickle, C. (1994). A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* **371**, 609–612.

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

Nybakken, K., Vokes, S. A., Lin, T.-Y., McMahon, A. P. and Perrimon, N. (2005). A genome-wide RNA interference screen in Drosophila melanogaster cells for new components of the Hh signaling pathway. *Nat Genet* **37**, 1323–1332.

Oh, H., Slattery, M., Ma, L., Crofts, A., White, K. P., Mann, R. S. and Irvine, K. D. (2013). Genome-wide association of Yorkie with chromatin and chromatin-remodeling complexes. *Cell Rep* **3**, 309–318.

Ohlen, Von, T. and Hooper, J. E. (1997). Hedgehog signaling regulates transcription through Gli/Ci binding sites in the wingless enhancer. *Mech. Dev.* **68**, 149–156.

Ohlen, Von, T., Lessing, D., Nusse, R. and Hooper, J. E. (1997). Hedgehog signaling regulates transcription through cubitus interruptus, a sequence-specific DNA binding protein. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2404–2409.

Oosterveen, T., Kurdija, S., Alekseenko, Z., Uhde, C. W., Bergsland, M., Sandberg, M., Andersson, E., Dias, J. M., Muhr, J. and Ericson, J. (2012). Mechanistic Differences in the Transcriptional Interpretation of Local and Long-Range Shh Morphogen Signaling. *Dev. Cell* **23**, 1006–1019.

Orenic, T. V., Slusarski, D. C., Kroll, K. L. and Holmgren, R. A. (1990). Cloning and characterization of the segment polarity gene cubitus interruptus Dominant of Drosophila. *Genes & Development* **4**, 1053–1067.

Parker, D. S., White, M. A., Ramos, A. I., Cohen, B. A. and Barolo, S. (2011). The cis-Regulatory Logic of Hedgehog Gradient Responses: Key Roles for Gli Binding Affinity, Competition, and Cooperativity. *Sci Signal* **4**, ra38.

Parker, S. C. J., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., van Bueren, K. L., Chines, P. S., Narisu, N., Program, N. C. S., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor

human disease risk variants. *Proceedings of the National Academy of Sciences* **110**, 17921–17926.

Pavesi, G., Mereghetti, P., Mauri, G. and Pesole, G. (2004). Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.* **32**, W199–203.

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502.

Perrimon, N. (1994). The genetic basis of patterned baldness in Drosophila. *Cell* **76**, 781–784.

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

Pertea, M. and Salzberg, S. L. (2010). Between a chicken and a grape: estimating the number of human genes. *Genome Biology* **11**, 206.

Peterson, K. A., Nishi, Y., Ma, W., Vedenko, A., Shokri, L., Zhang, X., McFarlane, M., Baizabal, J.-M., Junker, J. P., van Oudenaarden, A., et al. (2012). Neural-specific Sox2 input and differential Gli-binding affinity provide context and positional information in Shh-directed neural patterning. *Genes & Development* **26**, 2802–2816.

Pevny, L. and Placzek, M. (2005). SOX genes and neural progenitor identity. *Curr. Opin. Neurobiol.* **15**, 7–13.

Piepenburg, O., Vorbrüggen, G. and Jäckle, H. (2000). Drosophila segment borders result from unilateral repression of hedgehog activity by wingless signaling. *Mol. Cell* **6**, 203–209.

Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23**, 4878–4884.

Ramos, A. I. and Barolo, S. (2013). Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **368**, 20130018.

Rebeiz, M. and Posakony, J. W. (2004). GenePalette: a universal software tool for genome sequence visualization and analysis. *Dev. Biol.* **271**, 431–438.

Rebeiz, M., Castro, B., Liu, F., Yue, F. and Posakony, J. W. (2012). Ancestral and conserved cis-regulatory architectures in developmental control genes. *Dev. Biol.*

362, 282–294.

Rebeiz, M., Reeves, N. L. and Posakony, J. W. (2002). SCORE: a computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. Site clustering over random expectation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9888–9893.

Ribes, V., Balaskas, N., Sasai, N., Cruz, C., Dessaud, E., Cayuso, J., Tozer, S., Yang, L. L., Novitch, B., Marti, E., et al. (2010). Distinct Sonic Hedgehog signaling dynamics specify floor plate and ventral neuronal progenitors in the vertebrate neural tube. *Genes & Development* **24**, 1186–1200.

Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P., et al. (2005). Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. *Genome Res.* **15**, 1–18.

Roberts, R. B., Hu, Y., Albertson, R. C. and Kocher, T. D. (2011). Craniofacial divergence and ongoing adaptation via the hedgehog pathway. *Proceedings of the National Academy of Sciences* **108**, 13194–13199.

Rodriguez, I. and Basler, K. (1997). Control of compartmental affinity boundaries by hedgehog. *Nature* **389**, 614–618.

Rubin, G. M. and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. *Science* **218**, 348–353.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797–803.

Sahai-Hernandez, P. and Nystul, T. G. (2013). A dynamic population of stromal cells contributes to the follicle stem cell niche in the Drosophila ovary. *Development* **140**, 4490–4498.

Santagati, F., Abe, K., Schmidt, V., Schmitt-John, T., Suzuki, M., Yamamura, K.-I. and Imai, K. (2003). Identification of Cis-regulatory elements in the mouse Pax9/Nkx2-9 genomic region: implication for evolutionary conserved synteny. *Genetics* **165**, 235–242.

Sasaki, H., Hui, C., Nakafuku, M. and Kondoh, H. (1997). A binding site for Gli proteins is essential for HNF-3beta floor plate enhancer activity in transgenics and can respond to Shh in vitro. *Development* **124**, 1313–1322.

Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T. and Ashe, H. L. (2013). Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. *Genes & Development* **27**, 1146–1158.

Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M. A. and Gemmell, N. (2013). Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PLoS ONE* **8**, e54710.

Scales, S. J. and de Sauvage, F. J. (2009). Mechanisms of Hedgehog pathway activation in cancer and implications for therapy. *Trends Pharmacol. Sci.* **30**, 303–312.

Schaaf, C. A., Misulovin, Z., Gause, M., Koenig, A., Gohara, D. W., Watson, A. and Dorsett, D. (2013). Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. *PLoS Genet.* **9**, e1003560.

Schachter, K. A. and Krauss, R. S. (2008). Murine models of holoprosencephaly. *Curr. Top. Dev. Biol.* **84**, 139–170.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A. and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biol* **7**, e13.

Sekulic, A., Migden, M. R., Oro, A. E., Dirix, L., Lewis, K. D., Hainsworth, J. D., Solomon, J. A., Yoo, S., Arron, S. T., Friedlander, P. A., et al. (2012). Efficacy and safety of vismodegib in advanced basal-cell carcinoma. *N. Engl. J. Med.* **366**, 2171–2179.

Shahi, M. H., Zazpe, I., Afzal, M., Sinha, S., Rebhun, R. B., Meléndez, B., Rey, J. A. and Castresana, J. S. (2015). Epigenetic regulation of human hedgehog interacting protein in glioma cell lines and primary tumor samples. *Tumour Biol.* **36**, 2383–2391.

Shaner, N. C., Steinbach, P. A. and Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905–909.

Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135–1145.

Stamatoyannopoulos, J. A. (2012). What does our genome encode? *Genome Res.* **22**, 1602–1611.

Stapleton, M., Liao, G., Brokstein, P., Hong, L., Carninci, P., Shiraki, T., Hayashizaki, Y., Champe, M., Pacleb, J., Wan, K., et al. (2002). The Drosophila gene collection: identification of putative full-length cDNAs for 70% of D. melanogaster genes. *Genome Res.* **12**, 1294–1300.

Suryamohan, K. and Halfon, M. S. (2015). Identifying transcriptional cis-regulatory modules in animal genomes. *Wiley Interdiscip Rev Dev Biol* **4**, 59–84.

Swanson, C. I., Evans, N. C. and Barolo, S. (2010a). Structural Rules and Complex Regulatory Circuitry Constrain Expression of a Notch- and EGFR-Regulated Eye

Enhancer. *Dev. Cell* **18**, 359–370.

Swanson, C. I., Evans, N. C. and Barolo, S. (2010b). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev. Cell* **18**, 359–370.

Swanson, C. I., Hinrichs, T., Johnson, L. A., Zhao, Y. and Barolo, S. (2008). A directional recombination cloning system for restriction- and ligation-free construction of GFP, DsRed, and lacZ transgenic Drosophila reporters. *Gene* **408**, 180–186.

Swanson, C. I., Schwimmer, D. B. and Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr. Biol.* **21**, 1186–1196.

Swoboda, P., Adler, H. T. and Thomas, J. H. (2000). The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. elegans. *Mol. Cell* **5**, 411–421.

Tabata, T. and Kornberg, T. B. (1994). Hedgehog is a signaling protein with a key role in patterning Drosophila imaginal discs. *Cell* **76**, 89–102.

Taher, L., Narlikar, L. and Ovcharenko, I. (2015). Identification and computational analysis of gene regulatory elements. *Cold Spring Harb Protoc* **2015**, pdb.top083642.

Takashima, S. and Murakami, R. (2001). Regulation of pattern formation in the Drosophila hindgut by wg, hh, dpp, and en. *Mech. Dev.* **101**, 79–90.

Tang, J. Y., Mackay-Wiggan, J. M., Aszterbaum, M., Yauch, R. L., Lindgren, J., Chang, K., Coppola, C., Chanana, A. M., Marji, J., Bickers, D. R., et al. (2012). Inhibiting the hedgehog pathway in patients with the basal-cell nevus syndrome. *N. Engl. J. Med.* **366**, 2180–2188.

Tanimoto, K., Liu, Q., Bungert, J. and Engel, J. D. (1999). Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. *Nature* **398**, 344–348.

Teglund, S. and Toftgård, R. (2010). Hedgehog beyond medulloblastoma and basal cell carcinoma. *Biochim. Biophys. Acta* **1805**, 181–208.

Tenzen, T., Allen, B. L., Cole, F., Kang, J.-S., Krauss, R. S. and McMahon, A. P. (2006). The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Dev. Cell* **10**, 647–656.

Thomas, S., Li, X.-Y., Sabo, P. J., Sandstrom, R., Thurman, R. E., Canfield, T. K., Giste, E., Fisher, W., Hammonds, A., Celniker, S. E., et al. (2011). Dynamic reprogramming of chromatin accessibility during Drosophila embryo development.

Genome Biology **12**, R43.

Timmer, J., Johnson, J. and Niswander, L. (2001). The use of in ovo electroporation for the rapid analysis of neural-specific murine enhancers. *Genesis* **29**, 123–132.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**, 1453–1465.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E., et al. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. *Genome Biology* **3**, RESEARCH0088.

Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S. E. and Rubin, G. M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome Biology* **8**, R145.

Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510.

Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y. and Kondoh, H. (2003). Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Dev. Cell* **4**, 509–519.

Vadasz, S., Marquez, J., Tulloch, M., Shylo, N. A. and García-Castro, M. I. (2013). Pax7 is regulated by cMyb during early neural crest development through a novel enhancer. *Development* **140**, 3691–3702.

van Steensel, B., Delrow, J. and Bussemaker, H. J. (2003). Genomewide analysis of Drosophila GAGA factor target genes reveals context-dependent DNA binding. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2580–2585.

Vieillard, J., Duteyrat, J.-L., Cortier, E. and Durand, B. (2015). Imaging cilia in Drosophila melanogaster. *Methods Cell Biol.* **127**, 279–302.

Villicaña, C., Cruz, G. and Zurita, M. (2014). The basal transcription machinery as a target for cancer therapy. *Cancer Cell Int.* **14**, 18.

Vokes, S. A., Ji, H., McCuine, S., Tenzen, T., Giles, S., Zhong, S., Longabaugh, W. J. R., Davidson, E. H., Wong, W. H. and McMahon, A. P. (2007). Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. *Development* **134**, 1977–1989.

Vokes, S. A., Ji, H., Wong, W. H. and McMahon, A. P. (2008). A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes & Development* **22**, 2651–2663.

Wang, H., Lei, Q., Oosterveen, T., Ericson, J. and Matise, M. P. (2011). Tcf/Lef repressors differentially regulate Shh-Gli target gene activation thresholds to generate progenitor patterning in the developing CNS. *Development* **138**, 3711–3721.

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688.

Wang, X., Zhao, Z., Muller, J., Iyu, A. and Khng, A. J. (2013). Targeted inactivation and identification of targets of the Gli2a transcription factor in the zebrafish. *Biology ….*

WATSON, J. D. and CRICK, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.

Weirauch, M. T. and Hughes, T. R. (2010). Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet.* **26**, 66–74.

Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., Najafabadi, H. S., Lambert, S. A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443.

Whalen, S., Truty, R. M. and Pollard, K. S. (2016). Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* **48**, 488–496.

White, M. A., Parker, D. S., Barolo, S. and Cohen, B. A. (2012). A model of spatially restricted transcription in opposing gradients of activators and repressors. *Mol. Syst. Biol.* **8**, 614.

Winklmayr, M., Schmid, C., Laner-Plamberger, S., Kaser, A., Aberger, F., Eichberger, T. and Frischauf, A.-M. (2010). Non-consensus GLI binding sites in Hedgehog target gene regulation. *BMC Mol. Biol.* **11**, 2.

Wunderlich, Z., Bragdon, M. D. J., Vincent, B. J., White, J. A., Estrada, J. and DePace, A. H. (2015). Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. *Cell Rep* **12**, 1740–1747.

Xie, J., Murone, M., Luoh, S. M., Ryan, A., Gu, Q., Zhang, C., Bonifas, J. M., Lam, C. W., Hynes, M., Goddard, A., et al. (1998). Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* **391**, 90–92.

Xie, L., Hoffmann, A. D., Burnicka-Turek, O., Friedland-Little, J. M., Zhang, K. and Moskowitz, I. P. (2012). Tbx5-hedgehog molecular networks are essential in the second heart field for atrial septation. *Dev. Cell* **23**, 280–291.

Yavari, A., Nagaraj, R., Owusu-Ansah, E., Folick, A., Ngo, K., Hillman, T., Call, G., Rohatgi, R., Scott, M. P. and Banerjee, U. (2010). Role of lipid metabolism in smoothened derepression in hedgehog signaling. *Dev. Cell* **19**, 54–65.

Yavatkar, A. S., Lin, Y., Ross, J., Fann, Y., Brody, T. and Odenwald, W. F. (2008). Rapid detection and curation of conserved DNA via enhanced -BLAT and EvoPrinterHD analysis. *BMC Genomics 2008 9:1* **9**, 1.

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O. and Stark, A. (2014). Enhancer--core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*.

Zhang, Y. and Kalderon, D. (2001). Hedgehog acts as a somatic stem cell factor in the Drosophila ovary. *Nature* **410**, 599–604.

Zhu, L. J., Christensen, R. G., Kazemian, M., Hull, C. J., Enuameh, M. S., Basciotta, M. D., Brasefield, J. A., Zhu, C., Asriyan, Y., Lapointe, D. S., et al. (2011). FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* **39**, D111–7.