Mechanisms of Gene Transcription Across the Hedgehog Signaling Gradient

by

Andrea Isabel Ramos Serrano

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
2014

Doctoral Committee:

Associate Professor Scott E. Barolo, Chair
Assistant Professor Benjamin Allen
Professor David Arnosti
Professor Kenneth M. Cadigan
Professor Andrzej A. Dlugosz

To my family for their unconditional love and support, and

to Elliott for turning my grad school journey into an adventure.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The combinatorial binding of transcription factors to enhancers controls unique patterns of gene expression to establish cell fates and body plans. Enhancers regulated by the Hedgehog (Hh) signaling morphogen integrate inputs from the bifunctional transcription factor Ci to activate or repress the expression of target genes in response to Hh signaling. Despite the essential role played by this pathway during organogenesis, it is not clear how Hh/Ci-regulated enhancers control different transcriptional outputs in different developmental contexts. Here, I examined the role of Ci binding sites in four Hh-regulated enhancers of *dpp*, *patched (ptc)*, *stripe* and *wingless.* Counterintuitively, high-affinity enhancers with consensus Ci binding motifs respond in more restricted domains, whereas low-affinity enhancers with Ci site variants respond more broadly to Hh signaling. To further study the relationship between Ci binding sites and enhancer function, I characterized the role of predicted binding motifs in newly identified enhancers in the *ptc* locus. This analysis doubled the number of known Hh/Ci-regulated enhancers in *Drosophila,* and revealed that the vast majority of these enhancers require sites that deviate from the optimal Ci site to respond to Hh in different contexts. To better understand how the Ci binding site composition of these enhancers produce specific transcriptional outputs *in vivo,* I measured the transcriptional output of isolated Ci site variants. Using synthetic Hh-regulated enhancers with defined inputs, I found that certain Ci binding motifs are regulated in a Hh/Ci-independent manner. These findings may explain previous

unexpected results, and suggest that some Ci sites may be under complex selective pressures to balance inputs from activator Ci, repressor Ci, and other transcription factors with overlapping binding preferences. Competition for shared or partially overlapping binding sites might provide tissue-specificity to Hh signaling.

CHAPTER 1

Introduction

How do complex organisms arise from single-celled embryos? How are different cell-types specified during development? These fundamental questions are not fully answered yet, although much progress has been made to improve our understanding of the molecular mechanisms that control basic developmental processes. This dissertation aims to better understand how distinct sequences of DNA encode the critical information that facilitates proper development.

1.1 Enhancer history

The genome contains all the information to create functional organisms. The detailed instructions that are required to produce optimal amounts of proteins at the right time and place are arranged into units or modules known as *cis*-regulatory elements (Wittkopp and Kalay, 2012). Enhancers and core promoters are among these regulatory non-coding sequences (Levine et al., 2014). For many years, a gene-centric (coding sequence) view of transcription predominated, and non-coding sequences were neglected to a secondary role. In the last ten years, the availability of many sequenced genomes and the development of ground-breaking high-throughput sequencing technologies have increased the general awareness that non-coding regions of the genome play critical roles in transcription, disease and morphological diversity (Levine, 2010;

Levo and Segal, 2014; Ong and Corces, 2011; Plank and Dean, 2014; Shlyueva et al., 2014; Smith and Shilatifard, 2014; Wittkopp and Kalay, 2012). Yet, the importance of non-coding sequences, in gene transcription— particularly the function of enhancer elements— has been known for almost 35 years (Banerji et al., 1981).

1.2 Enhancer structure and function

Enhancers integrate inputs from different cellular and developmental contexts to produce tissue-specific responses critical during tissue differentiation, proliferation and maintenance (Lagha et al., 2012). Enhancers work as platforms for combinatorial binding of transcription factors (TFs) that recognize short motifs in these sequences (Spitz and Furlong, 2012). Unique combinations of TF inputs provide enhancers the ability to precisely control when, where and how much of a gene is transcribed. Binding of TFs to enhancer sequences activates or represses transcription by recruiting or evicting the transcriptional machinery from the core-promoter (Levo and Segal, 2014).

Enhancers vary in size from several hundred base pairs to thousands of base pairs (Evans et al., 2012). Enhancers are sometimes described as “minimal enhancers” when describing shorter fragments or as “super-enhancers” when describing long segments of regulatory sequence (Hnisz et al., 2013; Lovén et al., 2013; Whyte et al., 2013). Originally, enhancers were defined as DNA fragments that augmented transcription independently of their orientation with respect of the promoter (Banerji et al., 1981).

Enhancers can be located within the vicinity of the gene they regulate, either upstream of the transcription start site (TSS) or in downstream intronic regions. Enhancers can also regulate transcription from far away (Krivega and

Dean, 2012). For example, the *dpp* gene contains enhancers that are located at least 20 kilobase pairs downstream of the promoter (Blackman et al., 1991). In other cases enhancers that regulate a particular gene are located in other loci (Lettice et al., 2003; Shlyueva et al., 2014).

Multiple enhancers can regulate the same gene. The *even-skipped* (*eve*) locus is a great example of a gene that contains multiple enhancers that are active in separate domains that combined produce the final endogenous *eve* pattern (Fujioka et al., 1999; Small et al., 1996). In other examples, separate enhancers have been shown to drive redundant expression patterns (Barolo, 2011; Hong et al., 2008). These "shadow" enhancers have been shown, in some cases, to be functional by conferring robustness during stress (Frankel et al., 2010; Perry et al., 2010).

Enhancers have been commonly identified by evolutionary conservation and/or clustering of predicted TF binding motifs (Berman et al., 2004; Goode et al., 2005; Pennacchio et al., 2006; Rebeiz, 2002). Testing the sufficiency to drive gene expression in reporter assays functionally validates these sequences (Wittkopp and Kalay, 2012). Although these methods have been successful in the past, they are by no means infallible because not all functional enhancers show evolutionary sequence conservation and not all highly conserved sequences display regulatory activity (Blow et al., 2010; Swanson et al., 2011; Vavouri and Lehner, 2009; Wittkopp et al., 2008).

More recently, enhancers have been identified genome-wide by TF or co-activator binding through ChIP-seq and related techniques (Arnold et al., 2013; Junion et al., 2012; Visel et al., 2009; Vokes et al., 2007; Yu et al., 2013; Zeitlinger et al., 2007). These methods have had success in identifying regulatory sequences, although many TF-bound regions do not appear to function as enhancers (Biggin, 2011; Li et al., 2011; Maston et al., 2012).

Other studies use biochemical signatures, including histone tail modifications, co-activator binding, and DNase accessibility to identify putative enhancers (Ong and Corces, 2011; Stamatoyannopoulos, 2012). Because of the large numbers of predicted enhancers, most of these sequences have yet to be functionally validated (Göke et al., 2011; Ng et al., 2013; Shen et al., 2012; Yu et al., 2013).

1.3 Signaling-regulated enhancers

A surprisingly small number of signaling pathways mediate cell-to-cell communication to establish unique patterns of gene expression throughout metazoan development (Barolo and Posakony, 2002; Perrimon et al., 2012; Pires-daSilva and Sommer, 2003). Despite the major differences between these signaling pathways, they follow similar signaling logics to transmit patterning information. Secreted signals (in most cases) bind cognate receptors to trigger signaling cascades that modify the function of dedicated TFs, which directly control the expression of target genes.

Signaling-regulated enhancers have unique architectures that include binding sites for the dedicated TFs and binding sites for additional signal-independent, tissue-specific or broadly expressed TFs (Spitz and Furlong, 2012). Many signal-regulated genes are turned on upon pathway activation, and actively turned off in the absence of signaling. This switch-like regulation of target genes is known as a transcriptional switch, and many signaling pathways use common regulatory strategies to tightly control gene expression (Barolo and Posakony, 2002; Ptashne, 2011).

The first common strategy is known as *activator insufficiency* (Barolo and Posakony, 2002). This occurs when signaling-regulated enhancers are unable to

activate transcription in cells with active signaling. Activator insufficiency is commonly seen in instances where minimal enhancers containing multimerized binding sites for dedicated TFs are insufficient to drive transcription in tissues where the cognate pathway is functional (Barolo, 2006; Ramos and Barolo, 2013). This transcriptional strategy prevents spurious target gene expression and provides tissue-specificity (Ramos and Barolo, 2013).

The second strategy also provides spatiotemporal specificity to gene expression. Signaling-regulated enhancers generally require extra inputs from selector genes or local activators, in addition to dedicated TFs inputs, to promote transcription (Mann and Carroll, 2002). This combinatorial regulation at the enhancer level is known as *cooperative activation* (Barolo and Posakony, 2002). Cooperativity among TFs at enhancers is an elegant and widely used strategy to tightly control gene transcription (Ptashne and Gann, 2002). The spatiotemporal overlap of distinct combinations of TFs are required to form complex patterns of gene expression.

Lastly, many signaling-regulated enhancers actively repress target gene transcription in the absence of signaling by using dedicated transcriptional repressors. This is known as *default repression* (Barolo and Posakony, 2002), This strategy prevents ectopic activation of target genes by local activator inputs.

Many signaling pathways used these conserved transcriptional strategies to tightly control the expression of key target genes during development. However, the precise molecular mechanisms that mediate these responses are not fully understood.

1.4 Hedgehog signaling directly regulates enhancers that contain Ci/Gli binding sites

Hedgehog (Hh) signaling is among the pathways that utilizes the transcriptional strategies that were previously discussed (Barolo and Posakony, 2002). Hh is a secreted morphogen that provides positional information primarily by forming a signaling gradient to differentially regulate gene expression at different signal intensities, which are interpreted by enhancers that contain binding sites for the TF Cubitus interruptus (Ci). In *Drosophila*, Hh-receiving cells modify Ci, a member of the Gli family of TFs that activates or represses transcription of key target genes (Hui and Angers, 2011). In the presence of the Hh signal, the activator isoform of Ci (Ci^{Act}) stimulates transcription of Hh target genes; but in the absence of signaling, a repressor isoform of the same protein (Ci^{Rep}) inhibits transcription of those genes (Figure 1.1). Ci recognizes enhancers that contain the same optimal consensus sequence as mammalian Gli factors, GACCACCCA—but, like many other TFs, it also binds sequences that deviate from this consensus site (Figure 1.2) (Hallikas et al., 2006; Müller and Basler, 2000; Ohlen and Hooper, 1997; Ohlen et al., 1997; Parker et al., 2011; Piepenburg et al., 2000). These deviant Ci binding sites are also known as non-consensus sites.

The DNA binding preference of Ci was obtained mainly from experiments studying the human Gli transcription factors, and since these factors are highly conserved through evolution, the results are readily applicable to fly Ci. The first seminal paper identified the consensus motif by isolating human DNA sequences that were bound by Gli (Kinzler and Vogelstein, 1990). Then, several years later, the crystal structure of a Gli-DNA complex was characterized (Pavletich and Pabo, 1993). These results provided the first evidence that the Gli family of transcription factors recognize distinct sequences in the genome, mostly

comprising a conserved 9-mer core. Later on, enhancers that contain these sequences were identified, and the direct regulation by Hh signaling was tested (Agren et al., 2004; Alexandre et al., 1996; Sasaki et al., 1997). More recently, a high-throughput approach was used to measure *in vitro* the binding preferences of Ci and the other human Gli factors (Hallikas et al., 2006). These results show that Ci and Glis recognize the consensus sequence more strongly, but these factors also bind to other degenerate sequences, although these interactions are weaker and potentially more transient (Figure 1.2) (Hallikas et al., 2006).

1.5 Known Hedgehog target enhancers

Most of the functionally characterized Hh/Ci-regulated enhancers in *Drosophila* respond to Hh signaling through non-consensus Ci binding sites (Ramos and Barolo, 2013; Winklmayr et al., 2010). The only known enhancer that contains consensus Ci sites is *ptc*, which encodes the Hh receptor (Chen and Struhl, 1996; Ingham et al., 1991). This enhancer contains a cluster of three highly conserved consensus Ci binding sites of optimal binding affinity (Alexandre et al., 1996; Parker et al., 2011).

Another example of a Hedgehog-regulated enhancer is *dpp*D, which is both activated and repressed by Ci in imaginal discs (Alexandre et al., 1996; Müller and Basler, 2000; Parker et al., 2011). The *dppD* enhancer is regulated by a cluster of Ci binding sites that are required for proper spatial patterning by Hh/Ci in the developing wing (Freeman, 2000; Müller and Basler, 2000; Parker et al., 2011). This enhancer drives expression of the long-range morphogen *decapentaplegic* (*dpp*), which encodes a BMP family member that controls wing growth and patterning (Wartlick et al., 2011). The *dppD* enhancer is not the only known element that controls the expression of *dpp* in imaginal discs. The *dppHO*

enhancer also drives expression in these tissues, but only during early larval stages (Hepker et al., 1999). Not surprisingly, this enhancer also requires non-consensus Ci sites to respond to Hh signaling.

The *knot* (*kn*) enhancer is also expressed in imaginal tissues in a pattern similar to the *ptc* enhancer (Alexandre et al., 1996; Hersh and Carroll, 2005). This gene, also known as *collier*, coordinates the activity of several signaling pathways including Hh, Dpp and EGF signaling to pattern the *Drosophila* wing (Crozatier et al., 2002). The *kn* enhancer relies on non-consensus binding sites to respond to Hh signaling (Hersh and Carroll, 2005). Other examples of Hh-regulated enhancers that drive gene expression in imaginal discs include two enhancers for the target gene *blistered,* which plays a role in cell differentiation (Nussbaumer et al., 2000).

Two Hh-regulated enhancers that drive expression during embryogenesis have been identified. These enhancers are *wg*1.0 and *sr*1.9, which also employ non-consensus Ci binding sites to drive precise expression patterns in the embryonic ectoderm. The *wg*1.0 enhancer responds to Hh via four non-consensus binding sites (Ohlen and Hooper, 1997; White et al., 2012) to control the expression of the *wingless* (*wg*) gene, which encodes a Wnt-family morphogen (Swarup and Verheyen, 2012). The *sr*1.9 enhancer relies on two non-consensus Ci binding sites to regulate the expression of *stripe* (*sr*), a gene required for muscle-pattern formation during embryogenesis (Frommer et al., 1996).

1.6 Developing *Drosophila* wing: a model to study enhancer function

Hh signaling patterns multiple tissues during development including the larval wing, which gives rise to the adult body wall appendage that is specialized

for flying. The developing larval wing is also known as the wing imaginal disc. In general, imaginal discs are sac-like structures that give rise to distinct adult cuticular structures (Morata, 2001).

The wing imaginal disc contains two separate cell lineages, which form the anterior and posterior compartments as a result of the activity of the Engrailed homeobox protein (Morata, 2001). The expression of Engrailed in the posterior compartment activates the expression of Hh, which triggers the Hh signaling cascade to pattern the developing wing. The posteriorly-produced Hh protein can move across the anterior/posterior (A/P) border to the anterior compartment to activate the expression of several Hh target genes like *dpp, knot and ptc* (Neto-Silva et al., 2009).

The wing imaginal disc has been widely used to study Hh signaling. In fact, much of our current knowledge about mammalian Hh signaling was directly influenced by findings in *Drosophila* (Hartl and Scott, 2014).

1.7 Aims and main findings

Despite the essential role that Hh signaling plays during development and adulthood, it is not clear how Hh/Ci-regulated enhancers control discrete transcriptional outputs at different developmental contexts. Therefore, in my thesis, I attempted to understand the transcriptional regulation of Hh target enhancers using the wing imaginal disc as model system. In Chapter 2, I studied previously known enhancers to better understand how distinct Ci binding sites integrate Hh inputs. In Chapter 3, I identified and characterized new Hh-responsive enhancers. Finally, in Chapter 4, I analyzed the intrinsic nature of different Ci site variants using synthetic Hh-regulated enhancers as a model to

ask questions about the relationship between sequence variants and *in vivo* transcriptional outputs. Together, the findings presented in this dissertation highlight the pivotal role of non-consensus Ci sites in the interpretation of the Hh signaling gradient, and provide a new toolkit of multiple novel Hh target enhancers that are useful to further understand the regulation of these regulatory sequences during development.

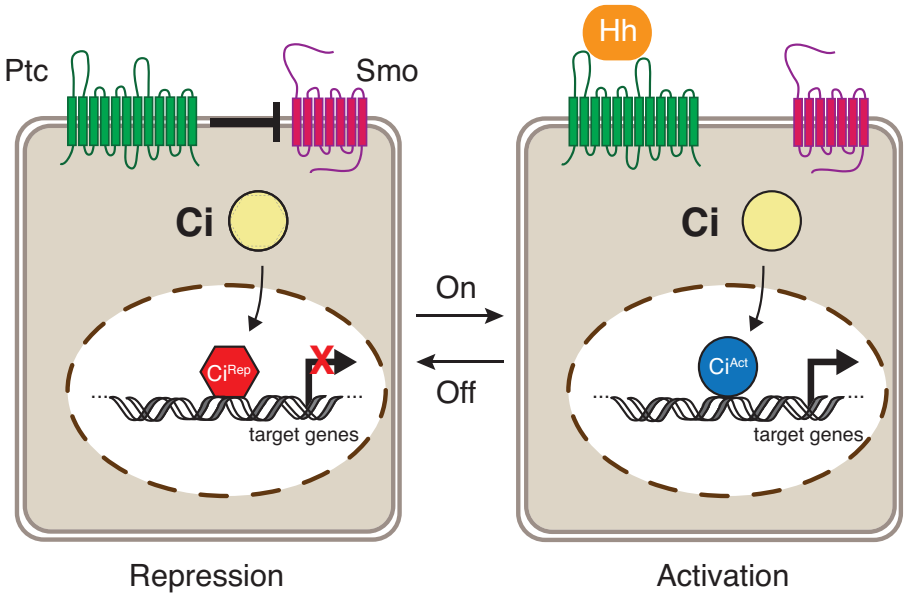


Figure 1.1 – The Hedgehog signaling pathway.

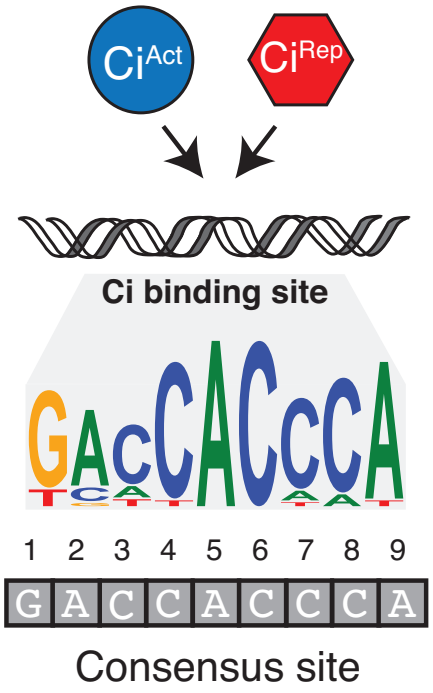


Figure 1.2 – The DNA binding preference of the transcription factor Cubitus interruptus (Ci).

1.8 References

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.

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

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

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

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

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 Biol. *5*, R61.

Biggin, M.D. (2011). Animal transcription networks as highly connected, quantitative continua. Dev. Cell *21*, 611–626.

Blackman, R.K., Sanicola, M., Raftery, L.A., Gillevet, T., and Gelbart, W.M. (1991). An extensive 3' *cis*-regulatory region directs the imaginal disk expression of *decapentaplegic*, a member of the TGF-beta family in *Drosophila*. Development (Cambridge, England) *111*, 657–666.

Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-

Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2010). ChIP-Seq identification of weakly conserved heart enhancers. Nat. Genet. *42*, 806–810.

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

Crozatier, M., Glise, B., and Vincent, A. (2002). Connecting Hh, Dpp and EGF signalling in patterning of the Drosophila wing; the pivotal role of collier/knot in the AP organiser. Development (Cambridge, England).

Evans, N.C., Swanson, C.I., and Barolo, S. (2012). Sparkling insights into enhancer structure, function, and evolution. Curr. Top. Dev. Biol. *98*, 97–120.

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.

Freeman, M. (2000). Feedback control of intercellular signalling in development. Nature *408*, 313–319.

Frommer, G., Vorbrüggen, G., Pasca, G., Jäckle, H., and Volk, T. (1996). Epidermal egr-like zinc finger protein of *Drosophila* participates in myotube guidance. The EMBO Journal *15*, 1642.

Fujioka, M., Emi-Sarker, Y., Yusibova, G.L., Goto, T., and Jaynes, J.B. (1999). Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. Development (Cambridge, England) *126*, 2527–2538.

Goode, D., Snell, P., Smith, D., Cooke, J., and Elgar, G. (2005). Highly conserved regulatory elements around the gene may contribute to the maintenance of conserved synteny across human chromosome 7q36.3. Genomics *86*, 172–181.

Göke, J., Jung, M., Behrens, S., Chavez, L., O'Keeffe, S., Timmermann, B., Lehrach, H., Adjaye, J., and Vingron, M. (2011). Combinatorial Binding in Human and Mouse Embryonic Stem Cells Identifies Conserved Enhancers Active in Early Embryonic Development. PLoS Comput. Biol. *7*, e1002304.

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.

Hartl, T.A., and Scott, M.P. (2014). Wing tips: The wing disc as a platform for studying Hedgehog signaling. Methods *68*, 199–206.

Hepker, J., Blackman, R.K., and Holmgren, R. (1999). Cubitus interruptus is necessary but not sufficient for direct activation of a wing-specific *decapentaplegic* enhancer. Development (Cambridge, England) *126*, 3669–3677.

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 (Cambridge, England) *132*, 1567–1577.

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.

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

Hui, C.-C., and Angers, S. (2011). Gli Proteins in Development and Disease. Annu. Rev. Cell Dev. Biol. *27*, 513–537.

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

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E.M. (2012). A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History. Cell *148*, 473–486.

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

Krivega, I., and Dean, A. (2012). Enhancer and promoter interactions—long distance calls. Curr. Opin. Genet. Dev. *22*, 79–85.

Lagha, M., Bothma, J.P., and Levine, M. (2012). Mechanisms of transcriptional precision in animal development. Trends Genet. *28*, 409–416.

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. Human Molecular Genetics *12*, 1725–1735.

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

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763.

Levo, M., and Segal, E. (2014). In pursuit of design principles of regulatory

sequences. Nat. Rev. Genet. *15*, 453–468.

Li, X.-Y., Thomas, S., Sabo, P.J., Eisen, M.B., Stamatoyannopoulos, J.A., and Biggin, M.D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding. Genome Biol. *12*, R34.

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. Cell *153*, 320–334.

Mann, R.S., and Carroll, S.B. (2002). Molecular mechanisms of selector gene function and evolution. Curr. Opin. Genet. Dev. *12*, 592–600.

Maston, G.A., Landt, S.G., Snyder, M., and Green, M.R. (2012). Characterization of Enhancer Function from Genome-Wide Analyses. Annual Review Genomics Human Genetics *13*, 29–57.

Morata, G. (2001). How Drosophila appendages develop. Nat Rev Mol Cell Biol *2*, 89–97.

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 (Cambridge, England) *127*, 2999–3007.

Neto-Silva, R.M., Wells, B.S., and Johnston, L.A. (2009). Mechanisms of Growth and Homeostasis in the Drosophila Wing. Annu. Rev. Cell Dev. Biol. *25*, 197–220.

Ng, J.-H., Kumar, V., Muratani, M., Kraus, P., Yeo, J.-C., Yaw, L.-P., Xue, K., Lufkin, T., Prabhakar, S., and Ng, H.-H. (2013). In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. Dev. Cell *24*, 324–333.

Nussbaumer, U., Halder, G., Groppe, J., and Affolter, M. (2000). Expression of the *blistered/DSRF* gene is controlled by different morphogens during *Drosophila* trachea and wing development. Mechanisms of Development *96*, 27–36.

Ohlen, Von, T., and Hooper, J.E. (1997). Hedgehog signaling regulates transcription through Gli/Ci binding sites in the *wingless* enhancer. Mechanisms of Development *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.

Ong, C.-T., and Corces, V.G. (2011). Enhancer function: new insights into the

regulation of tissue-specific gene expression. Nat. Rev. Genet. *12*, 283–293.

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. Science Signaling *4*, ra38–ra38.

Pavletich, N., and Pabo, C. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. Science *261*, 1701–1707.

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., Pitsouli, C., and Shilo, B.-Z. (2012). Signaling mechanisms controlling cell fate and embryonic patterning. Cold Spring Harbor Perspectives in Biology *4*, a005975.

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

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

Pires-daSilva, A., and Sommer, R.J. (2003). The evolution of signalling pathways in animal development. Nat. Rev. Genet.

Plank, J.L., and Dean, A. (2014). Enhancer Function: Mechanistic and Genome-Wide Insights Come Together. Molecular Cell *55*, 5–14.

Ptashne, M. (2011). Principles of a switch. Nat. Chem. Biol. *7*, 484–487.

Ptashne, M., and Gann, A. (2002). Genes & Signals (CSHL Press).

Ramos, A.I., and Barolo, S. (2013). Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. Philosophical Transactions of the Royal Society B: Biological Sciences *368*, 20130018.

Rebeiz, M. (2002). SCORE: A computational approach to the identification of *cis*-regulatory modules and target genes in whole-genome sequence data. Proceedings of the National Academy of Sciences *99*, 9888–9893.

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 (Cambridge, England) *124*, 1313–

1322.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the *cis*-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286.

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

Smith, E., and Shilatifard, A. (2014). Enhancer biology and enhanceropathies. Nat. Struct. Mol. Biol. *21*, 210–219.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

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

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

Swarup, S., and Verheyen, E.M. (2012). Wnt/Wingless Signaling in *Drosophila*. Cold Spring Harbor Perspectives in Biology *4*, a007930–a007930.

Vavouri, T., and Lehner, B. (2009). Conserved noncoding elements and the evolution of animal body plans. Bioessays *31*, 727–735.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

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 (Cambridge, England) *134*, 1977–1989.

Wartlick, O., Mumcu, P., Jülicher, F., and González-Gaitán, M. (2011). Understanding morphogenetic growth control -- lessons from flies. Nat Rev Mol Cell Biol *12*, 594–604.

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.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., and Lin, C.Y. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell.

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.

Wittkopp, P.J., and Kalay, G. (2012). *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. *13*, 59–69.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2008). Regulatory changes underlying expression differences within and between *Drosophila* species. Nat. Genet. *40*, 346–350.

Yu, Y., Chen, Y., Kim, B., Wang, H., Zhao, C., He, X., Liu, L., Liu, W., Wu, L.M.N., Mao, M., et al. (2013). Olig2 targets chromatin remodelers to enhancers to initiate oligodendrocyte differentiation. Cell *152*, 248–261.

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

CHAPTER 2

Low-affinity Transcription Factor Binding Sites Shape Morphogen Responses and Enhancer Evolution

2.1 Abstract

In the era of functional genomics, the role of transcription factor-DNA binding affinity is of increasing interest: for example, it has recently been proposed that low-affinity genomic binding events, though frequent, are functionally irrelevant. Here, we investigate the role of binding site affinity in the transcriptional interpretation of Hedgehog morphogen gradients. We noticed that enhancers of several Hedgehog-responsive *Drosophila* genes have low predicted affinity for Ci, the Gli-family transcription factor that transduces Hedgehog signaling in the fly. Contrary to our initial hypothesis, improving the affinity of Ci/Gli sites in enhancers of *dpp*, *wingless*, and *stripe*, by transplanting optimal sites from the *patched* gene, did not result in ectopic responses to Hedgehog signaling. However, we found that the enhancers require low-affinity binding sites for normal activation in regions of relatively low signaling. When Ci/Gli sites in these enhancers were altered to improve their binding affinity, we observed patterning defects in the transcriptional response that are consistent

with a switch from Ci-mediated activation to Ci-mediated repression. Synthetic transgenic reporters containing isolated Ci/Gli sites confirmed this finding in wings. We propose that the requirement for gene activation by Ci in regions of low-to-moderate Hedgehog signaling results in evolutionary pressure favoring weak binding sites in enhancers of certain Hedgehog target genes.

2.2 Introduction

Enhancers, also known as *cis*-regulatory elements or modules, are genomic DNA elements in command of the timing, location and levels of gene transcription. These transcriptional regulatory sequences integrate signaling and tissue-specific inputs through binding sites for a myriad of transcription factors (TFs) to specify spatiotemporal patterns of gene expression (Levine, 2010). Traditionally, enhancers have been identified functionally, in most cases by directly testing the sufficiency of stretches of DNA to drive gene expression in reporter assays. Nowadays, putative enhancers can be mined on a genome-wide basis by biochemical signatures, including histone tail modifications, co-activator binding, and DNase accessibility (Ong and Corces, 2011; Stamatoyannopoulos, 2012). Because hundreds or thousands of chromosomal sites cannot be easily tested for transcriptional activity, some genomic studies accept chromatin signatures associated with enhancer activity as self-validating evidence of enhancer function (*e.g.*, (Göke et al., 2011; Ng et al., 2013; Shen et al., 2012; Yu et al., 2013)). Another potential biochemical indicator of enhancers is TF or co-activator binding, as assessed on a genome-wide level by ChIP-seq and related

techniques (*e.g.*, (Arnold et al., 2013; Junion et al., 2012; Visel et al., 2009; Vokes et al., 2007; Yu et al., 2013; Zeitlinger et al., 2007)). These methods have had success in identifying regulatory sequences, although many TF-bound regions do not appear to function as enhancers (Biggin, 2011; Li et al., 2011; Maston et al., 2012). Other studies use DNA sequence signatures, mainly evolutionary conservation and/or clustering of predicted TF binding motifs, to screen genomes for enhancers (*e.g.*, (Berman et al., 2004; Goode et al., 2005; Pennacchio et al., 2006; Rebeiz, 2002)). These methods have also been successful, although again, they are by no means foolproof: for example, not all functional enhancers show evidence of evolutionary sequence conservation—even if their function is conserved—and conversely, not all highly conserved sequences display regulatory activity (*e.g.*, (Blow et al., 2010; Swanson et al., 2011; Vavouri and Lehner, 2009; Wittkopp et al., 2008)).

Enhancers are increasingly prominent in evolutionary thinking, as they have been shown to be the main agents of morphological diversity during evolution (Haag and Lenski, 2011; Peter and Davidson, 2011; Wittkopp et al., 2008). Changes that affect TF binding to enhancers have the potential to modify pleiotropic genes in a tissue-specific manner without compromising the survival of the organism. Sequence alterations such as deletions, insertions and nucleotide substitutions in enhancers have been shown to be responsible for morphological diversity (Wittkopp and Kalay, 2012). Because of the complex arrangement of TF binding motifs at enhancers, even tiny changes in regulatory

sequences can have significant effects in the transcriptional output by altering the TF input by modifying the binding affinity, binding site number, or changing the spacing between TFs, among many other possible scenarios (Levine, 2010).

Enhancers integrate inputs from different cellular and developmental contexts to produce tissue-specific responses critical during tissue differentiation, proliferation and maintenance. A small number of signaling pathways provide instructive inputs that are used in multiple developmental contexts (Barolo and Posakony, 2002; Pires-daSilva and Sommer, 2003). The highly conserved Hedgehog (Hh) signaling pathway is one of the key regulatory networks mediating cell communication during the development of most animals (Ingham et al., 2011). The Hh morphogen provides instructive positional information by establishing a signaling gradient that promotes different cell fates at different signal intensities, which are interpreted by enhancers, that contain binding sites for the effector of the pathway, the transcription factor Cubitus interruptus (Ci). In *Drosophila*, Hedgehog-receiving cells post-translationally modify Ci, a member of the Gli family of TFs, which activates or represses transcription of key target genes (Hui and Angers, 2011). In the presence of the Hedgehog signal, the activator isoform of Ci (Ci-Act) stimulates transcription of Hh target genes, but in the absence of signaling, a repressor isoform of the same protein (Ci-Rep) inhibits transcription of those genes. Ci recognizes enhancers that contain the same optimal consensus sequence as mammalian Gli factors, GACCACCCA—but, like many other TFs, it can also bind to sequences that deviate from this

consensus site (Hallikas et al., 2006; Ohlen et al., 1997). Thus, Ci activates or actively represses the transcription of Hh responsive genes depending on the state of signaling.

The Hh signaling gradient has been extensively characterized in the context of the developing wing of *Drosophila melanogaster* (Figure 3.1). In the third-instar larval wing imaginal disc, which gives rise to the adult wing, cells in the posterior compartment secrete the Hh morphogen: this signal is received and interpreted by cells of the anterior compartment that express Ci. The short-range Hh signal generates opposing reciprocal gradients of Ci-Act and Ci-Rep (Figure 3.1a) (Ashe and Briscoe, 2006; Parker et al., 2011). Anterior-compartment cells near the A/P compartment boundary receive maximal levels of Hh signaling and thus form Ci-Act exclusively, hence Hh/Ci regulated enhancers are active: these cells form what we will call the “activator zone.” Cells far from the source of Hh do not encounter the ligand and form Ci-Rep only, which represses target enhancers. These cells can be classified into the “repressor zone,” which comprises most of the anterior compartment of the wing. Between the activator and repressor zones there exists a region that receives moderate levels of Hh and produces both Ci-Act and Ci-Rep. We will refer to this region as the “mixed zone.” Here, the morphogen response becomes more complex, as Ci binding sites in Hh-responsive enhancers integrate competing inputs with opposing transcriptional functions. How *cis*-regulatory elements “decide” whether to be active or repressed by Ci in this zone is not well understood, but recent findings

(Parker et al., 2011; White et al., 2012), as well as the results presented here, show that the decision relies in part on the number and sequence of their Ci binding motifs. Bicoid and Dorsal, two morphogens that form signaling gradients during embryogenesis, also regulate key target genes in response to differences in binding site number and affinity (Driever et al., 1989; Jiang and Levine, 1993). However, because of the reciprocal gradients of Ci-Act and Ci-Rep, Hh/Ci-regulated enhancers interpret these differences unconventionally, and drive gene expression in unexpected domains across the gradient (Parker et al., 2011; White et al., 2012). A classic response is displayed by the Dorsal target gene *twist* which has a proximal enhancer with two low-affinity binding sites that drive limited gene expression in cells with high levels of the morphogen (Jiang and Levine, 1993). Improving the affinity of those sites resulted in higher levels of gene expression in a broader domain of the *Drosophila* embryo (Jiang and Levine, 1993). In the case of several Hh/Ci-regulated enhancers, the transcriptional response to changes in affinity is opposite to what is expected from the morphogen gradient model (these observations will be described in more detail below) (Wolpert, 1969).

A limited number of direct Hh/Ci target enhancers have been identified in *Drosophila* over the past two decades (Table 2.1). More recently, new elements have been characterized in vertebrates (Agren et al., 2004; Eichberger et al., 2008; Kasper et al., 2006; Oosterveen et al., 2012; Peterson et al., 2012; Sasaki et al., 1997). The highest standard for identification of a direct Ci/Gli target

enhancer consists of the following pieces of evidence: (a) the enhancer and parent gene are activated in a pattern consistent with Hh/Gli regulation; (b) the enhancer contains sites that are biochemically demonstrated to be bound by Gli proteins *in vitro* or *in vivo*; and (c) destruction of Gli sites diminishes the response of the enhancer and/or gene to Hh/Gli *in vivo*. Most, but not all, of the targets cited above meet that standard of evidence and can be regarded as confirmed direct Hh/Gli targets. Regardless of the species of origin, these enhancers respond to Hh signaling through variations on the same optimal Ci/Gli binding consensus (Hallikas et al., 2006; Hersh and Carroll, 2005; Müller and Basler, 2000; Ohlen and Hooper, 1997; Oosterveen et al., 2012; Peterson et al., 2012; Piepenburg et al., 2000; Sasaki et al., 1997).

Enhancers of the *Drosophila* genes *patched* (*ptc*) and *decapentaplegic* (*dpp*) were two of the earliest-identified direct Hh target sequences (Alexandre et al., 1996; Müller and Basler, 2000). The *ptc* enhancer is directly activated by Hh/Ci in larval imaginal discs via high-affinity Ci sites which perfectly match the optimal Gli binding consensus (Figure 3.1b-c and (Table 2.1) (Alexandre et al., 1996; Parker et al., 2011). By contrast, *dpp* is activated in the same tissues by an enhancer (designated here as *dpp*D) containing Ci sites of significantly lower affinity, with multiple mismatches to the optimal consensus (Figure 3.1b-c and Table 2.1) (Müller and Basler, 2000; Parker et al., 2011). In the wing imaginal disc, *ptc* is expressed in a narrow stripe of cells in the activator zone receiving maximal levels of Hh signaling, while *dpp* is expressed in a broader stripe in the

mixed zone, farther from the source of morphogen (Figure 3.1a) (Capdevila et al., 1994). These observations present a puzzle: why is a low-affinity Ci target gene like *dpp* activated more broadly across the Hh morphogen gradient than a high-affinity target gene like *ptc*? These results contrast with previous observations of the responsiveness of Bicoid and Dorsal target enhancers with low- and high-affinity sites (Driever et al., 1989; Jiang and Levine, 1993). Wolpert's French Flag model of positional information, which has been invoked (in modernized forms) to explain transcriptional responses to Hedgehog signaling (Ashe and Briscoe, 2006; Dessaud et al., 2008; Wolpert, 1969; 2011), would seem to predict that high-affinity targets should be more sensitive to signaling and as a result be expressed in a relatively broad domain across the gradient; by comparison, low-affinity target genes might be expected to have a higher response threshold and thus a more restricted expression domain. A model has been recently proposed to explain transcriptional responses to Hh/Gli in the vertebrate neural tube (Peterson et al., 2012). Yet the expression patterns of *ptc* and *dpp* in the wing suggest that different mechanisms may be at work. Furthermore, the effects of opposing activator/repressor transcription factor gradients, acting through the same *cis*-regulatory sites, have not been satisfactorily explained in any system.

We set out to explore how Ci binding site affinity affects the interpretation of Hh gradients in the developing *Drosophila* wing and embryonic ectoderm. Here, we present new data that corroborate recent findings (Parker et al., 2011; White et al., 2012) that some Hh-responsive enhancers require low-affinity

binding sites for normal activation in regions of relatively low signaling. Not only are these sites important, but their low affinity is equally important: when these non-consensus sites were upgraded to optimal Ci binding motifs, the result is gene expression patterning defects that are consistent with a switch from Ci-mediated activation to Ci-mediated repression (Parker et al., 2011). We present evidence consistent with a model in which selective pressure maintains non-consensus, low-affinity Ci binding sites in Hh-responsive enhancers, and propose that this is an evolutionary mechanism for maximizing Hh/Ci-mediated transcriptional activation in regions of Hh morphogen gradients where Ci-Act and Ci-Rep compete for enhancer binding.

2.3 Results

2.3.1 Many enhancers are regulated by non-consensus Ci binding sites

Most of the functionally characterized Hh/Ci-regulated enhancers in *Drosophila* respond to Hh signaling through non-consensus Ci binding sites (Table 2.1), some of which have been shown to exhibit relatively poor Ci binding affinity *in vitro* (Parker et al., 2011; White et al., 2012). The only known exception is *ptc*, which encodes the Hh receptor (Chen and Struhl, 1996; Ingham et al., 1991). *ptc* is unique among the known direct Hh/Ci target genes in two ways. First, *ptc* is regulated by a cluster of highly conserved consensus Ci binding sites of optimal binding affinity (Figure 3.1b-c and Figure 2.6a) (Alexandre et al., 1996; Parker et al., 2011). Second, unlike all other known Hh targets in the fly, which

respond to Hh in a tissue-restricted pattern, *ptc* is transcriptionally activated by Hh signaling universally (*i.e.*, in all tissues where Hh signaling occurs), as part of a negative feedback mechanism that regulates the range of signaling (Freeman, 2000).

Among the enhancers listed in is *dpp*D, which is both activated and repressed by Ci in imaginal discs (Alexandre et al., 1996; Müller and Basler, 2000; Parker et al., 2011). The *dppD* enhancer is regulated by a cluster of Ci binding sites which, though they deviate considerably from the optimal consensus and have low Ci binding affinity *in vitro*, are required for proper spatial patterning by Hh/Ci in the developing wing (Figure 3.1b-c and Figure 2.6b) (Freeman, 2000; Müller and Basler, 2000; Parker et al., 2011). This enhancer drives wing and leg expression of the long-range morphogen *decapentaplegic* (*dpp*), which encodes a BMP family member that controls wing growth and patterning (Wartlick et al., 2011). Two other Hh-regulated enhancers, *wg*1.0 and *sr*1.9, employ non-consensus Ci binding sites to drive precise expression patterns in the embryonic ectoderm (Table 2.1 and Figure 2.2). The *wg*1.0 enhancer responds to Hh via four non-consensus, low-affinity Ci binding sites (Figure 2.2b-c and Figure 2.6c) (Ohlen and Hooper, 1997; White et al., 2012) to control the expression of the *wingless* (*wg*) gene, which encodes a Wnt-family morphogen (Swarup and Verheyen, 2012). The *sr*1.9 enhancer relies on two non-consensus Ci binding sites (Table 2.1 and Figure 3.1b-c) to regulate the

expression of *stripe* (*sr*), a gene required for muscle-pattern formation during embryogenesis (Frommer et al., 1996).

Many of these functionally significant non-consensus Ci binding sites are strongly conserved throughout the evolution of the genus *Drosophila* (Figure 3.1b and Figure 2.6). This suggests the possibility of evolutionary pressures maintaining functional low-affinity Ci interactions with enhancers that interpret developmental Hh signaling gradients.

2.3.2 **Improving the binding affinity of Ci in the *dpp*D enhancer restricts expression to the activator zone**

We noticed that the *ptc* and *dpp*D enhancers, which are regulated by Ci binding sites of very different affinity, drive gene expression in distinct Hh signaling zones of the developing wing (Figure 3.1a and Figure 2.7) (Capdevila et al., 1994). The *ptc* enhancer, which contains optimal sites, responds to Hh only in the activator zone, while *dpp*D, with its non-consensus, low-affinity sites, responds to Hh in the mixed zone, farther from the source of morphogen (Figure 2.7) (Gallet et al., 2008) . To determine whether the low affinity of the Ci binding motifs in *dpp*D (which is evolutionarily conserved: see Figure 2.3 and Figure 2.6b) is important for responding Hh/Ci in the mixed zone, we converted the three low-affinity sites into high-affinity sites taken from the *ptc* enhancer (Parker et al., 2011). We observed that this “upgraded” enhancer, *dpp*D[Ci-ptc], which differs from the wild-type enhancer by only seven nucleotide positions, drives maximal gene expression in the activator zone instead of the mixed zone, similarly to *ptc*

(Figure 3.1d). To more precisely determine the transcriptional effect of changes in Ci binding affinity, we used a quantitative reporter gene assay (Parker et al., 2011) to measure GFP fluorescence across the dorsal portion of the wing pouch and normalized it to a *dpp*D[Ci-ptc]-DsRed reference transgene as an internal control for potential variations in age, fixation, and wing shape. We compared normalized GFP transgene expression levels driven by three versions of *dpp*D: wild type (wt); Ci-KO, in which the Ci sites were destroyed; and Ci-ptc, in which the binding affinity of the sites was improved by targeted base substitutions (Parker et al., 2011).

In accordance with previous work (Müller and Basler, 2000), we found that *dpp*D[Ci-KO] drove a broad expression pattern in the wing that differs from that of the wild-type enhancer in two respects: de-repression in anterior cells, and partial loss of activation in the mixed zone (Figure 3.1e) (Parker et al., 2011) . We used the expression of *dpp*D[Ci-KO] as a baseline, and measured the difference in fluorescence intensity between it and *dpp*D[wt] or *dpp*D[Ci-ptc] to determine the direct effect mediated by those three Ci binding sites at each position along the A/P axis of the wing disc (Figure 3.1f) (Parker et al., 2011). Although the *dpp*D[Ci-KO] expression pattern clearly shows reduced sensitivity to Ci activation and repression, its expression still suggests some regulation by Hh signaling: this is likely due to indirect regulation via a non-Ci input that is itself regulated by Hh/Ci (Müller and Basler, 2000), but it could also reflect input from uncharacterized Ci binding sites (Figure 2.3 and Figure 2.6b). Increased Ci

binding affinity provided stronger activation in the activator zone, as well as stronger repression in the repression zone, as expected—but unexpectedly, it also caused a switch from activation to repression in the mixed zone, where *dpp* (but not *ptc*) is normally activated (Figure 3.1f) (Parker et al., 2011).

2.3.3 **Low-affinity Ci binding sites diversify the Hh response**

As we proposed previously, the ectopic repression of *dpp*D[Ci-ptc] in the activator zone may be explained by two biophysical mechanisms (Parker et al., 2011). First, it is possible that Ci-Act and Ci-Rep have different binding preferences for distinct Ci motifs, such that Ci-Act prefers certain non-consensus sites while Ci-Rep prefers consensus sites. This scenario may seem unlikely because Ci-Act and Ci-Rep share the same DNA-binding domain (Orenic et al., 1990; Pavletich and Pabo, 1993), but it has not been directly ruled out. An alternative possibility is that strong cooperative interactions occur between Ci-Rep (but not Ci-Act) that result in lower-threshold levels for Ci-Rep (schematics of these models can be found elsewhere (Parker et al., 2011; Whitington et al., 2011)). Cooperative interactions are pervasive in gene regulation (Ptashne and Gann, 2002) and have been shown to lower threshold responses to other morphogens (Burz et al., 1998; Jiang and Levine, 1993). Fortuitously, these two models predict dramatically different transcriptional outputs for a modified *dpp*D enhancer with a single high-affinity site (*dpp*D[Ci1-ptc]) (Figure 3.1d). If the sequence motif itself dictates binding of Ci-Rep vs. Ci-Act, then the transcriptional profile of *dpp*D[Ci1-ptc] will be similar to *dpp*D[Ci-ptc], as both

enhancers contain only optimal consensus sites of identical sequence. On the other hand, if cooperative interactions between Ci-Rep are responsible for the restricted expression pattern of *dpp*D[Ci-ptc], then *dpp*D[Ci1-ptc] will behave more like *dpp*D[wt], because a single Ci site cannot mediate homomeric cooperative interactions.

We found that *dpp*D[Ci1-ptc] generates a broad stripe that is active in both the activator zone and the mixed zone (Figure 3.1f), which is consistent with the repressor-cooperativity model and inconsistent with the binding-preferences model (Parker et al., 2011). These results, and the deep evolutionary conservation of some of the low-affinity Ci sites in *dpp*D[wt], suggest the presence of selective evolutionary pressure maintaining low Ci occupancy at the *dpp*D enhancer. We speculate that *dpp* requires low-affinity Ci sites, which allow for activation by Hh/Ci but avoid invoking strong cooperative Ci repression in the mixed zone, in order to establish an organizing center in the middle of the wing for symmetric growth (Wartlick et al., 2011).

2.3.4 *wg* and *sr* require low-affinity Ci binding sites to respond optimally to Hh/Ci

To determine whether our observations regarding the effects of Ci binding site affinity are unique to *dpp*D or to the developing wing, we examined two other Hh/Ci-regulated enhancers, both of which respond to Hh signaling in the embryonic ectoderm but not the wing. We first tested a 1.0 kb enhancer of the *wingless* (*wg*) gene which drives Hh-responsive embryonic stripes anterior to

segmental stripes of Hh expression (Figure 2.2a,d) (Ohlen and Hooper, 1997; White et al., 2012). Four Ci binding sites in the *wg*1.0 enhancer (Table 2.1) have been reported to contribute to activation in Hh-responsive cells [43]. We improved the affinity of the three best-conserved Ci sites (*wg*1.0[3xCi-opt]) (Figure 2.2b-c and Figure 2.6c) (White et al., 2012). We observed that, rather than enhancing the transcriptional response to Hh, *wg*1.0[3xCi-opt] drives reduced expression levels in the embryonic ectoderm (Figure 2.2d) (White et al., 2012).

We also examined the *sr*1.9 enhancer, which is expressed in Hh-responsive embryonic stripes to the posterior of Hh-positive cells. This element has three non-consensus Ci binding motifs showing significant sequence conservation (Figure 2.2b-c), two of which had been previously identified (Table 2.1) (Piepenburg et al., 2000) . Destroying two of the predicted Ci sites has been reported to abolish the activity of this element (Piepenburg et al., 2000), but we found that improving the affinity of these sites, rather than augmenting gene expression, greatly reduced it (Figure 2.2e).

Taken together, these observations are consistent with the idea that *wg* and *sr*, like *dpp*, have Hh-responsive enhancers whose Ci occupancy is tuned at sub-maximal levels for optimal transcriptional activation in the proper zone of expression. We propose that this regulatory strategy stems from the dual nature of Ci as both an activator and a repressor, and the fact that these opposing activities are exerted through shared binding sites.

2.3.5 **Increasing the binding affinity of Ci does not induce significant ectopic expression**

We hypothesized that the relatively low binding affinity of these Ci-regulated enhancers might be important, not just for shaping responses to Hh morphogen gradients, but also for maintaining tissue-specificity of the Hh response. If this were the case, improving Ci binding affinity in these enhancers might be expected to sensitize them to Hh signaling, and thus might induce ectopic transcriptional responses to Hh/Ci outside of each gene's normal expression pattern. To address this point, we examined our high-affinity versions of the *dpp*D, *wg*1.0 and *sr*1.9 enhancers in tissues and at developmental stages where active Hh signaling occurs, but where the gene and enhancer do not normally respond to that signal.

The *dpp*D enhancer normally responds to Hh/Ci in the wing, leg, and antennal discs, but not in the embryonic ectoderm (where other genes such as *wg*, *sr*, and *ptc* respond to Hh signaling), and not in the morphogenetic furrow of the developing retina (where *dpp* expression is induced by Hh/Ci, but not through the *dpp*D enhancer) (Figure 2.8a) (Heberlein et al., 1995; 1993). We did not observe significant ectopic activity of *dpp*D[Ci-ptc] in Hh-responding cells of the embryonic ectoderm, nor in the morphogenetic furrow of the eye (Figure 2.8a). The only ectopic expression we observed was in part of the dorsal margin of the retina (Figure 2.8a), which might receive signaling from nearby Hh-positive photoreceptors (Lee et al., 1992), although this is not part of the normal *dpp* expression pattern (Blackman et al., 1991).

We next examined the expression of *wg*1.0[3xCi-opt] and *sr*1.9[2xCi-opt] in wing imaginal discs, where *wg* and *sr* do not normally respond to Hh/Ci, and found that improving Ci affinity did not activate ectopic transcriptional responses to Hh (Figure 2.8b,c). Consistently with these results, it was previously shown that adding consensus Ci sites to the wing-specific enhancer of *vestigial* (not a Hh/Ci target gene) fails to induce ectopic Hh responses (Halder et al., 1998; Kim et al., 1996). Our results demonstrate that the tissue-specific Hh responses of enhancers of *dpp*, *wg*, and *sr* cannot be explained by low binding affinity for Ci.

2.3.6 **Functionally significant non-consensus Ci sites display conservation of motif quality, even in the absence of strict sequence conservation**

Evolutionary sequence alignments of functional non-consensus Ci sites reveal multiple possible mechanisms by which the strength of Ci regulatory input into Hh-regulated enhancers may be maintained over evolutionary time, despite significant sequence turnover. Ci site 1 in the *dpp*D enhancer is perfectly preserved across 12 *Drosophila* species, but this is an exception: most non-consensus Ci motifs, even those for which regulatory function has been demonstrated, are not so strongly conserved, and many have undergone rapid and extensive sequence changes (Figure 2.3). For example, the sequence that comprises Ci site 2 in *dpp*D is conserved and aligned only in the three species most closely related to *D. melanogaster*; yet examination of nearby sequences reveals that the same motif (CGGGCGGTC) is found nearby in six additional *Drosophila* species, though it is not aligned with the *D. melanogaster* motif

(Figure 2.3a). In most cases, these motifs share no recognizable flanking sequence with the *D. melanogaster* site, so it cannot be determined whether this motif is an island of high conservation amid rapidly changing and expanding/contracting flanking sequence, or (probably less likely) the same motif has been independently acquired multiple times during *Drosophila* evolution.

Ci site 2 in the *wg*1.0 enhancer has a different evolutionary history: a predicted Ci site is present in all *Drosophila* species at this position, but the sequence itself is not highly conserved. Three different motifs, with similar predicted affinities, occur at this site (Figure 2.3b), suggesting that although the sequence of the site is evolving rapidly, the quality, or predicted affinity, of the site is constrained. A similar case of apparent quality constraint coupled with sequence flux occurs at Ci site 2 of the *sr*1.9 enhancer, where, for example, sequence changes in the *D. pseudoobscura* lineage diminish the quality of the site, while at the same time creating a new overlapping motif of very similar quality to the *D. melanogaster* motif (Figure 2.3c).

Ci site 1 in the *wg*1.0 enhancer seems to have undergone a triplet repeat expansion/contraction in the middle of the site, along with other changes (Figure 2.3b), with the result that some species, such as *D. melanogaster*, have a single moderate-affinity site, while other species have a weaker motif at that position but have gained additional nearby sites. These may be examples of compensatory changes that maintain levels of local Ci occupancy within a region of the enhancer. Another possible case of compensation occurs in the vicinity of Ci site

1 of *sr*1.9, which is poorly conserved—eight distinct sequence variants occur at this position across 12 species—yet in most cases, overall site quality appears to be well preserved, especially if a neighboring motif and its variations are taken into account. For example, *D. pseudoobscura* and *D. persimilis* have a motif in the position of site 1 that is considerably farther removed from the consensus than that in *D. melanogaster* (scoring 52.5 compared to 71.8), but have simultaneously acquired changes in a neighboring sequence that significantly improves its quality as a Ci motif (scoring 83.6 compared to 61.0 in *D. melanogaster*).

These are anecdotal cases, and the functional significance of these motifs in species other than *D. melanogaster* has not yet been tested. Nevertheless, careful sequence analysis appears to provide support for our speculation that the poor overall sequence conservation of many low-to-moderate-affinity Ci binding motifs may be deceptive: these local genomic regions may be under selective pressure to maintain Ci occupancies within a certain range, while at the same time allowing a great deal of change at the level of DNA sequence.

2.3.7 **Ci is insufficient to activate gene expression *in vivo***

To determine whether Ci binding sites, isolated from normal enhancer contexts, are capable of producing a transcriptional response to normal Hh signaling *in vivo*, we created a transgenic synthetic reporter in which three optimal Ci binding sites lie upstream of a minimal promoter driving GFP expression. This cluster of high-affinity sites (designated HHH) was not sufficient

to activate expression in regions of active Hh signaling in imaginal wing discs or in embryos (Figure 1.1). A similar construct bearing four high-affinity sites was previously shown to fail to respond to Hh in leg discs (Hepker et al., 1999). Our results exemplify a conserved transcriptional strategy known as "activator insufficiency," which is shared by multiple signaling pathways and is thought to be an evolutionary mechanism for preventing ectopic responses to highly pleiotropic signals such as Hh (Barolo and Posakony, 2002).

2.3.8 **Synthetic enhancers recapitulate endogenous expression patterns in the wing**

In order to study the functional properties of Ci binding sites outside the context of a complex enhancer sequence, we needed to circumvent the insufficiency of Ci sites alone (Figure 1.1a) to activate gene transcription *in vivo*. We borrowed a clever strategy (Jennings et al., 1999) that combines binding sites for the broadly expressed transcriptional activator Grainyhead (Grh) (Uv et al., 1997; 1994) with binding sites for Ci. Grh binding sites have been shown to be sufficient to activate gene transcription in the wing (Furriols and Bray, 2001). Using this approach, we were able to create a baseline level of transcription that allowed us to detect activating and repressive inputs from Ci sites, which can then be measured as changes in gene expression in Grh+Ci reporters, relative to a Grh-alone reporter. We generated four versions of synthetic enhancers with three Grh binding sites (GGG) upstream of three high-affinity sites (HHH), three low-affinity sites (LLL), one high-affinity site (H), and three mutant Ci sites (KO) to

preserve the spacing between the promoter and Grh (GGG) (Figure 1.1a). All of these transgenic constructs drove Hh/Ci-regulated stripes of different width and strength, with the exception of the 3xGrh-only construct (GGG), which, as expected, drove basal levels of expression throughout the wing disc (Figure 1.1a) (Uv et al., 1997). We quantitated, normalized, and compared GFP fluorescence data from these synthetic reporters as described for Figure 3.1, and observed that GGGHHH is expressed at high levels in the activator zone, GGGLLL is weakly expressed in the mixed zone, and GGGH is expressed at moderate levels in the activator and mixed zones (Figure 1.1b).

Next, we subtracted the Grh-only (GGG) expression levels from that of the Grh+Ci reporters to measure Ci-mediated activation and repression across the Hh gradient (Jennings et al., 1999; Parker et al., 2011). We found that GGGHHH, the synthetic counterpart of *dpp*D[Ci-ptc], is strongly activated by Ci in the activator zone but is repressed by Ci in the mixed zone, whereas the activity of GGGLLL peaks in the mixed zone and is weaker in the activator zone (Figure 1.1c). GGGH (analogous to *dpp*D[Ci1-ptc]) is activated by Ci in both the activator and mixed zones (Figure 1.1c) (Parker et al., 2011). The fact that these synthetic results are strikingly consistent with our observations with *dpp*D (compare Figure 1.1c and Figure 3.1e) indicates that the observed effects of Ci affinity on Hh responses in the wing are not dependent on a particular enhancer context, and demonstrates the utility of synthetic reporters for the quantitative analysis of Ci-regulated transcription in a simple and well-controlled sequence context.

The weak response of GGGLLL in the activator zone, compared to its expression in the mixed zone, is noteworthy (Figure 1.1c). In the case of the native *dpp*D enhancer, diminished expression in the activator zone has been attributed to repression by the homeodomain transcription factor Engrailed (En), which is expressed in a narrow stripe of anterior cells abutting the A/P boundary during late larval stages (Blair, 1992; Raftery et al., 1991). We analyzed the sequences of the synthetics to determine whether we had unknowingly introduced En binding sites (Figure 2.9e-h), and found a single predicted En site that overlaps with the first Ci site in GGGLLL and GGG (the site is destroyed in GGGHHH and GGGH). This En motif might be responsible for repressing GGGLLL in the activator zone. However, since we did not observe repression of GGG, which has the same En motif, in En-positive cells of the activator zone and the posterior compartment, and since GGGLLL was not repressed in the En-positive posterior compartment (Figure 1.1b), we conclude that these reporters are not directly repressed by En. The restricted activity of the low-affinity Ci binding sites in the mixed zone therefore seems to be encoded in the sequence of the Ci sites themselves. If true, this implies an as-yet-unknown mechanism for interpretation of the Hh gradient in the wing via Ci binding sites, but further research is required.

2.3.9 **Synthetic enhancers drive *ptc*-like expression in embryos**

To investigate whether the ability of these synthetic enhancers to respond to Hh/Ci is limited to imaginal tissues, we examined embryos at stages when Hh

signaling occurs (Figure 1.1d). The Grh, activator is expressed in the epidermis of mid- to late-stage embryos (Bray and Kafatos, 1991). At stage 11, our GGG synthetic construct, containing three Grh binding sites, reported low levels of Grh input in the dorsal ectoderm (Figure 1.1d). At that same stage, our synthetic Grh+Ci reporters (but not GGG) were activated in stripes to the anterior and posterior of each stripe of Hh-expressing cells (Figure 1.1d). This pattern differs from those of the natural Hh/Ci-activated enhancers of *wg* and *sr*, whose response is restricted to one side (the anterior and posterior, respectively) of each Hh-positive stripe (Ohlen and Hooper, 1997; Piepenburg et al., 2000); instead, it more closely resembles that of the *ptc* gene, which responds symmetrically to stripes of Hh signaling in embryos (Hooper and Scott, 1989; Nakano et al., 1989). GGGHHH drove high levels of expression in stripes that span both the dorsal and ventral sides of the embryo, while GGGLLL drove moderate levels of expression in dorsal stripes in cells that have the strongest Grh input (Figure 1.1d). GGGH drove activity in a similar pattern to that of GGGHHH, but at lower levels (Figure 1.1d). Contrary to what we observed by improving the affinity of *wg*1.0 and *sr*1.9, the high-affinity reporter GGGHHH was not more restricted in its expression than the low-affinity reporter GGGLLL, but instead was more strongly activated (Figure 2.2c,d compared to Figure 1.1d). Therefore, the strongly negative effect of high-affinity Ci sites on expression of the *wg* and *sr* enhancers may depend on the sequence context of those regulatory elements.

2.3.10 **Deep evolutionary conservation of putative homeodomain binding sites in *dpp*D**

The *dpp*D enhancer integrates inputs from other unknown factors besides Ci: this is demonstrated by the *dpp*D[Ci-KO] construct, which is active throughout the anterior compartment of the wing (Figure 3.1a) (Müller and Basler, 2000; Parker et al., 2011). To investigate the other inputs controlling *dpp*D, we examined the sequence conservation of this element across twelve *Drosophila* species (Figure 2.6b). Conserved TF binding motifs are considered likely to be functionally significant (Kheradpour et al., 2007; Meireles-Filho and Stark, 2009), although there are significant exceptions (Swanson et al., 2011; 2010; Wittkopp et al., 2008). The *dpp*D enhancer contains seven core homeodomain (HD) binding motifs (`TAAT`), of which six are perfectly conserved throughout the genus (Figure 2.6b). Overrepresentation of conserved HD binding sites has been also shown in some Hh-regulated enhancers in vertebrates (Oosterveen et al., 2012). All of the largest blocks of sequence conservation in *dpp*D include at least one HD core motif (*e.g.*, Figure 2.3a). Among these conserved potential HD binding sites is a previously identified site (designated as H^E in Figure 2.6b) which was shown to repress *dpp*D in posterior cells and has been proposed to mediate repression by En (Müller and Basler, 2000).

2.3.11 ***dpp*D integrates inputs from conserved putative HD binding sites**

To determine whether these potential HD binding sites contribute to the regulation of *dpp*D, we first tested the contribution of the previously identified En

binding site with a targeted mutation (*dpp*D[En-KO]). Consistent with prior findings (Müller and Basler, 2000), this mutation resulted in mild de-repression in the posterior compartment, where En is expressed (Figure 2.5a). We next mutated all seven core HD motifs in *dpp*D (7xHD-KO). This mutant enhancer drove a weak, incomplete wing stripe (Figure 2.5a). We quantitated the GFP fluorescence activated by these constructs across the wing to determine the regulatory contribution of putative HD binding sites. By comparing our measurements with wild type *dpp*D (Figure 2.5b), we found that the predicted En site, in addition to its known role in repression of *dpp*D in posterior cells (Müller and Basler, 2000), also contributed to *dpp*D activation in the anterior compartment, in cells lacking En (Figure 2.5b). We also found that at least one of the HD motifs is responsible for activating *dpp*D[En-KO] in posterior cells, as *dpp*D[7xHD-KO] was not active in that compartment. In anterior-compartment cells where *dpp* is normally expressed, we observed that the loss in activity in *dpp*D[7xHD-KO] was more severe than that caused by mutating three Ci binding sites (*dpp*D[Ci-KO]) (Figure 2.5b). The role of HD protein in activating *dpp*D contrasts with the repressive role of some HD binding proteins in Hh-regulated enhancers in the mouse neural tube (Oosterveen et al., 2012). However, we noticed that *dpp*D[7xHD-KO] was de-repressed in the eye retina (data not shown) where Hh signaling is active but *dpp*D[wt] is normally not expressed (Figure 2.8a).

Although the identities of the additional *dpp*D inputs remain a mystery, we speculate that the HD transcription factors Aristaless and Distal-less might be among these factors, based on their expression patterns in the wing and their known genetic relationship with *dpp* (Campbell and Tomlinson, 1998; Campbell et al., 1993). Our results are consistent with a model in which complex regulatory inputs from HD binding proteins (Han et al., 1989) act through highly conserved sites in the *dpp*D enhancer. They also highlight the critical role of low-affinity Ci binding sites, which cooperate with these positive and negative inputs to specify *dpp* expression in the proper segment of the Hh morphogen gradient (Figure 2.5c). Such a view contrasts sharply with characterizations of low-affinity TF-DNA interactions as functionally inconsequential (Fisher et al., 2012); to the contrary, certain types of regulatory circuits—especially those regulated by signaling pathways that use activator/repressor switch mechanisms, like Hh, Wnt, Notch, and others—may acquire and maintain low-affinity interactions to extract the maximum amount of information from developmental signaling events (Barolo and Posakony, 2002; Swanson et al., 2011; Tanay, 2006).

2.4 Discussion

In this study, we have presented *in vivo* evidence corroborating previous findings (Parker et al., 2011; White et al., 2012) that multiple tissue-specific enhancers require low-affinity Ci binding sites for optimal activation by Hh/Ci. Most of the Hh target enhancers identified up to this point in *Drosophila* and mouse are regulated by degenerate Ci/Gli binding sites of predicted lower affinity

(Table 2.1) (Hallikas et al., 2006). The prevalence of these non-consensus sites in Hh target enhancers across species demonstrates their importance in regulating the Hh response. The transcriptional relevance of low-affinity TF binding is not limited to Hh/Ci regulated enhancers. For instance, two phylogenetically conserved low-affinity binding sites in the mouse *Pax6* lens enhancer have been shown to be critical to promote gene expression at the right stage of development (Rowan et al., 2010).

We also provide a mechanistic explanation as to why these Hh/Ci-regulated elements require low-affinity sites to activate transcription in cells with moderate signaling levels. We showed that clusters of high-affinity sites mediate a restricted response in cells with high levels of Hh signaling, most likely as a result of cooperative interactions among Ci-Rep molecules in highly occupied Ci binding sites, while clusters of low-affinity sites mediate a broader response by having lower occupancy by Ci (Parker et al., 2011; White et al., 2012). Using synthetic enhancer reporters with high- or low-affinity Ci binding Ci sites, we confirmed this effect in the wing, but not in embryos. This tissue-specific discrepancy may imply a context-dependent function for some non-consensus Ci binding sites. As in the *Pax6* lens enhancer (Rowan et al., 2010), it is possible that some low-affinity binding sites are required specifically during earlier stages of development to interpret overall lower levels of Hh signaling (Balaskas et al., 2012; Nahmad and Lander, 2011).

Finally, we provided clues as to additional regulatory inputs into *dpp*D by showing a requirement for conserved consensus homeodomain binding sites. Cooperation between Glis and HD proteins has been recently shown in the mouse neural tube (Oosterveen et al., 2012). In this case, HD proteins are critical to repress Hh-regulated neural tube enhancers, while in *dpp*D they are critical to activate gene expression.

The limited number of known, experimentally confirmed, direct Hh/Gli target enhancers may reflect the widespread, practical tendency to search for consensus or near-consensus motifs, and to focus on the highest peaks of TF-DNA binding, when hunting for *cis*-regulatory sequences. From a biochemical standpoint—for example, when mining ChIP-seq data—low-affinity DNA-binding interactions are troublesome because they are much more common, by definition, than the top 1% of peaks. It is important to note that we do not mean to strictly equate ChIP peak height with TF binding affinity, nor to equate *in vitro* binding or *in silico* "motif quality" with *in vivo* TF occupancy, though these properties may often be roughly correlated. Separating the weak but functional binding events from weak and non-functional binding events is extremely challenging, and some have proposed that low-affinity genome-binding interactions can be categorically ignored (Fisher et al., 2012; Stamatoyannopoulos, 2012). This certainly simplifies the problem from a computational perspective, but the findings discussed here and elsewhere (Jaeger et al., 2010; Rowan et al., 2010) suggest a risk of discarding functional

sequences. Similar challenges confront *in silico* genomic screens to identify clusters of predicted TF binding sites: these necessarily filter out binding events of low predicted affinity, because there are many more predicted low-affinity binding motifs than consensus high-affinity motifs in any given sequence (Michelson, 2002). Binding site predictions have been supported by taking evolutionary sequence conservation into account (Hallikas et al., 2006; Vokes et al., 2007), but this risks filtering out true positives: as shown in our Ci motif alignments, lower-affinity binding sites seem to be less constrained with respect to sequence variation, even in cases when the presence of the site itself is highly conserved. This is presumably because, for each non-consensus binding motif, there are multiple alternative sequences with similar affinity and thus equivalent functionality. Importantly, this type of degenerate motif conservation is easily missed: for example, some of the well-conserved Ci motifs described here are not properly aligned in the UCSC Genome Browser, because they do not constitute contiguous blocks of perfect sequence identity. To avoid these pitfalls, it is important to use phylofootprinting approaches that account for these alignment flaws, such as the one described in (Kheradpour et al., 2007). In contrast to most of the low-affinity binding sites discussed here, optimal-affinity Ci motifs in the *ptc* enhancer have been preserved throughout the evolution of the genus *Drosophila*, and perhaps much farther: `GACCACCCA` motifs occur in promoter-proximal regions of multiple vertebrate orthologs of *ptc* (Agren et al., 2004; Vokes et al., 2007) (additional data not shown).

Evolutionary enhancer sequence alignments, along with limited experimental data, also suggest that, although many predicted low-affinity sites are poorly conserved, overall TF occupancy on an enhancer may be maintained despite significant sequence turnover. This may occur either through the rapid gain and loss of individual sites, or through the maintenance of relatively weak binding affinity at a site that is unstable at the level of DNA sequence (Swanson et al., 2011; Tuch et al., 2008). While this last idea requires further direct testing, it is consistent with the fact that Gli sites of moderate predicted affinity have many sequence variants of similar quality, while the highest-affinity motifs have far fewer alternatives of similar quality. In other words, there are many more ways to be a weak binding site than a strong site. For example, among all possible 9-mer sequences, there are 654 motifs with Ci matrix similarity scores between 70 and 75 (inclusive), but only 12 motifs with scores between 90 and 95, and one motif with a score above 95. Therefore, weaker binding sites, and the enhancers containing them, have a far greater volume of sequence space in which to roam without strongly impacting transcriptional output [22]. A thermodynamics-based simulation of enhancer evolution has shown that there is a greater number of fit solutions using weak TF sites than using high-affinity sites for a given gene expression problem (He et al., 2012).

Equally consistent with our view of TF binding site evolution is the fact that it is much easier (that is, more likely) to create a low-affinity, non-consensus binding motif with a single mutation than a high-affinity consensus motif. An

enhancer-sized DNA sequence can acquire a weak Gli motif with single nucleotide substitutions at any of a large number of positions, as demonstrated by our simulations (Figure 2.10). These arguments may help to explain why sequence conservation is not a foolproof test of the functional relevance of non-consensus TF binding sites.

While we cannot offer a simple answer to the technical challenges facing those who hunt enhancers, the findings described in this report lead us to conclude that low-affinity TF-DNA interactions, mediated by non-consensus and often poorly conserved sequence motifs, play important and widespread roles in developmental patterning and *cis*-regulatory evolution, and therefore cannot be safely ignored.

2.5 Materials and methods

Ci binding site prediction, scoring, and ranking

A mono-nucleotide distribution matrix for Ci binding sites, derived from competitive DNA binding assays (Hallikas et al., 2006) was downloaded from the Genomatix Software Suite (www.genomatix.de; Genomatix, Germany). Matrix similarity scores (Quandt et al., 1995) were calculated using data from the first nine nucleotide positions of the Ci matrix, which contain the majority of the information content. The matrix similarity score plots in Figure 3.1c and Figure 2.2c were generated with Apple Numbers and modified with Adobe Illustrator. Ci site rankings are determined by sorting all possible 9-mers in order of matrix

similarity score, such that the optimal motif (GACCACCCA), with a score of 100, has a rank of 1. 9-mers with a lower matrix score than their reverse-complement sequences, such as TGGGTGGTC, are removed from the ranking so that each high-scoring site is included only once.

DNA cloning and mutagenesis

Wild type *ptc*, *dpp*D, *sr*1.9 and *wg*1.0 enhancers were amplified by standard PCR from w^{1118} genomic DNA. Enhancer constructs were sub-cloned into the pENTR/D-TOPO plasmid (Invitrogen) by TOPO cloning. Enhancers were subsequently cloned into the pHPdesteGFP transgenesis vector (Boy et al., 2010) by LR Cloning (Invitrogen), or into the pEAB transgenesis vector (N.C. Evans and S.B., unpublished work) by traditional cloning methods. Targeted binding site mutations were created by overlap extension PCR (Swanson et al., 2010). Synthetic Hh-responsive enhancers were generated by assembly PCR (Swanson et al., 2010). See Figure 2.9 for full sequences of wild type and mutated enhancers investigated in this study.

***Drosophila* transgenesis**

Site-directed transformation by embryo injection was performed as previously described (Bischof et al., 2007). Reporter transgenes were integrated into a *phi*C31 landing site at genomic position 86FB.

Immunohistochemistry and microscopy

Embryos were fixed and stained using standard methods as previously described (White et al., 2012). Third-instar wing imaginal discs were dissected and fixed as described (Parker et al., 2011). Confocal images were captured on an Olympus FluoView 500 Laser Scanning Confocal Microscope mounted on an Olympus IX-71 inverted microscope. Samples to be directly compared were fixed, prepared, and imaged under identical confocal microscopy conditions and settings. The primary antibodies used included rabbit anti-EGFP (Invitrogen), diluted 1:100, and mouse anti-En (Developmental Studies Hybridoma Bank), diluted 1:50. Embryos were staged as described (Campos-Ortega and Hartenstein, 1985).

Quantitation of transgenic reporter expression data

Wing confocal images were collected and quantified as previously described (Parker et al., 2011). The Matlab program *Icarus* (E. Ortiz-Soto, A.I.R., and S.B., manuscript in preparation) was used to process and plot wing imaginal disc fluorescence data.

Evolutionary sequence alignments

Alignments of enhancer-orthologous sequences from 12 sequenced *Drosophila* genomes were obtained from the UCSC Genome Browser (genome.ucsc.edu), except for the *dpp*D enhancer, for which the UCSC alignment was incomplete; this alignment was performed with Clustal Omega

(www.ebi.ac.uk/Tools/msa/clustalo), using sequences identified with the EvoPrinter HD online tool (Yavatkar et al., 2008). Predicted binding motifs were identified with the GenePalette program (Rebeiz and Posakony, 2004); alignment graphics were then modified with Adobe Illustrator.

2.6 Acknowledgements

We are particularly grateful to Elliott Ortiz-Soto, Katherine Gurdziel, Dave Parker, Michael White and Barak Cohen for their exceptional contributions to this project. We thank Ingrid Lohmann for generously providing the pHPdesteGFP destination vector, and Niki Evans for making the pEAB transgenic vector. We thank Rachna Pannu for her assistance with the *wg* enhancer project, and David Lorberbaum and Lisa Johnson for their comments on the manuscript. This work was supported by the Cellular and Molecular Biology Training Grant (NIH T32-GM007315) and a Center for Organogenesis predoctoral fellowship (NIH T32-HD007505) to A.I.R. and by NIH grant GM076509 and NSF grant MCB-1157800 to S.B.

The text and figures presented in this chapter were previously published: Ramos, A.I., and Barolo, S. (2013). Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. Philosophical Transactions of the Royal Society B: Biological Sciences *368*, 20130018.

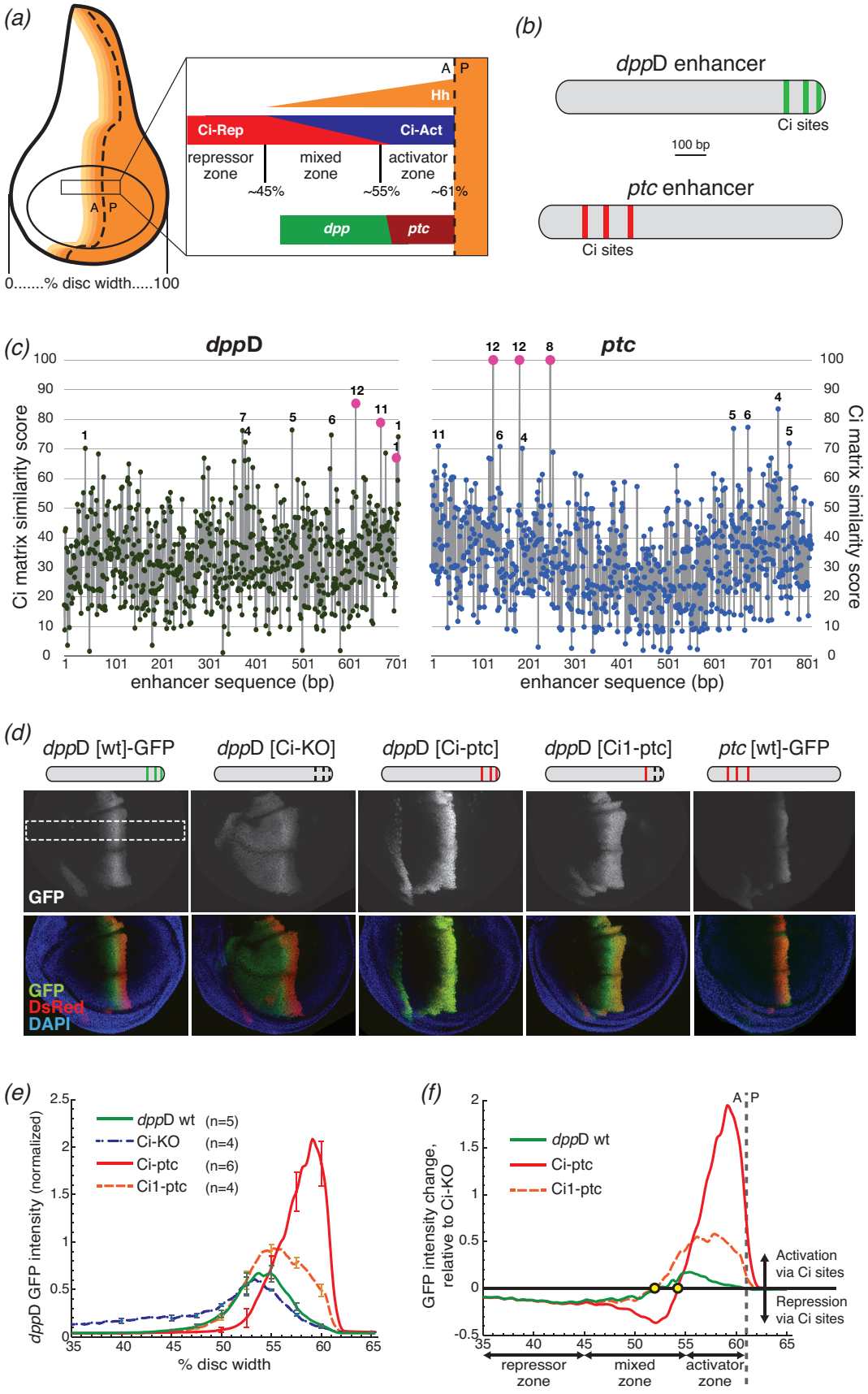

(a)
A
P
Hh
Ci-Rep
Ci-Act
repressor zone
mixed zone
activator zone
~45%
~55%
~61%
dpp
ptc
0.....% disc width.....100
(b)
dppD enhancer
Ci sites
100 bp
ptc enhancer
Ci sites
(c)
dppD
ptc
Ci matrix similarity score
enhancer sequence (bp)
(d)
dppD [wt]-GFP
dppD [Ci-KO]
dppD [Ci-ptc]
dppD [Ci1-ptc]
ptc [wt]-GFP
GFP
GFP
DsRed
DAPI
(e)
dppD wt (n=5)
Ci-KO (n=4)
Ci-ptc (n=6)
Ci1-ptc (n=4)
dppD GFP intensity (normalized)
% disc width
(f)
dppD wt
Ci-ptc
Ci1-ptc
GFP intensity change, relative to Ci-KO
Activation via Ci sites
Repression via Ci sites
repressor zone
mixed zone
activator zone

Figure 2.1 – The *dpp*D enhancer requires conserved low-affinity Ci/Gli sites to respond optimally to Hh and Ci in the developing wing.

(a) Diagram of the *Drosophila* third-instar wing imaginal disc, showing the distribution of the Hedgehog signaling gradient across the anterior compartment.The dashed line indicates the anterior-posterior (A/P) boundary separating posterior cells, which secrete Hh, from anterior cells, which express the Ci transcription factor. Magnification of a segment of the wing pouch across the compartment boundary shows distinct zones (repressor, mixed and activator) based on their distance from the source of Hh morphogen. The Hh target genes *dpp* and *ptc* respond differently to the gradient; *dpp* is expressed maximally in the mixed zone, while *ptc* expression is restricted to the activator zone. *(b)* Ci binding motifs in the *dpp*D and *ptc* enhancers. *(c)* Estimated Ci binding affinity and evolutionary conservation across the *dpp*D and *ptc* enhancers of *D. melanogaster*. Ci matrix similarity score (see Methods) was plotted for every 9-mer. Known or proposed Ci sites (see Table 2.1) are shown as larger dots. For each 9-mer with a score ≥70, numerals indicate the number of *Drosophila* species (out of 12) in which that sequence is present at or near the orthologous position. *(d)* Top, diagrams of *dpp*D enhancer constructs, with defined Ci binding sites as vertical bars (broken bars indicate mutated sites). Middle and bottom, confocal images of third-instar larval wing imaginal discs, showing GFP expression driven by *dpp*D-GFP or *ptc*-GFP reporter transgenes. Red fluorescence is driven by a *dpp*D[Ci-ptc]-DsRed transgene used for GFP fluorescence normalization and positional reference. In *dpp*D[Ci-KO]-GFP, three Ci sites were destroyed by targeted mutation; in *dpp*D[Ci-ptc]-GFP, three Ci sites were converted to optimal motifs taken from *ptc*; in *dpp*D[Ci1-ptc]-GFP, the 5' Ci site was optimized while site 2 and 3 were destroyed. The white dashed rectangle indicates the section of the dorsal wing pouch that is measured in the following panels. *(e)* Normalized GFP fluorescence data collected from the dorsal section of the wing pouch. Error bars indicate one standard deviation. *(f)* Net effect of wild type or high-affinity Ci sites on *dpp*D expression (calculated as the normalized transgene expression of *dpp*D[wt] or *dpp*D[Ci-ptc] minus normalized *dpp*D[Ci-KO] expression). Circles indicate the positions on the A/P axis at which Ci input switches from net activation to net repression for each enhancer.

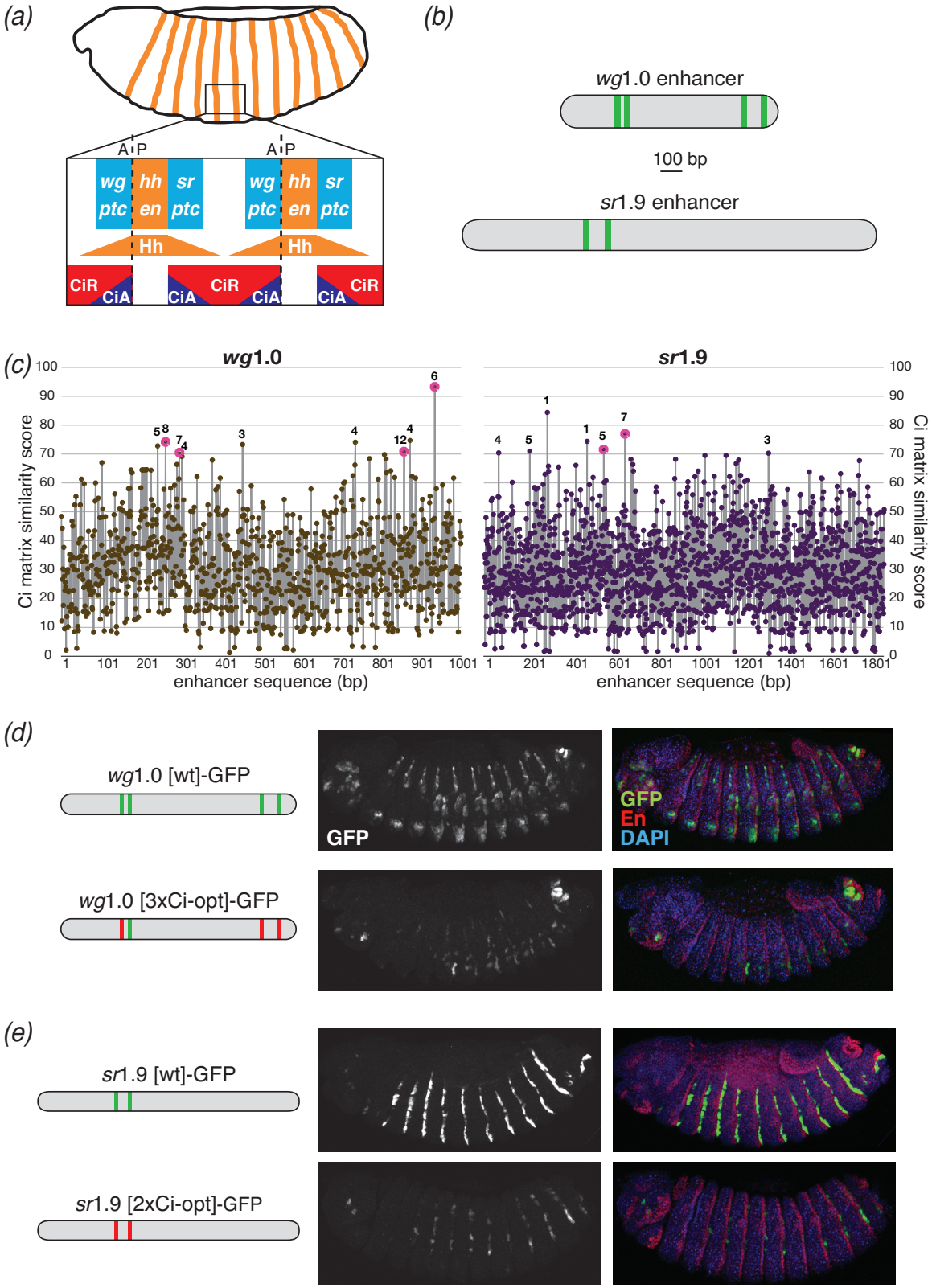
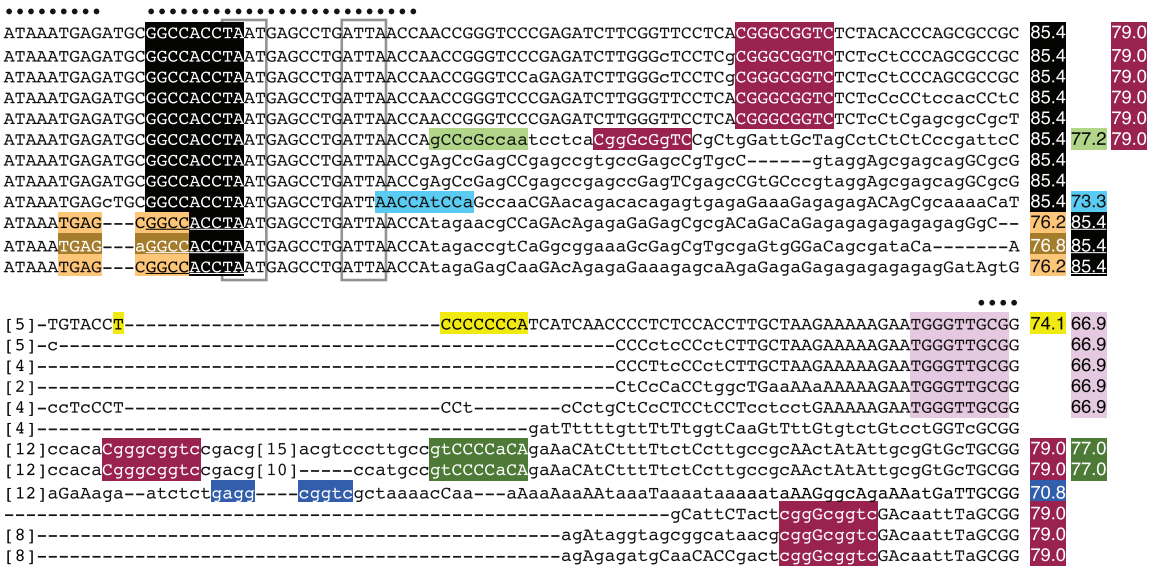

(a)
A P
A P
wg ptc
hh en
sr ptc
wg ptc
hh en
sr ptc
Hh
Hh
CiR
CiA
CiR
CiA
CiR
CiA
(b)
wg1.0 enhancer
100 bp
sr1.9 enhancer
(c)
wg1.0
sr1.9
Ci matrix similarity score
enhancer sequence (bp)
(d)
wg1.0 [wt]-GFP
wg1.0 [3xCi-opt]-GFP
GFP
GFP
En
DAPI
(e)
sr1.9 [wt]-GFP
sr1.9 [2xCi-opt]-GFP

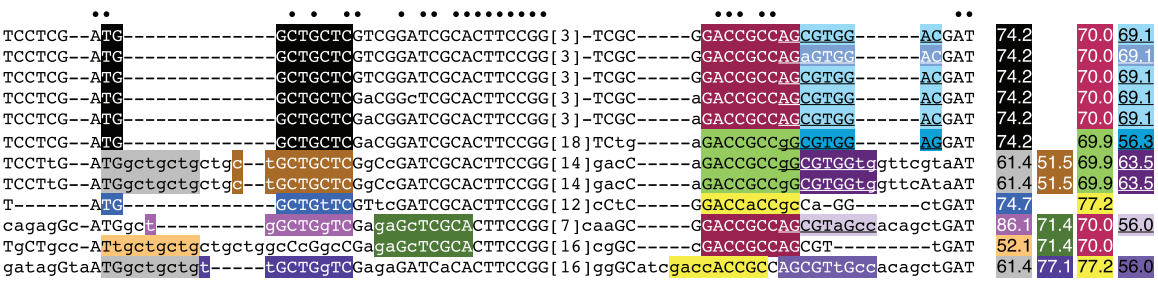
Figure 2.2 – Optimizing the Ci binding affinity in the *wingless* and *stripe* embryonic enhancers results in reduced levels of gene expression.

(a) Diagram of a stage 14 *Drosophila* embryo. A closer view of parasegments 5 and 6 shows bidirectional Hh signaling gradients which modulate the transcriptional activity of Ci. Known Hh/Ci target genes respond symmetrically (*patched* (Hooper and Scott, 1989; Nakano et al., 1989)) or asymmetrically (*e.g.*, *wingless (wg)* (Baker, 1988) and *stripe (sr)* (Piepenburg et al., 2000)) to the Hh signal. *(b)* Diagrams of the *wg*1.0[wt] and *sr*1.9[wt] enhancers showing Ci binding motifs as vertical lines. *(c)* Estimated Ci binding affinity across the *wg*1.0 and *sr*1.9 enhancers of *D. melanogaster*. Ci matrix similarity scores and conservation data are indicated as in Figure 1c. In *(d)* and *(e)*, diagrams on the left show *wg*1.0 and *sr*1.9D enhancer constructs; images of transgenic embryos show GFP alone (middle) and merged GFP, En (which marks Hh-producing cells), and DAPI nuclear stain (right). *(d)* Confocal images of stage 14 transgenic embryos carrying *wg*1.0[wt] and *wg*1.0[3xCi-opt], in which three Ci sites have been converted to optimal Ci binding motifs, driving GFP. *(e)* Confocal images of stage 14 transgenic embryos carrying *sr*1.9[wt] and *sr*1.9[2xCi-opt], in which two Ci sites have been converted to optimal Ci binding motifs, driving GFP. A, anterior; P, posterior.

*(a) dpp*D enhancer, Ci sites 1, 2, and 3



*(b) wg*1.0 enhancer, Ci sites 1 and 2



*(c) sr*1.9 enhancer, Ci sites 1 and 2

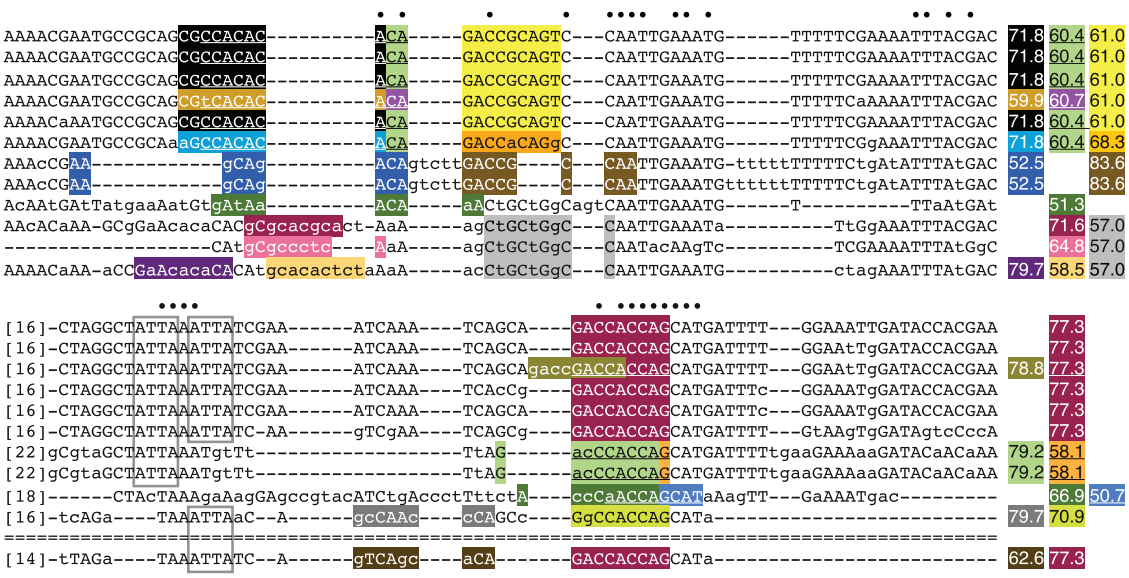


Figure 2.3 – Evolutionary conservation of predicted binding affinity, but not of sequence identity, at many low-to-moderate-affinity Ci binding sites.

Twelve-species *Drosophila* sequence alignments are shown for selected regions of the *dpp*D, *wg*1.0, and *sr*1.9 enhancers. Selected 9-mers are shaded, and Ci matrix similarity scores for those motifs are shown to the right. Sequences are from the following Drosophila species, from top to bottom: *D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, D. persimilis, D. willistoni, D. virilis, D. mojavensis*, and *D. grimshawi*. Dashes indicate gaps; double-dashes (=) indicate a lack of alignable sequence. Bracketed numbers indicate the number of bases deleted at that position to conserve space. Conserved homeodomain binding motifs are in gray boxes.

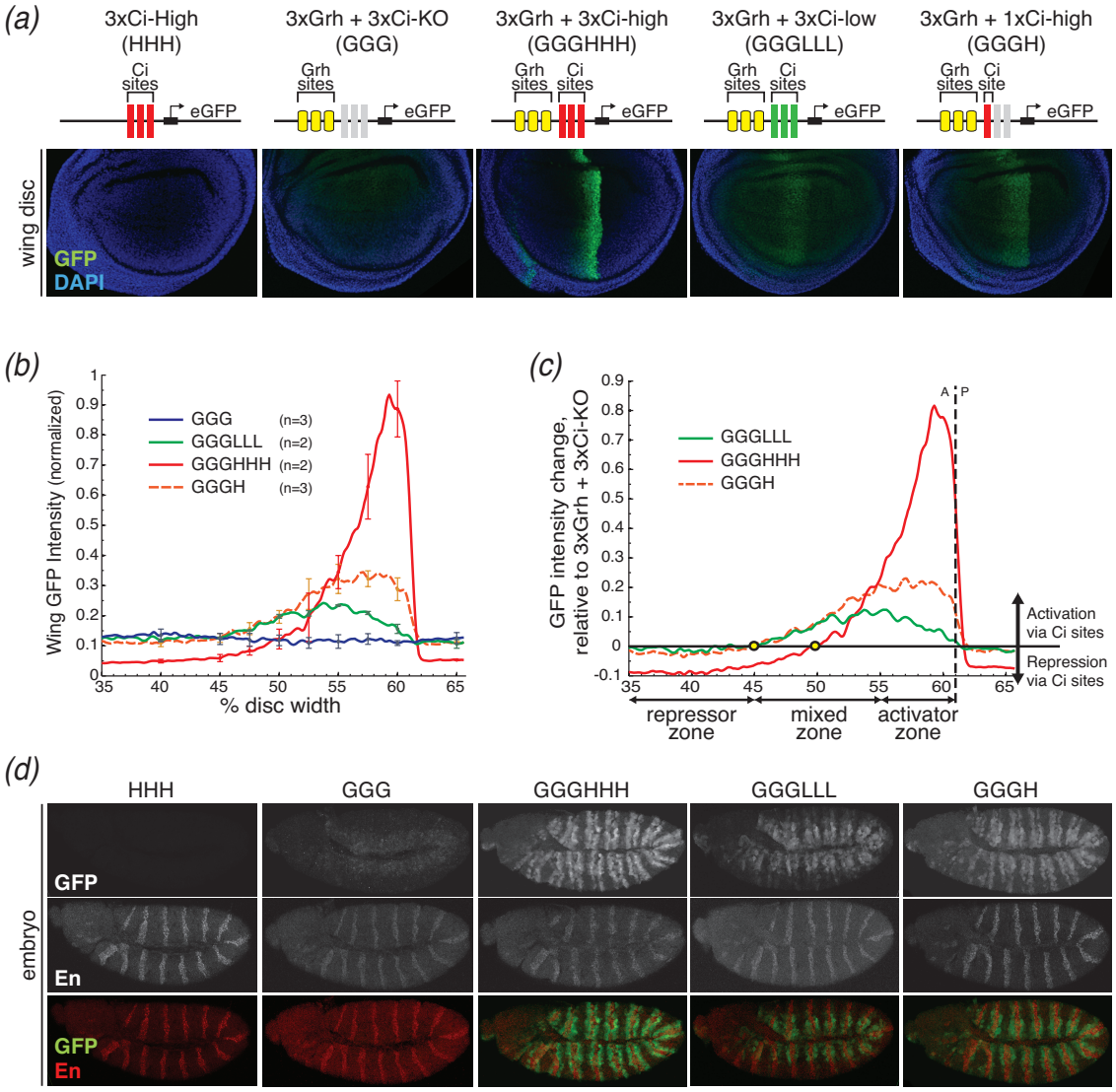


Figure 2.4 – Hedgehog gradient responses of synthetic enhancers in the wing and embryo.

(a) Diagrams of synthetic transgenic reporters, with Grainyhead (Grh) sites and low-affinity, high-affinity, and mutant Ci sites indicated. Confocal images of the pouch regions of wing imaginal discs from transgenic third-instar larvae are shown. GFP fluorescence data is internally normalized to a *dpp*D[Ci-*ptc*]-DsRed reference transgene (not shown). *(b)* Normalized GFP fluorescence wing data across the dorsal wing pouch. Error bars indicate one standard deviation. *(c)* Net effect of Ci sites on synthetic transgene expression (normalized GFP expression minus normalized GGG-GFP expression). Circles indicate the positions at which Ci input switches from net activation to net repression. *(d)* Confocal images of stage 11 transgenic embryos carrying the same synthetic GFP reporters. Top, GFP expression; middle, Engrailed (En) antibody, which marks Hh-producing cells; bottom, merge.

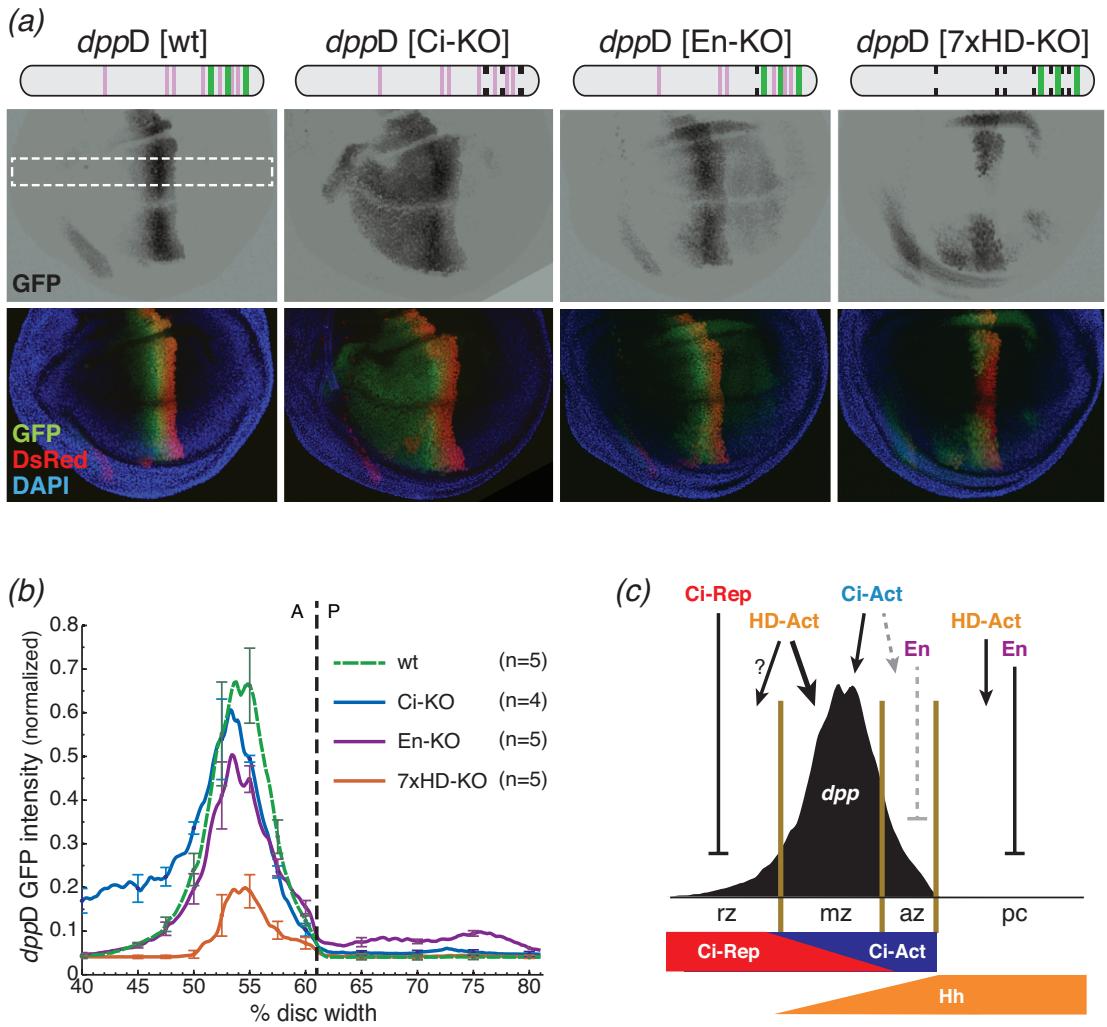
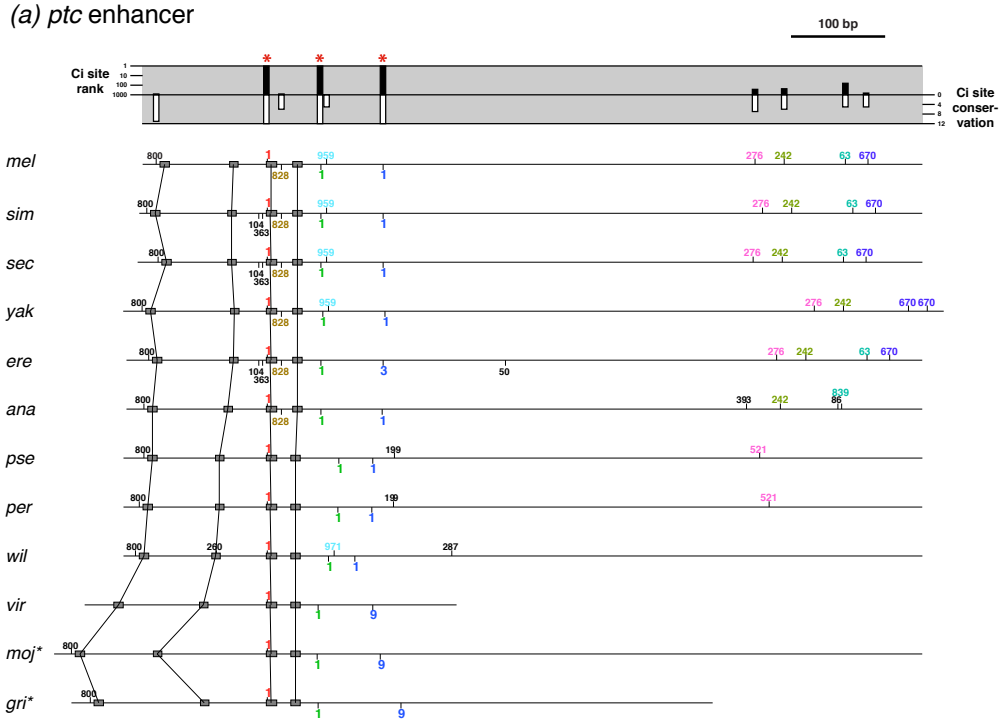


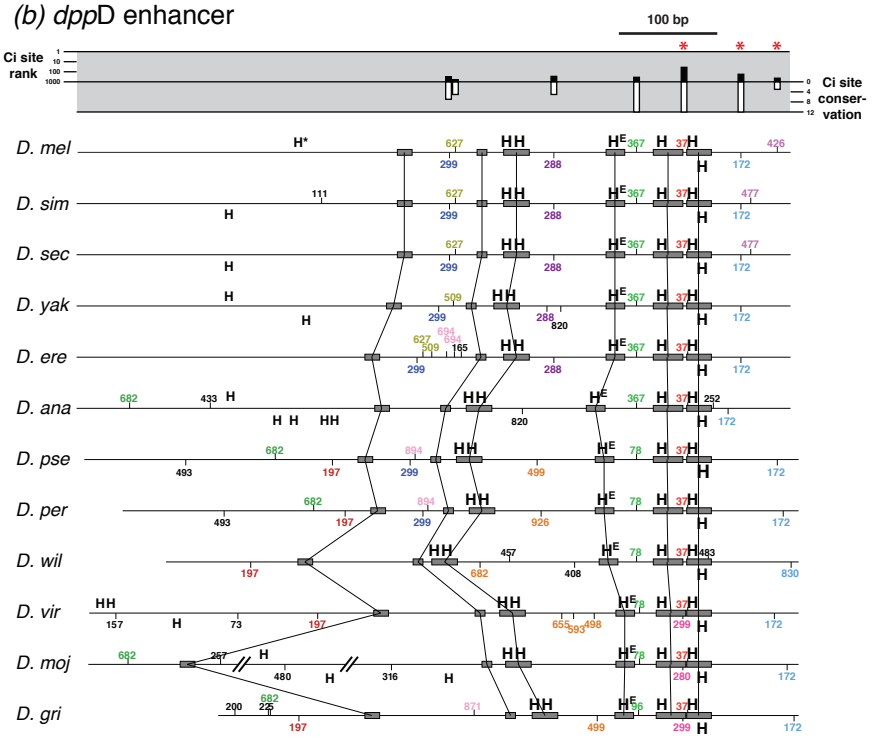
Figure 2.5 – Conserved homeodomain binding motifs integrate regulatory inputs into *dpp*D.

(a) Top, diagrams of *dpp*D enhancer constructs, with defined Ci and homeodomain (HD) binding sites as vertical bars (wide bars, wild type low-affinity Ci binding sites; narrow bars, HD motifs; broken bars, mutated sites). Middle and bottom, confocal images of wing imaginal discs from transgenic third-instar larvae are shown. Middle, GFP alone; bottom, GFP merged with *dpp*D[Ci-ptc]-DsRed reference transgene and DAPI nuclear stain. *(b)* Normalized GFP fluorescence data from transgenic dorsal wing pouches. Error bars indicate one standard deviation. Dashed line shows A/P compartment boundary. ***(c)*** Proposed model of the *dpp*D regulatory network across the wing A/P axis. A, anterior; P, posterior.

(a) ptc enhancer



*(b) dpp*D enhancer



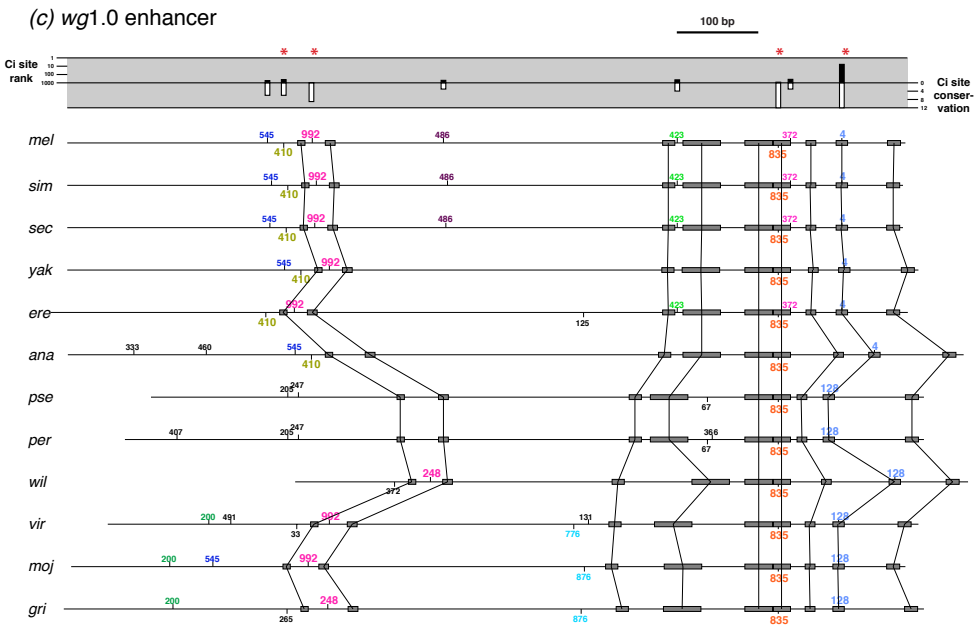


Figure 2.6 – Evolutionary sequence alignments of Hh/Ci-regulated enhancers across 12 sequenced *Drosophila* species genomes.

Red asterisks highlight previously described Ci binding sites (Alexandre et al., 1996; Müller and Basler, 2000; Ohlen and Hooper, 1997) (See Table 2.1). At the top of each alignment, black bars show ranking of Ci binding motifs based on predicted binding affinity; rank #1 is the optimal Gli binding motif. White bars show motif sequence conservation, calculated as the number of *Drosophila* species in which a similar motif (6/9 match or better) appears in a similar relative position. Grey rectangles show sequence blocks of 9 bp or greater that are shared among all 12 species, identified by the GenePalette program (Rebeiz and Posakony, 2004). Asterisks indicate functionally tested Ci sites (Müller and Basler, 2000; Parker et al., 2011). (bottom) Comparison of *dpp*D enhancer sequence similarity among 12 sequenced *Drosophila* species genomes. Species abbreviations: *mel, melanogaster; sim, simulans; sec, sechellia; yak, yakuba; ere, erecta; ana, ananassae; pse, pseudoobscura; per, persimilis; wil, willistoni; vir, virilis; moj, mojavensis; gri, grimshawi.* H, conserved HD binding motif (TAAT); HE, previously identified En site (Müller and Basler, 2000). Smaller H symbols denote non-conserved TAAT motifs. H* indicates a TAATCA motif found in our enhancer sequence which is given as TTATCA in official genome sequence. Asterisks next to species designations indicate that EvoPrinter was unable to find the *mel*-orthologous enhancer sequence in that genome; in those cases, species-specific BLAST was used to identify the orthologous sequence.

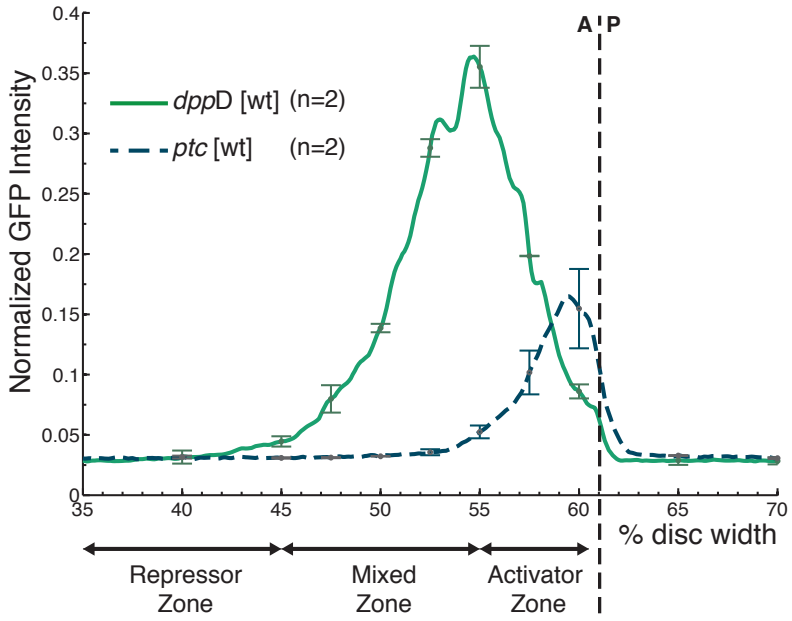


Figure 2.7 – Quantitative comparison of wing disc expression patterns driven by the *dpp*D and *ptc* enhancers.

The *dpp*D and *ptc* enhancers drive maximal expression in distinct signaling zones and at different levels. GFP fluorescence data were collected from third-instar imaginal wing discs and normalized to expression of a DsRed reporter transgene driven by *dpp*D[Ci-ptc]. Error bars indicate 1 standard deviation. A, anterior; P, posterior.

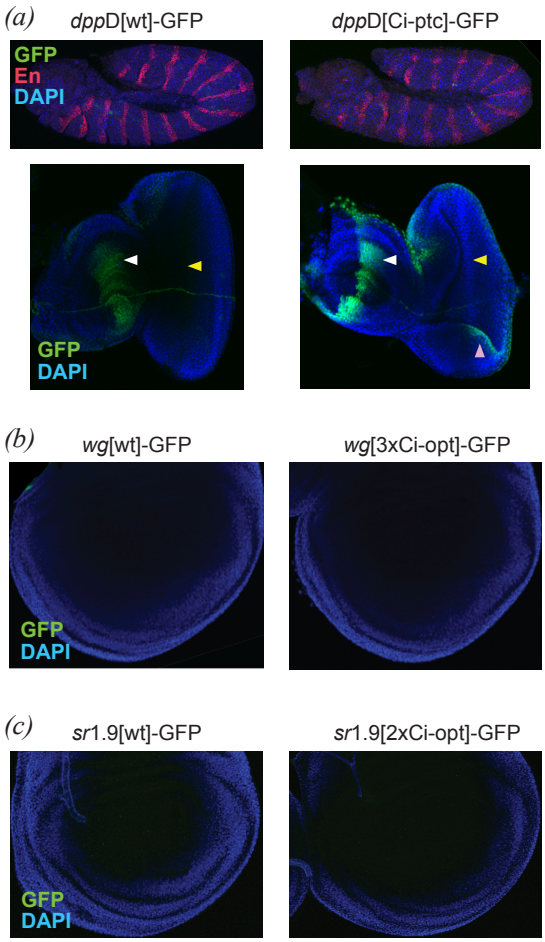


Figure 2.8 – Enhancers of dpp, wg, and sr respond to Hedgehog/Ci in a tissue-specific manner.

Improving Ci binding sites in the *dpp*D, wg and sr1.9 enhancers does not induce significant ectopic expression in wing discs or embryos. (a) Stage 11 embryos (top) and third-instar imaginal eye-antenna discs (bottom) from transgenic flies carrying *dpp*D[wt]-GFP and *dpp*D[Ci-ptc]-GFP. White arrowheads indicate antennal expression patterns; yellow arrowheads indicate the location of the morphogenetic furrow where dpp is normally expressed (Blackman et al., 1991), but the *dpp*D enhancer does not respond; pink arrowhead shows ectopic activity of *dpp*D[Ci-ptc] in the posterior margin of the eye disc. (b) *wg*[wt]-GFP and *wg*[3xCi-opt]-GFP do not respond to Hh/Ci in transgenic wing imaginal discs. (c) sr1.9[wt]-GFP and sr1.9[2xCi-opt]-GFP do not respond to Hh/Ci in transgenic wing imaginal discs.

(a) *ptc* enhancer (827 bp) 2R: 4,536,264-4,537,090

```
atgcatgcgcAGCCTGCCACGCACGCGCTtCCCCcAAACAaATaCACaCACACacactgagacgAAAGCT
CCAttgggcagcgctGCCGACGCTGAAGGCCGACatcggcagagctgaacgtttgggtagGGGACCACCCACAtCG
CTTGGCGGTTTCAGTTTAATGAAGGCAGaAACAaATTTATTTTTGGGTGGTCCaCACTGCAgCGAAAATAAACtAC
aGtgGcAAcAAcAAACcAGCAGccaAgGCacTTTGGGTGGTCCatgCaaaAAaaaaacaaattacggcatgcgaat
AACAATaGAAATTAGCGctCTcgtggcggagctatttgggtatattagagctacatatttatttgtttataaaaa
gtataaatgtaaacaatgagttccaagcattaagtccgtatgctcaacaattacattatcattattattatcactt
aaatatttacaaaggatatttaaacagtaatagatatatatttatttcttaatttctgttaacatatgtatttac
attggtagttattctttattttgcaacaagcattcataaattttatataacaaacttggtattttctcggaaaaac
tcctgaAtCacCCCTCGGTATTTTgtgcgttgagctatcgttaaagcagCCCTCGCAGAGAGCGTtCTCAAACCAA
AATGGCCGCACACGAaACAAGAGAGCGAGtGagagtagggagaGCGTcTgtgttgtgtgttgagtGTCgcCCAcGC
ACACAGGCGCAaAACaGtgCACACAGACgcccgctgggcaagagagagtgagagagagaaacagcggcgcgcg
```

(b.1) *dpp*D enhancer (724 bp) 2L: 2,481,154-2,481,870

```
aatattttgttcaattttGTAACaGtaGAGAGAGAGCAAaatgggttccactcacCTTGTCAGCCAGTCAGTCGCA
CATCCAGTTCCTTgGcCAtgtgcccctttcccctttgcgcttctcctccgtgttgccgattccgccccacacggag
ttagttttgtcgtacactgaaaaaaatgaacggaaatcatgaatttatgttttagtctgaaagggagagttaaagg
acacagcatattaatcaggaacataacttcccttaaacttctttcaacttctttcaGTGCAGCTGCAGGTGTGTGT
GTtTatGgaGGaCTGTGCGTCTCAAGTTTTCAACAACAAGATATAAGCCAATAAAGgAGGAACAcCgGCGAAAAGG
ATGAGCGGCCAGcccAGCAcacagggcACAAAAAGAAAGCGCAGGCAGGAGAATATACCTTAATTACGGTTAATGG
AGCGTTCGAAAAAACAAAAcCGATGGcttTATATGTGgcccaGTGTGTGTATCATATGTTGGATCTTCGGCCGAGT
GCCACGGCGAAATAACTTAATCACATTTCGAGAgGAGACGACCGCAAAAATCTGCGAGCCATGTTCGTAATTTTGT
ATATAAATGAGATGCGGCCACCTAATGAGCCTGATTAACCAaCCGggtcccgagatcttcggtTCCTCaCGGGCGG
TCtctacacccagcgccgctcccttgtacctccccccat
```

(b.2) *dpp*D [Ci-KO]

```
AATATTTTGTTCAATTTTGTAACAGTAGAGAGAGAGCAAAATGGGTTCCACTCACCTTGTCAGCCAGTCA
GTCGCACATCCAGTTCCTTGGCCATGTGCCCCTTTCCCCTTTGCGCTTCTCCTCCGTGTTGCCGATTCCGCCCCAC
ACGGAGTTAGTTTTGTCGTACACTGAAAAAAATGAACGGAAATCATGAATTTATGTTTTAGTCTGAAAGGGAGAGT
TAAAGGACACAGCATATTAATCAGGAACATAACTTCCCTTAAACTTCTTTCAACTTCTTTCAGTGCAGCTGCAGGT
GTGTGTGTTTATGGAGGACTGTGCGTCTCAAGTTTTCAACAACAAGATATAAGCCAATAAAGGAGGAACACCGGCG
AAAAGGATGAGCGGCCAGCCCAGCACACAGGGCACAAAAAGAAAGCGCAGGCAGGAGAATATACCTTAATTACGGT
TAATGGAGCGTTCGAAAAAACAAAACCGATGGCTTTATATGTGGCCCAGTGTGTGTATCATATGTTGGATCTTCGG
CCGAGTGCCACGGCGAAATAACTTAATCACATTTCGAGAGGAGACGACCGCAAAAATCTGCGAGCCATGTTCGTAA
TTTTGTATATAAATGAGATGCGGaaAaCTAATGAGCCTGATTAACCAACCGGGTCCCGAGATCTTCGGTTCCTCAC
GGtCttTCTCTACACCCAGCGCCGCTCCCTTGTAaaTaCCCCCCAT
```

(b.3) *dpp*D [Ci-ptc]

```
AATATTTTGTTCAATTTTGTAACAGTAGAGAGAGAGCAAAATGGGTTCCACTCACCTTGTCAGCCAGTCA
GTCGCACATCCAGTTCCTTGGCCATGTGCCCCTTTCCCCTTTGCGCTTCTCCTCCGTGTTGCCGATTCCGCCCCAC
ACGGAGTTAGTTTTGTCGTACACTGAAAAAAATGAACGGAAATCATGAATTTATGTTTTAGTCTGAAAGGGAGAGT
TAAAGGACACAGCATATTAATCAGGAACATAACTTCCCTTAAACTTCTTTCAACTTCTTTCAGTGCAGCTGCAGGT
GTGTGTGTTTATGGAGGACTGTGCGTCTCAAGTTTTCAACAACAAGATATAAGCCAATAAAGGAGGAACACCGGCG
AAAAGGATGAGCGGCCAGCCCAGCACACAGGGCACAAAAAGAAAGCGCAGGCAGGAGAATATACCTTAATTACGGT
TAATGGAGCGTTCGAAAAAACAAAACCGATGGCTTTATATGTGGCCCAGTGTGTGTATCATATGTTGGATCTTCGG
CCGAGTGCCACGGCGAAATAACTTAATCACATTTCGAGAGGAGACGACCGCAAAAATCTGCGAGCCATGTTCGTAA
TTTTGTATATAAATGAGATGCGaCCACCcAATGAGCCTGATTAACCAACCGGGTCCCGAGATCTTCGGTTCCTCAt
GGGtGGTCTCTACACCCAGCGCCGCTCCCTTGgACCaCCCaCCCAT
```

(b.4) *dpp*D [Ci1-ptc]

```
AATATTTTGTTCAATTTTGTAACAGTAGAGAGAGAGCAAAATGGGTTCCACTCACCTTGTCAGCCAGTCA
GTCGCACATCCAGTTCCTTGGCCATGTGCCCCTTTCCCCTTTGCGCTTCTCCTCCGTGTTGCCGATTCCGCCCCAC
ACGGAGTTAGTTTTGTCGTACACTGAAAAAAATGAACGGAAATCATGAATTTATGTTTTAGTCTGAAAGGGAGAGT
TAAAGGACACAGCATATTAATCAGGAACATAACTTCCCTTAAACTTCTTTCAACTTCTTTCAGTGCAGCTGCAGGT
GTGTGTGTTTATGGAGGACTGTGCGTCTCAAGTTTTCAACAACAAGATATAAGCCAATAAAGGAGGAACACCGGCG
AAAAGGATGAGCGGCCAGCCCAGCACACAGGGCACAAAAAGAAAGCGCAGGCAGGAGAATATACCTTAATTACGGT
TAATGGAGCGTTCGAAAAAACAAAACCGATGGCTTTATATGTGGCCCAGTGTGTGTATCATATGTTGGATCTTCGG
CCGAGTGCCACGGCGAAATAACTTAATCACATTTCGAGAGGAGACGACCGCAAAAATCTGCGAGCCATGTTCGTAA
TTTTGTATATAAATGAGATGCGaCCACCcAATGAGCCTGATTAACCAACCGGGTCCCGAGATCTTCGGTTCCTCAC
GGtCttTCTCTACACCCAGCGCCGCTCCCTTGTAaaTaCCCCCCAT
```

*(b.5) dpp*D [En-KO]

AATATTTTGTTCAATTTTGTAACAGTAGAGAGAGAGCAAAATGGGTTCCACTCACCTTGTCAGCCAGTCA
GTCGCACATCCAGTTCCTTGGCCATGTGCCCCTTTCCCCTTTGCGCTTCTCCTCCGTGTTGCCGATTCCGCCCCAC
ACGGAGTTAGTTTTGTCGTACACTGAAAAAAATGAACGGAAATCATGAATTTATGTTTTAGTCTGAAAGGGAGAGT
TAAAGGACACAGCATATTAATCAGGAACATAACTTCCCTTAAACTTCTTTCAACTTCTTTCAGTGCAGCTGCAGGT
GTGTGTGTTTATGGAGGACTGTGCGTCTCAAGTTTTCAACAACAAGATATAAGCCAATAAAGGAGGAACACCGGCG
AAAAGGATGAGCGGCCAGCCCAGCACACAGGGCACAAAAAGAAAGCGCAGGCAGGAGAATATACCTTAATTACGGT
TAATGGAGCGTTCGAAAAAACAAAACCGATGGCTTTATATGTGGCCCAGTGTGTGTATCATATGTTGGATCTTCGG
CCGAGTGCCACGGCGAAATAACTactagtCATTTCGAGAGGAGACGACCGCAAAAATCTGCGAGCCATGTTCGTAA
TTTTGTATATAAATGAGATGCGGCCACCTAAATGAGCCTGATTAACCAACCGGGTCCCGAGATCTTCGGTTCCTCAC
GGGCGGTCTCTACACCCAGCGCCGCTCCCTTGTACCTCCCCCCCAT

*(b.6) dpp*D [HD-KO]

AATATTTTGTTCAATTTTGTAACAGTAGAGAGAGAGCAAAATGGGTTCCACTCACCTTGTCAGCCAGTCA
GTCGCACATCCAGTTCCTTGGCCATGTGCCCCTTTCCCCTTTGCGCTTCTCCTCCGTGTTGCCGATTCCGCCCCAC
ACGGAGTTAGTTTTGTCGTACACTGAAAAAAATGAACGGAAATCATGAATTTATGTTTTAGTCTGAAAGGGAGAGT
TAAAGGACACAGCATATTAcaCAGGAACATAACTTCCCTTAAACTTCTTTCAACTTCTTTCAGTGCAGCTGCAGGT
GTGTGTGTTTATGGAGGACTGTGCGTCTCAAGTTTTCAACAACAAGATATAAGCCAATAAAGGAGGAACACCGGCG
AAAAGGATGAGCGGCCAGCCCAGCACACAGGGCACAAAAAGAAAGCGCAGGCAGGAGAATATACCTTAcaTACGGT
TAcaGGAGCGTTCGAAAAAACAAAACCGATGGCTTTATATGTGGCCCAGTGTGTGTATCATATGTTGGATCTTCGG
CCGAGTGCCACGGCGAAATAACTTAcaCACATTTCGAGAGGAGACGACCGCAAAAATCTGCGAGCCATGTTCGTAc
aTTTGTATATAAATGAGATGCGGCCACCTAcaGAGCCTGcaTAACCAACCGGGTCCCGAGATCTTCGGTTCCTCAC
GGGCGGTCTCTACACCCAGCGCCGCTCCCTTGTACCTCCCCCCCAT

(c.1) wg embryonic ectoderm enhancer (1026 bp) 2L: 7,302,331-7,303,356

gatatctgattaatatcatttttcattgaaacacattaccaagtaatatcgacccctttccttccattt
gcctgatatttaccatttcggtgccatttcccagtgaacCCATTAtgGATCTTGTTACCAtAATGTCAGCACTAAA
ATGGCTTCCTCcgctgccgagaagagatcgCCACCACCAcTCtactcTTTTGCACATGccaccaCCAccaccacca
cACGCTCCgGTGGTTCCCCtTTCGCCCCGCCATgTTTCCTCGATGGCTGCTCGtCGGATCGCACTTCCGGAGTTcg
cgGACCGCCaGCGTGGAcGATGATAATGCATTATGCATTCCgCCGcCTGggTTTTAATTGTTATTAACTCGcGATA
AGgAGAACTTTAACACCTATaGTTAGGcCGGgCCaGAAAATGTAGCCgAGCATGctaagaTATAATTTcaACATTT
TATAAGCACTTAACAaccCAAAAAccgctggtaatgcaccgAAAAAAAgtagttagaAgTAAtgtgagtaataata
gaatattagtttttaaatagtatagaaaaaaaacattaaatgacaagaaagttcaaagaaaacgtttacaaatgca
ttttatccaattacaaaatatcttcctgatattgagcgatttaaatgtagttgcattcttttccctctgtgcata
aatttttcggtaacaaaaactggctgcctTTTCCTTtCGCTGGCGGGGGGAACGGGAAATATAAATTGTCGGACCCC
GAAATTCTACACTGTCATCAAAGCGTAACTAAACGCCCATTTTTCTCAACTTTTTGGAGTCTCTTTCGGACCGGAC
AACCATTTTCGGTAGTTATTAGTGGCATATTTTGGCCTAAAGTGgGTGgtCCATGTCGAACAGCCAGCGCAGGCAA
CAAAATTTATTGTTTAGTAGTGGAGTGAAGTTAACAACAAAATGGACCaCCCAGCGAAAGAGAAAGACACGCAACG
ATCCCAACGCGGACCTGGCCAGAAAAAAATATTAACGCCTCGAG

(d.1) sr 1.9 enhancer (1889 bp) 3R: 13,913,053-13,914,941

TGACATAGTTcactagtgaaTGCAggCaTGATttaa**AGGGCGTTGCACTAATTGAAATTGCAAAT**tTttg
gttgccatgtAAAATTGATAGgAaTTTTATtccgCTTCGttgcaGTGTGAAACtcttcttgtgcCTGCCAcatttt
caGATTTATGAtacaggCAaGGCAGT**GTCA**ta**TTTAT**T**AATTG**A**AGAATTATAAG**CTgtTAAtccA**AGAT**C**C**CTCa
GT**TATGATTATTATGA**G**AGAT**GT**TC**aacaatcGaA**G**AtcgaaG**ATA**TAcgttcggaaaccctagacaatcaaca**TG
TGTGGCcagcctAgTAAAACAATAtAAATtGgG**A**AAAT**GC**tC**AATA**AAtatgaacagcaacA**AA**AGgA**A**AATATcT
tTaAT**AAATTTT**CAA**C**CA**A**tGc**AATTTGCATA**t**A**g**TGAATG**gAA**C**GCGTTATCGTAAATCA**ccAATAAAgtAA**GC**A
AAACC**G**AC**CG**AAA**TAAAAGTGTCA**CCTttATAtgtccacttatTAgATGcgccaGTaTCTAGacCCAAtagagtTT
GaCtATATtGATGATgATGAAAACgAATGCCGCAgcGcCACACACA**GACC**GCAGt**CCAATTGAAATGTTTTT**CgaA
AATTTACGACttGAAAAggAAAA**TTTTCT**AG**GCTATTAAAT**TATCgAAaTCaAATCAgCa**GACCACCAG****CATGATT
TtGGAAattGATACcacgaatggaaggttgtctttctacacaaggttcctaatgaattaaggtatataaattaaatg
ttgatatatttcttatatagattccttgattttagttttaatttttgatgcaatgcgtacacatttgtcatgacaa
aaatgtataaaaacctaaatagctaaaaaaaaaaagcataaatgagcatttttctcttgaagaatgcacgtgggcgt
ttataactaattttcaacgcaattcaattaaaagacgtacactttgttggAGCCAaCTAAATAtaagTg**AAAA**AtC
C**AGTC**A**AAGGAGTTGTC**CAtAGTAACCaTCAagtAA**GC**aACGACA**ACAAC**ATcagttgG**CAGTCATGCTAATGAAT
GG**AACGCG**ttttTttT**GTAAATTGTCTG**C**CAA**g**CGActgaatga**ATG**aATGAATGAgTAt**GTGACT**g**AATGAGTGA
CTGACAGC**CAAATGGCAGAGGAGTTG**tCgt**TTAATGA**G**CA**ACCCCTTcttagcCAACCaGtcga**CAACAGAA**g**C**Ca
CaaacgATGTCGatgtcgacgtagccctagggtattcatttaatccgtcagtttgcaGACTTaAGaGTCT**CTCATT
AATTCAT**TAAA**a**GG**T**GTTAAAAACGGAAGCCA**t**CGTCA**G**TTG**tT**GT**gAAggacc**TAA**c**ATGGTCAAT**c**A**agCGccA
AaatGaTTGgAaTTgcaaacgtTTtCAaTaAtATaATtatcaaattaaggtgatggaaagttgttagctggatttt
aattactaatatttcatcattttcaaacagaatgcagttcaaaacgtacgagcacttcacgtattattattcatag
attattcatagttgtagtttgtcggaatacaacgtggtgtatcttgtatcttagattttgtatcttaatattgggt
ttttgggccatactattgtctctagtgccatttctcaaatttatttttcctaacaaagtataagcCTtataaggtgg
cGAGCTaAttgtgctggcctttgttattacT**ATTG**CaT**GT**G**T**CACAT**TCTGAATGAGCAATTA**C**AGTTA**A**TAAGA**G
TACAAATGTCAAATGtccGaT**GTGGGCG**ATt**CG**GC**TAAGTATAATTAATTTG**ACTtGgGaCTTgg**GTTTCG**C**TCAG
TGTAAGa**CACTTGAACTAATTGACAAG**T**TTCCATAT**a**TAA**G**CATTTTG**T**GAATGAA**tgaattcgagttttc

(d.2) sr 1.9 [2xCi-opt]

TGACATAGTTCACTAGTGAATGCAGGCATGATTTAAAGGGCGTTGCACTAATTGAAATTGCAAATTTTTG
GTTGCCATGTAAAATTGATAGGAATTTTATTCCGCTTCGTTGCAGTGTGAAACTCTTCTTGTGCCTGCCACATTTT
CAGATTTATGATACAGGCAAGGCAGTGTCATATTTATTAATTGAAGAATTATAAGCTGTTAATCCAAGATCCCTCA
GTTATGATTATTATGAGAGATGTTCAACAATCGAAGATCGAAGATATACGTTCGGAAACCCTAGACAATCAACATG
TGTGGCCAGCCTAGTAAAACAATATAAATTGGGAAAATGCTCAATAAATATGAACAGCAACAAAAGGAAAATATCT
TTAATAAATTTTCAACCAATGCAATTTGCATATAGTGAATGGAACGCGTTATCGTAAATCACCAATAAAGTAAGCA
AAACCGACCGAAATAAAAGTGTCACCTTTATATGTCCACTTATTAGATGCGCCAGTATCTAGACCCAATAGAGTTT
GACTATATTGATGATGATGAAAACGAATGCCGCAGCG**GACCACCCA**GACCGCAGTCCAATTGAAATGTTTTTCGAA
AATTTACGACTTGAAAAGGAAAATTTTCTAGGCTATTAAATTATCGAAATCAAATCAGCA**GACCACCCA**CATGATT
TTGGAAATTGATACCACGAATGGAAGGTTGTCTTTCTACACAAGGTTCCTAATGAATTAAGGTATATAATTAAATG
TTGATATATTTCTTATATAGATTCCTTGATTTTAGTTTTAATTTTTGATGCAATGCGTACACATTTGTCATGACAA
AAATGTATAAAAACCTAAATAGCTAAAAAAAAAAAGCATAAATGAGCATTTTTCTCTTGAAGAATGCACGTGGGCGT
TTATAACTAATTTTCAACGCAATTCAATTAAAAGACGTACACTTTGTTGGAGCCAACTAAATATAAGTGAAAAATC
CAGTCAAAGGAGTTGTCCATAGTAACCATCAAGTAAGCAACGACAACAACATCAGTTGGCAGTCATGCTAATGAAT
GGAACGCGTTTTTTTTTGTAAATTGTCTGCCAAGCGACTGAATGAATGAATGAATGAGTATGTGACTGAATGAGTGA
CTGACAGCCAAATGGCAGAGGAGTTGTCGTTTAATGAGCAACCCCTTCTTAGCCAACCAGTCGACAACAGAAGCCA
CAAACGATGTCGATGTCGACGTAGCCCTAGGGTATTCATTTAATCCGTCAGTTTGCAGACTTAAGAGTCTCTCATT
AATTCATTAAAAGGTGTTAAAAACGGAAGCCATCGTCAGTTGTTGTGAAGGACCTAACATGGTCAATCAAGCGCCA
AAATGATTGGAATTGCAAACGTTTTCAATAATATAATTATCAAATTAAGGTGATGGAAAGTTGTTAGCTGGATTTT
AATTACTAATATTTCATCATTTTCAAACAGAATGCAGTTCAAAACGTACGAGCACTTCACGTATTATTATTCATAG
ATTATTCATAGTTGTAGTTTGTCGGAATACAACGTGGTGTATCTTGTATCTTAGATTTTGTATCTTAATATTGGGT
TTTTGGGCCATACTATTGTCTCTAGTGCCATTTCTCAAATTTATTTTCCTAACAAAGTATAAGCCTTATAAGGTGG
CGAGCTAATTGTGCTGGCCTTTGTTATTACTATTGCATGTGTCACATTCTGAATGAGCAATTACAGTTAATAAGAG
TACAAATGTCAAATGTCCGATGTGGGCGATTCGGCTAAGTATAATTAATTTGACTTGGGACTTGGGTTTCGCTCAG
TGTAAGACACTTGAACTAATTGACAAGTTTCCATATATAAGCATTTTGTGAATGAATGAATTCGAGTTTTC

***(e)* GGGLLL: Synthetic 3xGrh, 3xCi-low (174 bp)**

GCTAGCAGATCTAAACCGGTTATGCGAGTCTAGACTTGGAAACCGGTTATGCGAGTCTAGACTTGGAAAC
CGGTTATGCGGCCTGGCGCGCCACTCGATGCGGCCACCTAATGAGCCGCGGCCTCACGGGCGGTCTCTACGTCGAC
CCTTGTACCTCCCCCCCATCATGACGTC

***(f)* GGGHHH: Synthetic 3xGrh, 3xCi-high (174 bp)**

GCTAGCAGATCTAAACCGGTTATGCGAGTCTAGACTTGGAAACCGGTTATGCGAGTCTAGACTTGGAAAC
CGGTTATGCGGCCTGGCGCGCCACTCGATGCGACCACCCAATGAGCCGCGGCCTCATGGGTGGTCTCTACGTCGAC
CCTTGTGGGTGGTCCCCATCATGACGTC

***(g)* GGGH: Synthetic 3xGrh, 1XCi-high (174 bp)**

GCTAGCAGATCTAAACCGGTTATGCGAGTCTAGACTTGGAAACCGGTTATGCGAGTCTAGACTTGGAAAC
CGGTTATGCGGCCTGGCGCGCCACTCGATGCGACCACCCAATGAGCCGCGGCCTCACGGGCTGTCTCTACGTCGAC
CCTTGTACATCCCCCCCATCATGACGTC

***(h)* GGG: Synthetic 3xGrh, 3xCi-KO (174 bp)**

GCTAGCAGATCTAAACCGGTTATGCGAGTCTAGACTTGGAAACCGGTTATGCGAGTCTAGACTTGGAAAC
CGGTTATGCGGCCTGGCGCGCCACTCGATGCGGCAACCTAATGAGCCGCGGCCTCACGGGCTGTCTCTACGTCGAC
CCTTGTACATCCCCCCCATCATGACGTC

***(i)* HHH: Synthetic 3xCi-high (90 bp)**

GGCGCGCCACTCGATGCGACCACCCAATGAGCCGCGGCCTCATGGGTGGTCTCTACGTCGACCCTTGTGGG
GTGGTCCCCATCATGACGTC

Figure 2.9 – Annotated sequences of enhancers and synthetic constructs.

Enhancer sequence conservation in *(a)*, *(b.1)*, *(c.1)*, and *(d.1)* was determined with the online sequence analysis tool EvoPrinterHD (Yavatkar et al., 2008). Blocks of sequence that are strongly conserved across the genus *Drosophila* are in bold caps; slightly less well conserved sequences are in non-bold caps. Ci binding sites are highlighted in yellow; consensus HD binding sites are highlighted in magenta. Targeted base substitutions in predicted binding sites are shown in lower case. *(a)* The *ptc* enhancer fragment as shown in Figure 1b (Alexandre et al., 1996). *(b.1)* The *dpp*D enhancer used in this study is a truncated version of the 800bp "fragment 10" enhancer (Figure 1b) (Muller and Basler, 2000). *(b.2)* Mutated version of *dpp*D (Parker et al., 2011). *(b.3) dpp*D with improved Ci binding affinity (Parker et al., 2011). *(b.4) dpp*D with a single high-affinity Ci binding site and 2 mutant sites. *(b.5) dpp*D with known Engrailed site mutated (Muller and Basler, 2000). *(b.6) dpp*D with all predicted HD sites mutated. *(c.1) wg* 1.0 kb embryonic enhancer (Ohlen and Hooper, 1997). *(c.2)* High-affinity version of the *wg* enhancer (White et al., 2012). *(d.1) sr*1.9 embryonic enhancer (Piepenburg et al., 2000). *(d.2)* High-affinity version of *sr*1.9. *(e-i)* Synthetic Ci reporter constructs. Ci sites are highlighted in yellow; Grh binding sites are highlighted in red. *(e)* Synthetic enhancer with 3Grh sites and 3 low-affinity Ci sites. *(f)* Synthetic enhancer with 3Grh sites and 3 high-affinity Ci sites. *(g)* Synthetic enhancer with 3Grh sites and one high-affinity site (Parker et al., 2011). *(h)* Synthetic enhancer with 3 Grh sites and 3 mutant Ci sites to maintain binding site spacing relative to the previous constructs. *(i)* Synthetic enhancer with 3 high-affinity Ci sites.

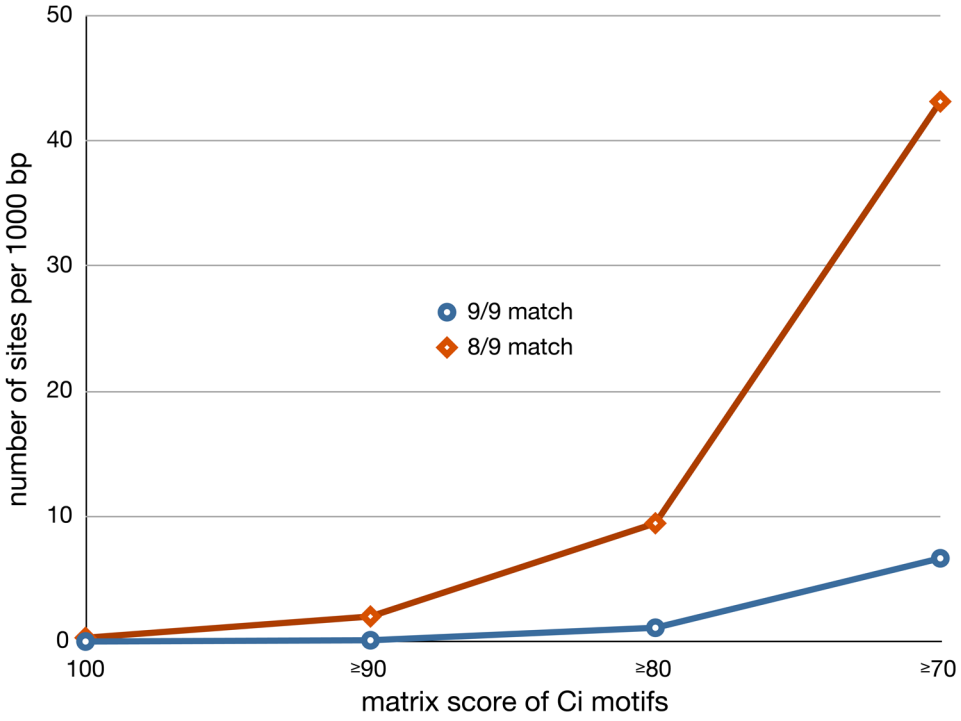


Figure 2.10 – Rates of occurrence of Ci motifs of various predicted affinities in genomic sequence.

50 randomly selected 1-kb segments of noncoding *D. melanogaster* genome sequence were analyzed for the presence of 9-bp Ci binding motifs with matrix similarity scores of 100, ≥90, ≥80, or ≥70, and for the presence of sequences with one mismatch (8/9 match) to those motifs. The average 1-kb sequence contains 9.5 sites, which with one nucleotide substitution, would be converted into Ci motifs with a matrix score of 80 or greater, and 43 sites that, with one substitution, would produce motifs with a score of 70 or greater. These estimates do not include other types of sequence evolution, such as indels, which present many more opportunities for acquiring Ci motifs. Non-consensus Ci motifs were detected with the GenePalette program (Rebeiz and Posakony, 2004).

Table 2.1 – Functionally characterized Hh/Ci-regulated enhancers in the *Drosophila* genome

Enhancer	Target gene	Sequence of CiBS	CiBS rank	Expression
ptc	*patched*	GACCACCCA*	1	wing disc
		GACCACCCA*	1	
		GACCACCCA*	1	
***dpp*D**	*decapentaplegic*	G**g**CCACC**t**A*	37	wing, leg and antennal disc
		GACC**g**CCC**g***	172	
		tACC**t**CCC**c***[1]	1512[2]	
***dpp*HO**	*decapentaplegic*	GACCACC**ag**	240	wing and haltere disc
		cgCCACCCA	103	
***wg*1.0**	*wingless*	GA**g**CA**g**CCA*	410	embryonic ectoderm
		G**t**CCAC**g**C**t***	1144[3]	
		G**tt**CAC**g**CA*	835	
		GACC**t**CCCA*	4	
***sr*1.9**	*stripe*	**cca**CAC**a**CA	4303[4]	embryonic ectoderm
		GACCACC**ag**	240	
knot	*knot*	G**c**CCACCCA	3[5,6]	wing disc
		G**a**CCACC**gc**	248[6]	
		G**g**CCAC**a**CA	43	
D-*h*	*hairy*	GACC**t**CCCA	4	leg disc
		GACCACC**at**	235	
oc7	*orthodenticle*	G**c**CC**t**CCCA	21	vertex primordium
pCB1.8	*blistered*	G**c**CCACC**ac**[1]	816	wing disc
		agCCACCCA[1]	101	
		GACCAC**agc**[1]	1259	
SRF-A	*blistered*	G**g**CCA**t**C**t**A[1]	735	wing disc

CiBS: Ci binding site(s). Bases deviating from the Ci/Gli consensus motif GACCACCCA (Hallikas et al., 2006; Kinzler and Vogelstein, 1990) are in bold and lower case. CiBS rank: rank of each 9-mer in order of predicted binding affinity for Ci (Hallikas et al., 2006).

* Relative *in vitro* binding affinity was experimentally measured (Parker et al., 2011; White et al., 2012).

[1] Potential binding site, not functionally and biochemically validated.

[2]This sequence, proposed as a possible Ci binding site by Müller et al. but not tested (Müller and Basler, 2000), overlaps another 9-mer with a better (but still poor) Ci site ranking (426). Neither site showed detectable binding to Ci *in vitro* (Parker et al., 2011) and neither is well-conserved evolutionarily.

[3]This sequence, proposed as a Ci binding site by Von Ohlen et al. (Ohlen et al., 1997), overlaps another 9-mer with a better (but still poor) Ci site ranking (992). A probe containing both overlapping sequences showed very weak binding to Ci *in vitro* (White et al., 2012), and the motif identified by Von Ohlen et al. (Ohlen and Hooper, 1997) is not evolutionarily well conserved (see Figure 3b and Supplementary figure S1c).

[4]This sequence, proposed as a Ci binding site by Piepenberg et al. (Piepenburg et al., 2000), overlaps another 9-mer with a better (but still poor) Ci site ranking of 681 (see Figure 3c).

[5]This site was reported as TGCCACCCA, a worse-ranking motif (rank 93), by Hersh et al. (Hersh and Carroll, 2005), but reference genomic sequence (BDGP R5/dm3) gives it as GCCCACCCA, and strong evolutionary sequence conservation supports the latter sequence.

[6]These two predicted sites were reported to be functionally irrelevant *in vivo* by Hersh et al. (Hersh and Carroll, 2005)—but see the previous footnote.

2.7 References

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.

Ashe, H.L., and Briscoe, J. (2006). The interpretation of morphogen gradients. Development (Cambridge, England) *133*, 385–394.

Baker, N.E. (1988). Localization of transcripts from the *wingless* gene in whole *Drosophila* embryos. Development (Cambridge, England) *103*, 289–298.

Balaskas, N., Ribeiro, A., Panovska, J., Dessaud, E., Sasai, N., Page, K.M., Briscoe, J., and Ribes, V. (2012). Gene regulatory logic for reading the sonic hedgehog signaling gradient in the vertebrate neural tube. Cell *148*, 273–284.

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

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 Biol. *5*, R61.

Biggin, M.D. (2011). Animal transcription networks as highly connected, quantitative continua. Dev. Cell *21*, 611–626.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific *phi*C31 integrases. Proc Natl Acad Sci U S A *104*, 3312–3317.

Blackman, R.K., Sanicola, M., Raftery, L.A., Gillevet, T., and Gelbart, W.M. (1991). An extensive 3' *cis*-regulatory region directs the imaginal disk expression of *decapentaplegic*, a member of the TGF-beta family in *Drosophila*. Development (Cambridge, England) *111*, 657–666.

Blair, S.S. (1992). Engrailed expression in the anterior lineage compartment of the developing wing blade of *Drosophila*. Development (Cambridge, England) *115*, 21–33.

Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2010). ChIP-Seq identification of weakly conserved heart enhancers. Nat. Genet. *42*, 806–810.

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.

Bray, S.J., and Kafatos, F.C. (1991). Developmental function of Elf-1: an essential transcription factor during embryogenesis in *Drosophila*. Genes & Development *5*, 1672–1683.

Burz, D.S., Rivera-Pomar, R., Jäckle, H., and Hanes, S.D. (1998). Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo. The EMBO Journal *17*, 5998–6009.

Campbell, G., and Tomlinson, A. (1998). The roles of the homeobox genes *aristaless* and *Distal-less* in patterning the legs and wings of *Drosophila*. Development (Cambridge, England) *125*, 4483–4493.

Campbell, G., Weaver, T., and Tomlinson, A. (1993). Axis specification in the developing *Drosophila* appendage: The role of *wingless*, *decapentaplegic*, and the homeobox gene *aristaless*. Cell *74*, 1113–1123.

Campos-Ortega, J.A., and Hartenstein, V. (1985). The Embryonic Development of Drosophila Melanogaster (Springer).

Capdevila, J., Estrada, M.P., Sánchez-Herrero, E., and Guerrero, I. (1994). The *Drosophila* segment polarity gene *patched* interacts with *decapentaplegic* in wing development. The EMBO Journal *13*, 71–82.

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

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

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

Eichberger, T., Kaser, A., Pixner, C., Schmid, C., Klingler, S., Winklmayr, M., Hauser-Kronberger, C., Aberger, F., and Frischauf, A.-M. (2008). GLI2-specific transcriptional activation of the bone morphogenetic protein/activin antagonist follistatin in human epidermal cells. J. Biol. Chem. *283*, 12426–12437.

Fisher, W.W., Li, J.J., Hammonds, A.S., Brown, J.B., Pfeiffer, B.D., Weiszmann, R., MacArthur, S., Thomas, S., Stamatoyannopoulos, J.A., Eisen, M.B., et al. (2012). DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. Pnas *109*, 21330–21335.

Freeman, M. (2000). Feedback control of intercellular signalling in development. Nature *408*, 313–319.

Frommer, G., Vorbrüggen, G., Pasca, G., Jäckle, H., and Volk, T. (1996). Epidermal egr-like zinc finger protein of *Drosophila* participates in myotube guidance. The EMBO Journal *15*, 1642.

Furriols, M., and Bray, S. (2001). A model Notch response element detects Suppressor of Hairless-dependent molecular switch. Curr. Biol. *11*, 60–64.

Gallet, A., Staccini-Lavenant, L., and Thérond, P.P. (2008). Cellular Trafficking of the Glypican Dally-like Is Required for Full-Strength Hedgehog Signaling and Wingless Transcytosis. Dev. Cell *14*, 712–725.

Goode, D., Snell, P., Smith, D., Cooke, J., and Elgar, G. (2005). Highly conserved regulatory elements around the gene may contribute to the maintenance of conserved synteny across human chromosome 7q36.3. Genomics *86*, 172–181.

Göke, J., Jung, M., Behrens, S., Chavez, L., O'Keeffe, S., Timmermann, B., Lehrach, H., Adjaye, J., and Vingron, M. (2011). Combinatorial Binding in Human and Mouse Embryonic Stem Cells Identifies Conserved Enhancers Active in Early Embryonic Development. PLoS Comput. Biol. *7*, e1002304.

Haag, E.S., and Lenski, R.E. (2011). L'enfant terrible at 30: the maturation of evolutionary developmental biology. Development (Cambridge, England) *138*, 2633–2637.

Halder, G., Polaczyk, P., Kraus, M.E., Hudson, A., Kim, J., Laughon, A., and Carroll, S. (1998). The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in *Drosophila*. Genes & Development *13*, 3900–3909.

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.

Han, K., Levine, M.S., and Manley, J.L. (1989). Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. Cell *56*, 573–583.

He, X., Duque, T.S.P.C., and Sinha, S. (2012). Evolutionary origins of transcription factor binding site clusters. Mol. Biol. Evol. *29*, 1059–1070.

Heberlein, U., Singh, C.M., Luk, A.Y., and Donohoe, T.J. (1995). Growth and differentiation in the *Drosophila* eye coordinated by *hedgehog*. Nature *373*, 709–711.

Heberlein, U., Wolff, T., and Rubin, G.M. (1993). The TGF beta homolog *dpp* and the segment polarity gene *hedgehog* are required for propagation of a morphogenetic wave in the *Drosophila* retina. Cell *75*, 913–926.

Hepker, J., Blackman, R.K., and Holmgren, R. (1999). Cubitus interruptus is necessary but not sufficient for direct activation of a wing-specific *decapentaplegic* enhancer. Development (Cambridge, England) *126*, 3669–3677.

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 (Cambridge, England) *132*, 1567–1577.

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.

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

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

Jaeger, S.A., Chan, E.T., Berger, M.F., Stottmann, R., Hughes, T.R., and Bulyk, M.L. (2010). Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. Genomics *95*, 185–195.

Jennings, B.H., Tyler, D.M., and Bray, S.J. (1999). Target Specificities of *Drosophila* Enhancer of split Basic Helix-Loop-Helix Proteins. Molecular and Cellular Biology *19*, 4600–4610.

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

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E.,

and Furlong, E.E.M. (2012). A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History. Cell *148*, 473–486.

Kasper, M., Schnidar, H., Neill, G.W., Hanneder, M., Klingler, S., Blaas, L., Schmid, C., Hauser-Kronberger, C., Regl, G., Philpott, M.P., et al. (2006). Selective modulation of hedgehog/GLI target gene expression by epidermal growth factor signaling in human keratinocytes. Molecular and Cellular Biology *26*, 6283–6298.

Kheradpour, P., Stark, A., Roy, S., and Kellis, M. (2007). Reliable prediction of regulator targets using 12 *Drosophila* genomes. Genome Res *17*, 1919–1931.

Kim, J., Sebring, A., Esch, J.J., Kraus, M.E., Vorwerk, K., Magee, J., and Carroll, S.B. (1996). Integration of positional signals and regulation of wing formation and identity by *Drosophila vestigial* gene. Nature *382*, 133–138.

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

Lee, J.J., Kessler, von, D.P., Parks, S., and Beachy, P.A. (1992). Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene *hedgehog*. Cell *71*, 33–50.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763.

Li, X.-Y., Thomas, S., Sabo, P.J., Eisen, M.B., Stamatoyannopoulos, J.A., and Biggin, M.D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding. Genome Biol. *12*, R34.

Maston, G.A., Landt, S.G., Snyder, M., and Green, M.R. (2012). Characterization of Enhancer Function from Genome-Wide Analyses. Annual Review Genomics Human Genetics *13*, 29–57.

Meireles-Filho, A.C.A., and Stark, A. (2009). Comparative genomics of gene regulation-conservation and divergence of *cis*-regulatory information. Curr. Opin. Genet. Dev. *19*, 565–570.

Michelson, A.M. (2002). Deciphering genetic regulatory codes: a challenge for functional genomics. Proc Natl Acad Sci U S A *99*, 546–548.

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

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 (Cambridge, England) *127*, 2999–3007.

Nahmad, M., and Lander, A.D. (2011). Spatiotemporal mechanisms of morphogen gradient interpretation. Curr. Opin. Genet. Dev. *21*, 726–731.

Nakano, Y., Guerrero, I., Hidalgo, A., Taylor, A., Whittle, J.R.S., 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.

Ng, J.-H., Kumar, V., Muratani, M., Kraus, P., Yeo, J.-C., Yaw, L.-P., Xue, K., Lufkin, T., Prabhakar, S., and Ng, H.-H. (2013). In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. Dev. Cell *24*, 324–333.

Ohlen, Von, T., and Hooper, J.E. (1997). Hedgehog signaling regulates transcription through Gli/Ci binding sites in the *wingless* enhancer. Mechanisms of Development *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.

Ong, C.-T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet. *12*, 283–293.

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. Science Signaling *4*, ra38–ra38.

Pavletich, N., and Pabo, C. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. Science *261*, 1701–1707.

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.

Peter, I.S., and Davidson, E.H. (2011). Evolution of gene regulatory networks controlling body plan development. Cell *144*, 970–985.

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.

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

Pires-daSilva, A., and Sommer, R.J. (2003). The evolution of signalling pathways in animal development. Nat. Rev. Genet.

Ptashne, M., and Gann, A. (2002). Genes & Signals (CSHL Press).

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.

Raftery, L.A., Sanicola, M., Blackman, R.K., and Gelbart, W.M. (1991). The relationship of *decapentaplegic* and *engrailed* expression in *Drosophila* imaginal disks: do these genes mark the anterior-posterior compartment boundary? Development (Cambridge, England) *113*, 27–33.

Rebeiz, M. (2002). SCORE: A computational approach to the identification of *cis*-regulatory modules and target genes in whole-genome sequence data. Proceedings of the National Academy of Sciences *99*, 9888–9893.

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

Rowan, S., Siggers, T., Lachke, S.A., Yue, Y., Bulyk, M.L., and Maas, R.L. (2010). Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. Genes & Development *24*, 980–985.

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 (Cambridge, England) *124*, 1313–1322.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U.,

Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the *cis*-regulatory sequences in the mouse genome. Nature *488*, 116–120.

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

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

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

Swarup, S., and Verheyen, E.M. (2012). Wnt/Wingless Signaling in *Drosophila*. Cold Spring Harbor Perspectives in Biology *4*, a007930–a007930.

Tanay, A. (2006). Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res *16*, 962–972.

Tuch, B.B., Li, H., and Johnson, A.D. (2008). Evolution of eukaryotic transcription circuits. Science *319*, 1797–1799.

Uv, A.E., Harrison, E.J., and Bray, S.J. (1997). Tissue-specific splicing and functions of the *Drosophila* transcription factor Grainyhead. Molecular and Cellular Biology *17*, 6727–6735.

Uv, A.E., Thompson, C.R., and Bray, S.J. (1994). The *Drosophila* tissue-specific factor Grainyhead contains novel DNA-binding and dimerization domains which are conserved in the human protein CP2. Molecular and Cellular Biology *14*, 4020–4031.

Vavouri, T., and Lehner, B. (2009). Conserved noncoding elements and the evolution of animal body plans. Bioessays *31*, 727–735.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

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 (Cambridge, England) *134*, 1977–1989.

Wartlick, O., Mumcu, P., Jülicher, F., and González-Gaitán, M. (2011). Understanding morphogenetic growth control -- lessons from flies. Nat Rev Mol Cell Biol *12*, 594–604.

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.

Whitington, T., Jolma, A., and Taipale, J. (2011). Beyond the balance of activator and repressor. *4*, pe29.

Wittkopp, P.J., and Kalay, G. (2012). *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. *13*, 59–69.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2008). Regulatory changes underlying expression differences within and between *Drosophila* species. Nat. Genet. *40*, 346–350.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. *25*, 1–47.

Wolpert, L. (2011). Positional information and patterning revisited. J. Theor. Biol. *269*, 359–365.

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 *9*, 106.

Yu, Y., Chen, Y., Kim, B., Wang, H., Zhao, C., He, X., Liu, L., Liu, W., Wu, L.M.N., Mao, M., et al. (2013). Olig2 targets chromatin remodelers to enhancers to initiate oligodendrocyte differentiation. Cell *152*, 248–261.

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

CHAPTER 3

Cis-regulatory Logic of Hedgehog/Gli Responses

3.1 Abstract

Hedgehog (Hh) signaling regulates tissue patterning and growth during development by precisely controlling the transcriptional output of key target genes using the Gli family of transcription factors. Despite the essential role of this pathway during organogenesis, it is not clear how Hh directs the transcription of different target genes at different developmental contexts. This gap in our understanding of Hh signaling events at the DNA level remains, in part, because of the limited number of known Hh-regulated enhancers. For instance, many target genes have been identified genetically, but the enhancers that mediate direct Hh/Gli-regulation are mostly unknown. Four enhancers have been identified for the target gene *patched* (*ptc*), which encodes a Hh receptor, one in flies and three in mammals. This gene is unique because, contrary to other targets, it is expressed in every tissue where Hh signaling occurs. When we examined the expression of the fly *ptc* enhancer in different Hh-regulated tissues we noticed that this enhancer drives tissue-specific expression, instead of general activation. Hence, we searched for additional regulatory elements in the *ptc* locus to better understand how this target gene responds to Hh/Gli in different contexts. Here we have identified and characterized a repertoire of Hh/Gli-regulated enhancers across the *ptc* locus in *Drosophila*. We also examined

enhancer-promoter communication, and found that the *ptc* promoter regulates the levels and spatial output of the enhancers to provide additional tissue-specificity to the Hh response. Based on previously identified *ptc* enhancers in mammals and our results, we propose that the *cis*-regulatory logic of *ptc* is conserved across species.

3.2 Introduction

The Hedgehog (Hh) signaling pathway patterns developing tissues across many species through a signaling cascade that ultimately controls gene transcription via the Gli family of transcription factors (Briscoe and Thérond, 2013). In the presence of the Hh ligand, target genes are transcribed, but in the absence of Hh, these genes are repressed. The *patched* (*ptc*) gene encodes one of the receptors of the pathway and it is a direct target in Hh-responsive tissues (Chen and Struhl, 1996; Marigo and Tabin, 1996). For this reason *ptc* expression is widely used as a functional readout of the pathway (Chen et al., 1999), and yet other Hh target genes have tissue-specific responses. For example, *decapentaplegic*, a *Drosophila* TGF-β growth factor, responds to Hh signaling in the developing wing but not in the embryonic ectoderm, even though both tissues rely on Hh signaling for patterning (Hursh et al., 1993; Müller and Basler, 2000). Because of the special transcriptional response of *ptc* to Hh signaling and the critical role that the Ptc receptor plays in transducing and controlling the range of the Hh morphogen gradient, a better understanding of the *cis*-regulation of this critical target gene is necessary.

Enhancers and the core promoter are among the most prevalent and commonly studied *cis*-regulatory elements in the genome (Levine et al., 2014). These non-coding sequences control the timing, levels and location of gene

transcription by containing clusters of binding sites for transcription factors (Yáñez-Cuna et al., 2013). Several fly and mammalian *ptc* enhancers have been identified. In mice, a limb enhancer (Lopez-Rios et al., 2014) and a neural tube enhancer (Vokes et al., 2007) that respond to Hh signaling and contain Gli binding sites (GBSs) were identified. In human cell lines, a promoter-proximal enhancer containing GBSs was also identified (Agren et al., 2004). In *Drosophila*, the previously identified ptc^{prox} enhancer (Figure 3.1A) contains a conserved cluster of three optimal, high-affinity Gli binding sites (GBSs) that respond to Hh signaling in the developing wing (Figure 3.1C) (Alexandre et al., 1996; Parker et al., 2011). In this tissue, Hedgehog is secreted by posterior cells, and received by anterior cells that express Gli/Ci and can transduce the signal. This directional gradient controls *ptc* expression along the A/P boundary as shown by the *ptc* enhancer trap (Figure 3.1B). A smaller version of this enhancer, designed around its three consensus GBSs is commonly used in luciferase assays carried out in mammalian and fly cells to measure Hh signaling activity (Chen et al., 1999). Mutation of the GBSs in the ptc^{prox} enhancer abolished expression in the wing which shows that this enhancer relies heavily on the high-affinity sites for activation (Figure 3.1D) (Alexandre et al., 1996). Since *ptc* is expressed in every Hh-responsive tissue, and the ptc^{prox} enhancer contains the best looking cluster of GBSs in the fly genome, we examined the expression of this enhancer in the embryonic ectoderm. In this tissue, Hh signaling establishes multiple signaling centers that are essential for early development (Nüsslein-Volhard and Wieschaus, 1980). *ptc* responds anteriorly and posteriorly of the Hh secreting cells (Figure 3.1E) (Hooper and Scott, 1989; Nakano et al., 1989). To our surprise, despite having extremely conserved high-affinity GBSs, this enhancer was not sufficient to respond in this context (Figure 3.1F). These data suggest

that *ptc* requires additional enhancers, possibly with different regulatory inputs, to respond in the embryonic ectoderm.

We screened about 27 kb of non-coding sequence in the *ptc* locus for additional tissue-specific enhancers, and found that most of this region contains regulatory information that fully or partially recapitulates the endogenous expression pattern of *ptc* in different developmental contexts (Figure 3.2). We also found that most of these enhancers respond to Hedgehog signaling via non-consensus GBSs for maximum levels of activation. Our *cis*-regulatory analysis went a step further from most studies, since we also characterized the core promoter. We found that the *ptc* promoter refines the output of each enhancer in a tissue-specific manner. The identification of these novel *cis*-regulatory elements highlight that although *ptc* transcription occurs in all Hh-responsive tissues, there is no general Hh-regulated *ptc* enhancer, and this seemingly simple response is most likely maintained by multiple *cis*-regulatory elements with diverse Hh/Gli-input.

3.3 Results

3.3.1 Multiple *ptc* enhancers drive expression in Hh responsive tissues

To identify novel enhancers in the *ptc* locus, we designed a mapping strategy consisting of 35 overlapping fragments that were tested *in vivo* for transcriptional activity (Figure 3.2A). We made GFP reporter transgenes containing these sequences upstream of the endogenous *ptc* promoter (fragment BA). This strategy was used to avoid disrupting specific enhancer-promoter interactions (Kadonaga, 2012). All these constructs were stably integrated into the fly genome, and GFP expression was scored in many Hh-responsive tissues.

We found that every tested fragment was active in at least one Hh-responsive tissue (representative results in Figure 3.2B-H).

In the larval wing and leg imaginal discs, secreted Hh controls *ptc* transcription in competent anterior cells in a stripe-like pattern along the anterior/posterior (A/P) boundary of these tissues (Briscoe and Thérond, 2013; Estella et al., 2012; Gradilla and Guerrero, 2013). We found multiple enhancers that drive *ptc*-like expression patterns in these tissues, including ZY and WU shown in Figure 3.2B and C, respectively.

Another site of Hh signaling is the larval intestine, where Hh is expressed in the rectum and small intestine (Chen and Struhl, 1996; Takashima and Murakami, 2001). *ptc* responds to this signal in the large intestine, or hindgut as shown by the enhancer trap (Figure 3.2D). We found that RP was active largely in this Ptc-positive domain (Figure 3.2D).

Hh signaling is also active in the eye-antennal imaginal disc. In the eye, Hh is secreted from differentiating photoreceptor cells and the posterior disc margin (Domínguez and Hafen, 1997). Fragment NM responds weakly in these cells, but drives a stripe-like pattern in the antennal disc, which is patterned as the wing and the leg discs (Figure 3.2E).

In addition to patterning larval tissues, Hh signaling is important to maintain stem cells during adulthood. For example, both the male and female reproductive organs require Hh signaling to maintain stem cell niches (Michel et al., 2012; Sahai-Hernandez and Nystul, 2013). In the testis, Hh is secreted from hub cells and received by nearby somatic cyst cells, where *ptc* is active (Michel et al., 2012). Interestingly, both the *ptc* enhancer trap and the Ptc antibody stain detect *ptc* expression in the hub as well (Michel et al., 2012). Our analysis confirmed that GC is active in these somatic cells and the hub itself (Figure 3.2F). This expression is unique from 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. Interestingly, autocrine Hh signaling has been reported to occur in many types of cancer (Liu et al., 2014; Tzelepi et al., 2011).

The adult ovary contains two stem cell niches: germ and follicle stem cells (FSCs). Hh signaling occurs in each of these cell types, and it is required for the proliferation of ovarian somatic cells, the follicle stem cells (FSCs) (Sahai-Hernandez and Nystul, 2013; Zhang and Kalderon, 2001). 1AC is most strongly activated in this region (Figure 3.2G). In addition, 1EH is also active in the adult ovary, but this expression occurs in the escort cells, a different cell type where *ptc* is also expressed (Figure 3.2H).

These non-overlapping, context specific enhancers represent an important theme of *ptc* regulation: multiple, separable regulatory inputs can pattern different tissues and are cell-type specific, even within a developmental context. Importantly, there was no single element that responded in all developmental contexts, strongly suggesting that there is tight control of context specific *ptc* transcription at the *cis*-regulatory level.

3.3.2 **Several enhancers contribute to *ptc* expression in the embryo**

Previous work has shown that a large fragment containing 12.5 kb of regulatory sequence upstream of the *ptc* TSS is sufficient to recapitulate *ptc* expression in the embryonic ectoderm from gastrulation on (Forbes et al., 1993). Here we show that fragments outside of this previously tested region are also able to pattern the embryo. We identified four non-overlapping DNA fragments that were active in cells known to respond to Hh (Figure 3.2J-M). For this analysis, we used the native *ptc* promoter, fragment BA, in order to maintain enhancer-promoter specificity (Butler and Kadanoga, 2001). As a control, we

tested the expression of the promoter fragment BA alone and found that it drives ubiquitous basal levels of expression (Figure 3.3I and I').

We found that fragments VR, PN, LJ, and 1EH are all expressed in cells flanking Hh positive cells, as marked by Engrailed (Figure 3.2J-M'). Two of these enhancers, LJ and PN, are located in the previously identified 12.5 kb transgene that is sufficient to respond in a *ptc*-like pattern (Forbes et al., 1993). Enhancer LJ is most reminiscent of endogenous *ptc* expression, although it does not perfectly replicate the *ptc* response, as the stripes are incomplete (Figure 3.1E compared to Figure 3.2K and K') (Hidalgo and Ingham, 1990). These data demonstrate that the previously identified 12.5 kb region contains two separable enhancers that are sufficient to activate expression in cells responding to Hh.

We found two additional enhancers outside of the proximal 12.5 kb region. VR and 1EH drove expression in Hh responsive cells in the embryonic ectoderm (Figure 3.2J-J' and M-M'). Enhancer VR was the least active enhancer *in vivo*, but still yielded a typical Hh response (Figure 3.2J-J'). Enhancer 1EH drives expression more broadly than any of the other embryo enhancers (Figure 3.2M-M'). In all cases, the enhancers were active in Hh-responsive cells and all recapitulated parts of the known *ptc* response to Hh signaling during germ band extension and retraction (Hooper and Scott, 1989). Based on the expression patterns, these enhancers are most likely regulated by Hh signaling through Gli.

3.3.3 **Embryonic *ptc* enhancers require low-affinity GBSs**

We identified putative GBSs in the embryo enhancers (Figure 3.2J-M). We found that all of the predicted GBSs deviated from the consensus sequence, except for one site. This is not surprising as it has been shown that non-consensus sites are the rule rather than the exception in most validated Hh-target

enhancers (Ramos and Barolo, 2013). We annotated and ranked all predicted GBSs in these enhancers using a positional weight matrix generated by Hallikas and colleagues (Hallikas et al., 2006). In this scale, the consensus GBS, GACCACCCA, is ranked 1 with an affinity matrix score of 100, representing that this 9mer has maximum affinity for Gli-binding *in vitro.* Each time a mismatch is introduced into the 9mer, *in vitro* affinity is weakened, so the matrix score decreases, as does the overall GBS rank (Figure 3.3B, bottom).

Using minimal versions of the newly identified embryo enhancers that were re-designed to contain evolutionarily conserved non-consensus GBSs, we generated additional GFP reporter constructs that used a minimal hsp70 promoter (hspmin) instead of the *ptc* promoter (Figure 3.3A). We switched promoters to determine the intrinsic activity of the enhancer without confounding effects from the endogenous promoter. These transgenes were integrated in the same genomic location to avoid changes in expression due to positional effects.

Enhancer LK, a subset of LJ, drives expression in cells posterior to the Hh positive cells, and it contains only non-consensus GBSs with lower predicted affinity (Figure 3B, top). When 6 of the 7 GBSs were mutated, leaving only the best predicted low-affinity site (Ci-16), we found that this enhancer loses most of its ability to respond to Hh, presumably because Gli can no longer recognize the mutated sites *in vivo* (Figure 3.3B, bottom). Furthermore, we increased levels of Hh by using a temperature sensitive promoter driving Hh cDNA (Ingham, 1993) and the expression domain of GFP activation was greatly expanded as compared to normal levels (data not shown). Together, these data establish LK as another example of a Hh responsive enhancer that functions via low-affinity Gli binding.

We next characterized 1EH, which contains 6 GBSs, 5 of which are non-consensus (Figure 3.3C, top). First we mutated the lone consensus GBS which resulted in minimal loss of activation. This result indicates that much of the

activity of 1EH comes from the remaining five non-consensus GBSs (Figure 3C, middle). We then mutated each of these low-affinity sites, and found that 1EH requires its low-affinity sites to respond to Hh (Figure 3.3C, middle). Furthermore, when the 3' low affinity sites (Ci82, Ci22) are optimized to match the consensus sequence, the level of activation was reduced (Figure 3.3C, top versus bottom panels). These results show that low-affinity GBSs are required for maximum levels of activation. These data offer further evidence that Gli occupancy at Hh-regulated embryo enhancers is likely being maintained at sub-maximum levels to coordinate gene expression in the embryo (Ramos and Barolo, 2013; White et al., 2012).

Finally, we examined fragment VT, a subset of VR that contains 8 non-consensus GBSs, ranging in affinity rank from Ci-5 to Ci-203 (Figure 3.3D, top). To determine if VT is a direct target of Hh, we mutated all of the predicted sites and found that the response is greatly reduced (Figure 3.3D, middle). As with 1EH, when the sequence of two low-affinity sites is altered to look like the optimal motif, the expression levels go down (Figure 3.3D, top versus bottom panels). These data suggest that embryonic ectoderm enhancers require low-affinity GBSs for maximum levels of activation. Interestingly, however, this response seems to be specific to that particular developmental context because when examined in the wing imaginal disc, we see a different response. The wild-type VT enhancer drives expression in the wing, and when the predicted GBSs were mutated the expression was lost (Figure 3.3D, top versus middle pannels). Upgrading the affinity of the enhancer results in increased expression (Figure 3.3D, bottom). These data suggest that while VT is able to respond in both the embryonic ectoderm and wing imaginal disc, it does so with different requirements for Gli occupancy.

We then tried to answer the question of how GBS affinity contributes to these Hh responsive enhancers in a heterologous luciferase assay in NIH/3T3 cells (Figure 3.4K). These cells contain all of the relevant Hh pathway components which allowed us to artificially stimulate the pathway using GLI1, the main transcriptional activator in mammals, to turn on Hh signaling in these cells. Our constructs were transfected in the presence or absence of GLI1 (Figure 3.3E). 1EH does not responsive to GLI1 in this assay, even though it clearly responded in the embryo (Figure 3.3C). Importantly, even after the affinity upgrade (1EH-GBS-opt), it is unable to respond to GLI1 induction (Figure 3.3E, right). Perhaps these binding sites are not accessible to GLI1 in this assay, or additional TFs that bind these sequences in the fly embryo are absent from these murine NIH/3T3 cells. Had we screened 1EH using this cell culture based method, as is frequently done in large enhancer screens, we would have called this an unresponsive enhancer and moved on, missing its strong response in the embryo.

Enhancer VT, on the other hand, is able to respond to induction with GLI1. Mutation of each low-affinity GBS resulted in reduced expression levels, as expected from *in vivo* wing and embryo data (Figure 3.3D). When the same low-affinity GBSs that were improved in the *in vivo* experiments were tested in this luciferase assay, we saw a better response as expected from this assay – better GBSs should give higher levels of activation.

3.3.4 **Characterization of novel *ptc* wing enhancers**

Since we identified many new Hh-regulated embryo enhancers in the *ptc* locus, we examined the expression patterns of multiple fragments in the wing imaginal disc (Figure 3.4A). We identified several enhancers that were sufficient

to generate *ptc*-like expression patterns in the wing (Figure 3.4A). In addition to DB (*ptc*^{prox}), we found that HF, VT, YU and ZY drove expression along the A/P boundary (Figure 3.4B).

To better understand the Hh/Gli input into these enhancers, GBSs were identified, as in the embryonic enhancers, most of the predicted motifs deviated from the consensus sequence (Figure 3.4D). DB contains the aforementioned cluster of three consensus sites in addition to a previously omitted non-consensus GBS. ZY contains one consensus site plus three non-consensus low-affinity sites. The rest of the fragments, HF, VT and YU, exclusively contain non-consensus sites.

Next we tested if the predicted GBSs in these enhancers were required to respond to Hh signaling by destroying these sites with targeted mutations. We found that the predicted GBSs are functionally required because the knockout constructs were unable to drive expression in the wing (Figure 3.5B). YU[GBS-KO]-GFP was the outlier as mutation of all the predicted GBSs still resulted in a stripe which was wider and weaker that the wild-type counterpart (Figure 3.4B and C). These data suggest that YU integrates additional positive inputs in the wing, and for this reason, the transcriptional logic of this enhancer is reminiscent of that of the Hh-regulated *dpp*D wing enhancer which also contains low-affinity GBSs (Muller and Basler, 2000; Parker et al., 2011).

In order to obtain a quantitative readout of the transcriptional output of these enhancers, and knowing the predicted Hh/Gli input into these sequences, we measured the GFP intensity levels at every position along the anterior-posterior axis of the wing (Figure 3.4C). Furthermore, we extracted the peak

intensity values for each construct to sort their responses more easily (Figure 3.4E, left). Surprisingly, DB was one of the least responsive wing enhancers *in vivo*, and opposite to what we expected, HF and YU, which contain only non-consensus GBS, produced the strongest wing stripes. ZY drives a stripe in the same cells as DB, but at much higher levels that are comparable to the expression of HF. VT produces the weakest stripe even though it is the sequence with the most predicted GBS. These data show that it is very hard to anticipate the transcriptional output of these enhancers from the predicted GBS input, and point to the contribution of other unknown tissue-specific inputs into these regulatory elements. Interestingly, regardless of the GBS composition, all the tested enhancers produced stripes that were abutting the A/P boundary, which is where *ptc* is normally expressed (Capdevila et al., 1994).

Next to determine if these enhancers were direct targets of Hedgehog signaling while simultaneously measuring the transcriptional output of these sequences in a different context, we carried out luciferase assays in NIH/3T3 cells that were treated with or without GLI1. All the wild-type enhancers are directly regulated by GLI1 because luciferase activity decreases when the predicted GBSs are mutated (Figure 3.4F). In cell culture, the transcriptional output of the enhancers is a better match to the predicted GBS input than it is in the wing. (Figure 3.4E versus F). For instance in this assay, high-affinity enhancers DB and ZY drive stronger outputs than the lower-affinity enhancers VT and YU, which have the weakest outputs. The exception in this context is HF,

which produces the strongest response although it has the fewest number of non-consensus GBSs.

3.3.5 The *ptc* proximal enhancer region is highly cooperative

To further understand how Hh/Gli-regulated enhancers integrate signaling inputs, we tested different fragments from the upstream proximal region of *ptc* (Alexandre et al., 1996; Forbes et al., 1993) (Figure 3.5A and B). All the fragments drove *ptc*-like stripes in the wing (Figure 3.5B). The position of these stripes was the same in all the constructs, but the patterns and levels of gene expression varied greatly. HF, GE, GB, HB, and JB produced full wing stripes, while DB, H and JI were mostly active in the wing pouch (Figure 3.5B).

We tested four non-overlapping fragments, DB, GE, H and JI, to determine the transcriptional output of these sequences in response to Hh/Gli. All of these fragments were sufficient to drive expression in the wing, which shows that this region contains four separate enhancers that, when combined (JB) produce the strongest and fullest wing stripe, and the contribution of these enhancers to the final pattern seen in JB is not additive (Figure 3.5B and E). These data show that the upstream *cis*-regulatory region of *ptc* is highly cooperative.

Since the relative order of these sequences to each other was altered, we compared the expression of DB, GB, HB and JB to determine the transcriptional output of these fragments when part the sequence context and spacing from the promoter is maintained. H and JI drive weak wing stripes, but when added to GB and HB respectively, they have a significant contribution in the final pattern of the stripe that is more that the individual outputs (Figure 3.5B and D-E). This analysis shows that the context of these enhancers has a dramatic effect on how these sequences interpret Hh signals, and that fragments that are appear to be

irrelevant, play important roles in refining the final pattern when placed in more endogenous contexts.

3.3.6 **Definition of the *ptc* core promoter**

In addition to the *cis*-regulatory information provided by the enhancers already described, we also found that the *ptc* promoter contains information that contributes to enhancer activity. For example, *ptc*BA contributes to broad activation in the embryonic ectoderm (Figure 3.2I and Figure 3.6A). This result potentially explains the low-level, Hh-independent activation of *ptc* that is required to keep the pathway off.

To identify which part of the promoter is responsible for this broad activity, we looked for promoter-specific binding sites in the region to identify a core promoter element. We found three highly conserved promoter elements in the 5' end of the BA fragment: initiator element (INR), downstream promoter element (DPE) and a motif ten element (MTE), all required for attracting the core promoter machinery associated with RNA Polymerase II (Butler and Kadonaga, 2002; Lim et al., 2004). We used these criteria to define region A, the *ptc* core promoter (Figure 3.6B). Upstream of A, we found several well conserved GAGA factor sites (GAF), known to open chromatin to allow transcription to occur (Wang et al., 2005). In fact ChIP-chip results demonstrate that GAGA factor is significantly associated with the *ptc* locus (Lee et al., 2008). We also identified a highly conserved PHO/YY1 binding site suggesting that this chromatin modifier, known to both activate and repress target gene transcription maybe also acting at the *ptc* locus (Thomas and Seto, 1999; Yant et al., 1995).

This sequence analysis suggested that *ptc*BA might be two elements with separable functions. We tested this hypothesis by placing *ptc*B upstream of the

hspmin promoter, which contains a TATA box, but no INR or DPE. When these constructs were examined in the absence of enhancers, *ptc*BA was the only element capable of driving broad levels of activation (Figure 3.6A, top row). These results suggest that *ptc*A and B work together to activate expression, but are not sufficient alone to activate transcription.

3.3.7 **The *ptc* promoter controls levels and spatial transcriptional outputs**

We tested different enhancers with the two promoters to determine whether enhancer-promoter communication was important to integrate Hh signaling at the *ptc* locus. We found that enhancer PN responds to Hh signaling when examined with *ptc*BA, but is mostly inactive with hspmin (Figure 3.6A). We next examined PN on the different promoter combinations. Neither PN+B+hspmin nor PN+A drives a pattern that is comparable to PN+BA in either GFP intensity or spatial patterning. These data suggest that both B and A are required to achieve maximum levels of activation and proper spatial patterning signals.

In addition, two more enhancers, LJ and 1EH, had major differences in expression patterns when examined on different promoters. LJ+hspmin responds posteriorly to the Hh signal but LJ+*ptc*BA responds both anteriorly and posteriorly – two different responses from the same enhancer (Figure 6A). We found that not only does the *ptc* promoter contribute to spatial patterning, it also affects expression levels, as demonstrated by 1EH. This enhancer responds to Hh both anteriorly and posteriorly, but it is much stronger with the hspmin promoter than with the ptcBA promoter (Figure 3.6A).

We examined the activity of GC, HF, YU and ZY with the two different promoters to determine if the enhancer-promoter specificity that we see in the

embryo also occurs in the wing. As in the embryo, we found that *ptc*BA alone drives ubiquitous expression whereas hspmin alone does not (Figure 3.6C and D). GC and HF produce similar expression patterns with either promoter in anterior cells, but drive low levels of GFP expression in posterior cells (Figure 3.6C and E). YU+hspmin drives the highest peak intensity levels of all the wing enhancers that we tested (Figure 3.4E), however with the *ptc*BA promoter the activity of this enhancer is dampened to levels similar to GC and HF (Figure 3.6E). The activity of ZY+*ptc*BA is also lower compared to ZY+hspmin. Overall in the wing, *ptc* enhancer+*ptc*BA constructs result in dampening of expression, which suggests that in this context *ptc*BA maintains transcription below a certain level. This regulatory function is probably critical to avoid the overexpression of *ptc* in the wing, which is known to alter the Hh gradient by reducing the range of signaling (Johnson et al., 1995).

3.4 Discussion

Our detailed mapping and characterization of the *cis*-regulatory sequences of *ptc* is the most in depth analysis of a Hh target gene to date. Since *ptc* encodes a receptor of the pathway and must be expressed in all tissues where Hh signaling occurs to transduce and attenuate the signaling gradient, this was the perfect case study to search for Hh-regulated enhancers locus-wide. Recent studies have identified *ptc* enhancers in mammals using different genome-wide techniques like Chip-seq and 4C analysis (Lopez-Rios et al., 2014; Vokes et al., 2007). However, these powerful techniques are limited to capture snapshots of protein-DNA complexes at a given time point in a specific cell-type. For that reason, they potentially miss critical regulatory regions that may contribute to expression in different cell-types or stages. Our strategy was completely

unbiased as we examined the expression patterns of numerous fragments of the *ptc* locus regardless of chromatin marks or predicted GBSs. It would have been interesting to compare Gli/Ci Chip data with our *in vivo* findings, but there is no reliable antibody to perform the experiments in *Drosophila*. Previous attempts have been made to investigate genome-wide binding of Gli/Ci using DamID, but the results were ambiguous, although the authors reported extensive binding at the *ptc* locus (Biehs et al., 2010).

Here, we have shown that the unique Hh-wide response of *ptc* is achieved through separate tissue-specific enhancers with distinct configurations of GBSs. These enhancers direct expression throughout development and are active in adult stem cell populations. We also characterized the endogenous *ptc* promoter which provides additional tissue-specificity to the intricate regulatory logic of the locus. Our analysis also included the first Hh-regulated sequence to be shown under direct control of the pathway via a cluster of consensus GBSs (ptc^{prox} or fragment DB) (Alexandre et al., 1996).

It was this discovery combined with the *in vitro* identification of the consensus Gli motif that shaped the way we think about how Hh target genes interpret the gradient (Alexandre et al., 1996; Kinzler and Vogelstein, 1990). Our data and the work of others changes this paradigm because now it is clear that most Hh-regulated enhancers contain functional GBSs that deviate from the optimal sequence (Oosterveen et al., 2012; Parker et al., 2011; Ramos and Barolo, 2013; Winklmayr et al., 2010). In most of these cases, however, the functional role of these enhancers remains to be determined. The validation of these sequences in their native context will be easier in the near future with groundbreaking genome engineering tools like the CRISPR/Cas9 system (Cong et al., 2013; Sebo et al., 2014). By using these techniques we will test if the newly identified enhancers are functionally redundant or if they are required for the final

pattern of *ptc* expression at multiple stages of development. The latter scenario makes sense from an evolutionary standpoint, because breaking down the transcriptional response to Hh signaling into separable units may provide a more robust network to avoid pleiotropic effects due to minimal changes to specific enhancer sequences (Barolo, 2011).

We propose that the *cis*-regulatory logic used to regulate *ptc* in *Drosophila*, is likely conserved in mammals. The recent finding of separable tissue-specific enhancers in the *ptch1,* the mammalian version of *ptc*, locus is analogous to what we have seen at the *ptc* locus in *Drosophila*. Since we found that most of the locus contains regulatory information, it will be interesting to take a closer look at the *ptch1* locus to determine if there are additional enhancers that drive tissue-specific expression. Another parallel between the fly and the mammalian *ptc* regulatory logic, is that all the known mammalian enhancers contain mostly non-consensus GBSs (Agren et al., 2004; Lopez-Rios et al., 2014; Vokes et al., 2007). Interestingly the relative location of some of these enhancers in the locus is also conserved. For example, the human, the limb and the *Drosophila* ptc^{prox} enhancers are located directly upstream of the TSS (Agren et al., 2004; Alexandre et al., 1996; Vokes et al., 2007).

Many examples of malfunctioning enhancers or enhanceropathies have been described in the literature, but none for *ptch1* (Herz et al., 2014; Smith and Shilatifard, 2014). All of the previous *ptch1* mutations are found in the coding sequence, but this is probably a result of mainly ignoring the non-coding sequences flanking the locus (Guo et al., 2013). In the future, it will be interesting if new molecular causes of *ptch1*-related diseases are associated to tissue-specific enhancers.

3.5 Materials and methods

GBS prediction and ranking

Matrix similarity scores were calculated as described (Quandt et al., 1995) with *in vitro* Gli binding data generated by Hallikas and colleagues (Hallikas et al., 2006). GBSs were identified *in silico* by screening the *ptc* locus for defined motifs using Genepalette (Rebeiz and Posakony, 2004).

DNA sequence alignments

Sequences and multi-species alignments were obtained from the UCSC Genome Browser.

DNA cloning and mutagenesis

Wild-type *ptc* enhancers were amplified by standard PCR (Roche Expand High Fidelity PCR System) from genomic DNA. Primers are listed on Table 1. Enhancer constructs were sub-cloned into the pENTR/D-TOPO plasmid (Life Technologies) by TOPO cloning. Enhancers tested with the hspmin promoter were subsequently cloned into the pHPdesteGFP transgenesis vector (Boy et al., 2010) by 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., 2010).

Drosophila transgenesis

P-element transformation was performed as described (http://sitemaker.umich.edu/barolo/protocols) in the w^{1118} strain. Site-directed

transformation by embryo injection was performed as previously described (Bischof et al., 2007). Reporter transgenes were integrated into a *phiC*31 landing site at genomic position 86FB.

Immunohistochemistry and confocal microscopy

Embryos and third-instar imaginal discs were fixed and stained using standard methods as previously described (Parker et al., 2011; White et al., 2012). Embryos were staged as described (Campos-Ortega and Hartenstein, 1985). Ovaries and testes were dissected between 0-2 days after eclosure and fixed in 4% PFA for 25 minutes at room temperature, washed 3 times in PBS with TritonX (0.1%) and subjected to antibody staining. Third-instar larval gut was dissected and fixed like the testes and ovaries. The primary antibodies used included rabbit anti-EGFP (Invitrogen) diluted 1:100, mouse 40-1a anti-beta-galactosidase (Developmental Studies Hybridoma Bank) diluted 1:200, mouse anti-En (Developmental Studies Hybridoma Bank) diluted 1:50. AlexaFluor488, AlexaFluor555, and AlexaFluo568 conjugates with secondary antibodies from Invitrogen were used at 1:2000 dilutions. Tissues were mounted in Prolong Gold antifade reagent with DAPI (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. Samples to be directly compared were fixed, prepared, and imaged under identical confocal microscopy conditions and settings.

Luciferase assays

NIH/3T3 cells were cultured at 37°C, 5% CO_2, 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 10^4 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. 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%$C0_2$ for an additional 48 hours. GLI1 DNA was added immediately after low-serum change where relevant to activate the HH pathway. Cells were harvested and luciferase and beta-galactosidase activities were measured using Luciferase Assay System (Promega Cat. # E1501) and BetaFluor β-gal assay kit (Novagen, Cat. #70979-3). Multiple assays were performed and each sample was assayed in triplicate.

Quantitation of transgenic reporter expression data

Fluorescence data from wing confocal images were collected and quantified as previously described using the Matlab program *Icarus* (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.

3.6 Acknowledgements

This research was supported by the Cellular and Molecular Biology Training Grant (NIH T32-GM007315) to D.S.L. and A.I.R., the RSP Training Grant to D.S.L., the Center for Organogenesis predoctoral fellowship (NIH T32-

HD007505) to A.I.R., NIH grant GM076509 and NSF grant MCB-1157800 to S.B. We thank, Lauren Hillers, Victoria Blake, Charlie Katzman, Annie Azrak, Shelby Peterson, Autumn Holmes, Jessica Frick, Katherine Gurdziel and Elliott Ortiz-Soto for research support. We thank the staff at the Microscopy and Image Analysis Laboratory (MIL) at the University of Michigan Medical School for their assistance obtaining confocal images. We are grateful to members of the Barolo laboratory, members of the University of Michigan developmental genetics group meeting, and the UMfly *Drosophila* community for helpful discussions.

The text and figures presented in this chapter are part of the manuscript: *Cis*-regulatory Logic of Hedgehog/Gli Responses. Lorberbaum, DS*, Ramos, AI*, Carpenter, B, Parker, DS, Hens, K, Chiang, A, Deplancke, B, Allen, B, and Barolo, S. (*Equal contribution).

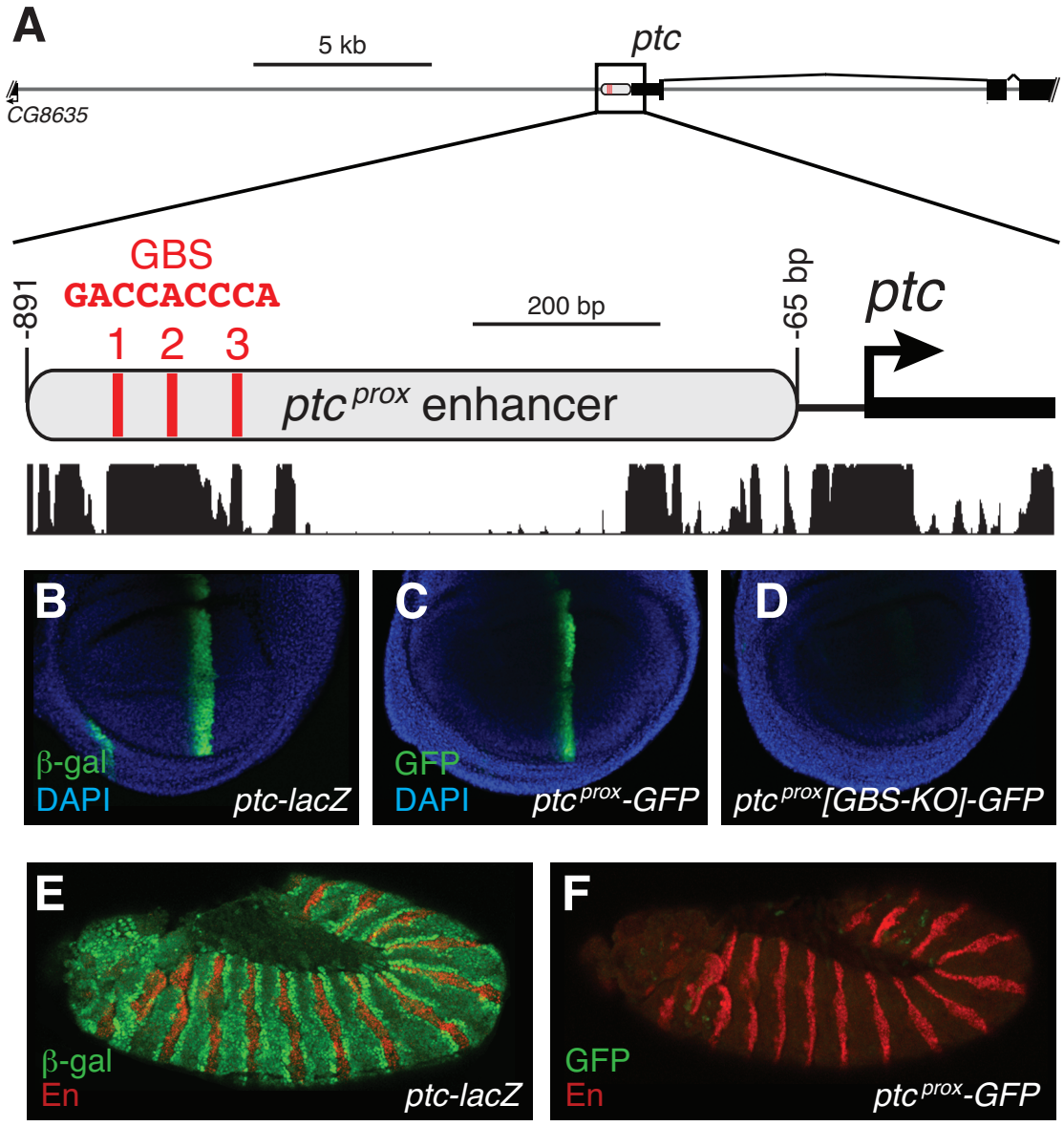


Figure 3.1 – The canonical *ptc*^{prox} enhancer is unable to respond to Hh signaling in the embryo despite its cluster of highly conserved optimal GBSs.

(A) Genomic location of the *ptc*^{prox} enhancer which has three GBSs shown as vertical red bars. The conservation of these sequences across 12 *Drosophila* species is shown below. (B) Third-instar wing imaginal disc showing the expression pattern of the *ptc* enhancer trap (*ptc-lacZ*). (C-D) Third-instar wing imaginal disc showing GFP expression driven by *ptc*^{prox}*-GFP* or *ptc*^{prox}*[GBS-KO]-GFP* reporter transgenes. (E) Stage 12 *ptc-lacZ* embryo. (F) Stage 12 transgenic embryo showing GFP expression driven by *ptc*^{prox}*-GFP*.

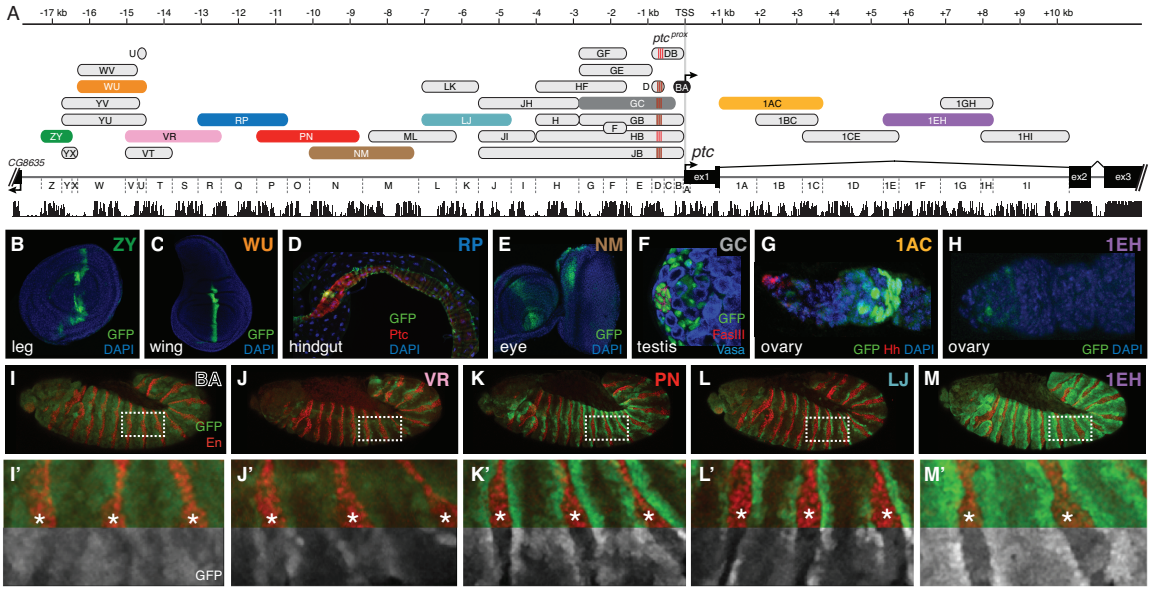


Figure 3.2 – The *ptc* locus is replete with enhancers that drive expression in multiple Hh-responsive tissues.

(A) The *ptc* locus showing the location of embryo enhancers displayed as colored cigars. The *ptc* coding region is shown as black rectangles. Evolutionary conservation across 12 *Drosophila* species is shown below the locus. (B-H) Selected enhancers that are active in Hh-responsive tissues (I-M) Enhancers that are active in the embryonic ectoderm (I'-M') magnification of selected regions in panels I-M.

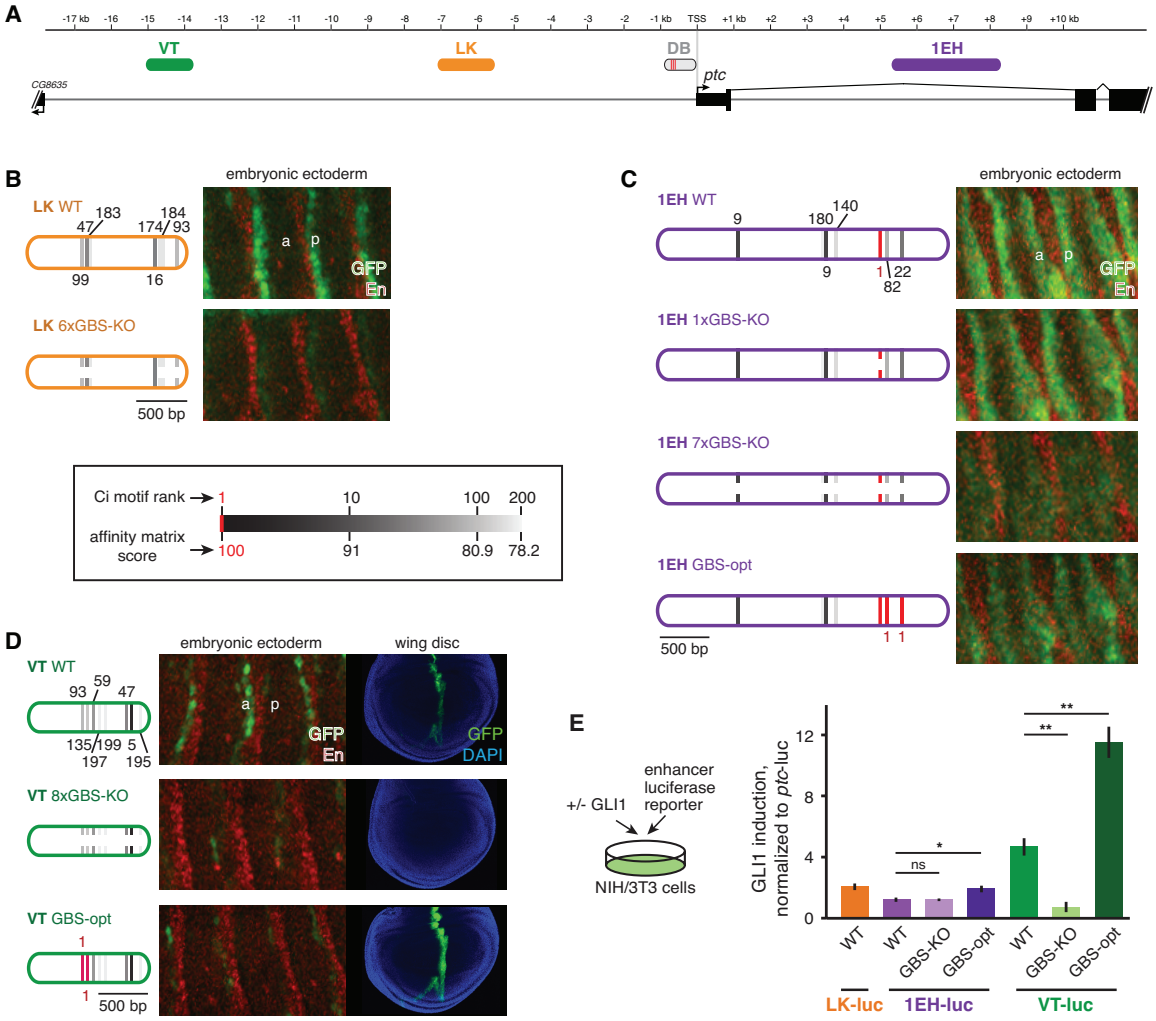


Figure 3.3 – Embryonic *ptc* enhancers require low-affinity GBSs to respond to Hh signaling.

(A) The *ptc* locus showing the location of embryo enhancers displayed as colored cigars. (B-C) Diagrams of the LK[wt] and LK[6xGBS-KO] enhancers showing GBSs as vertical lines; broken lines indicate mutated sites. On the right, dorsal sections from confocal images of embryos containing LK and LK[6xGBS-KO] visualized with GFP and En (which marks Hh-producing cells). Below is the color-scale used to visualize the predicted GBS based on Gli motif rank which was determined from an affinity matrix score (Hallikas et al., 2006). (D-G) Diagrams of the 1EH[wt], 1EH[1xGBS-KO], 1EH[7xGBS-KO], and 1EH[GBS-opt] enhancers showing GBSs as vertical lines; broken lines indicate mutated sites. On the right, dorsal sections from confocal images of embryos containing 1EH[wt], 1EH[1xGBS-KO], 1EH[7xGBS-KO], and 1EH[GBS-opt] visualized with GFP and En. (H-J) Diagrams of the VT [wt], VT [8xGBS-KO], and VT[GBS-opt] enhancers showing GBSs as vertical lines; broken lines indicate mutated sites. On the right, dorsal sections from confocal images of embryos containing VT [wt], VT [8xGBS-KO], and VT[GBS-opt] visualized with GFP and En. (H'-J') Confocal images of third-instar larval wing imaginal discs, showing GFP expression driven by VT [wt], VT [8xGBS-KO], and VT[GBS-opt] (K) Diagram of the experimental conditions used in the luciferase assay. (L) Luciferase assays using LK, 1EH, and VT in the presence of GLI1 and upon mutation or affinity upgrade of the GBSs.

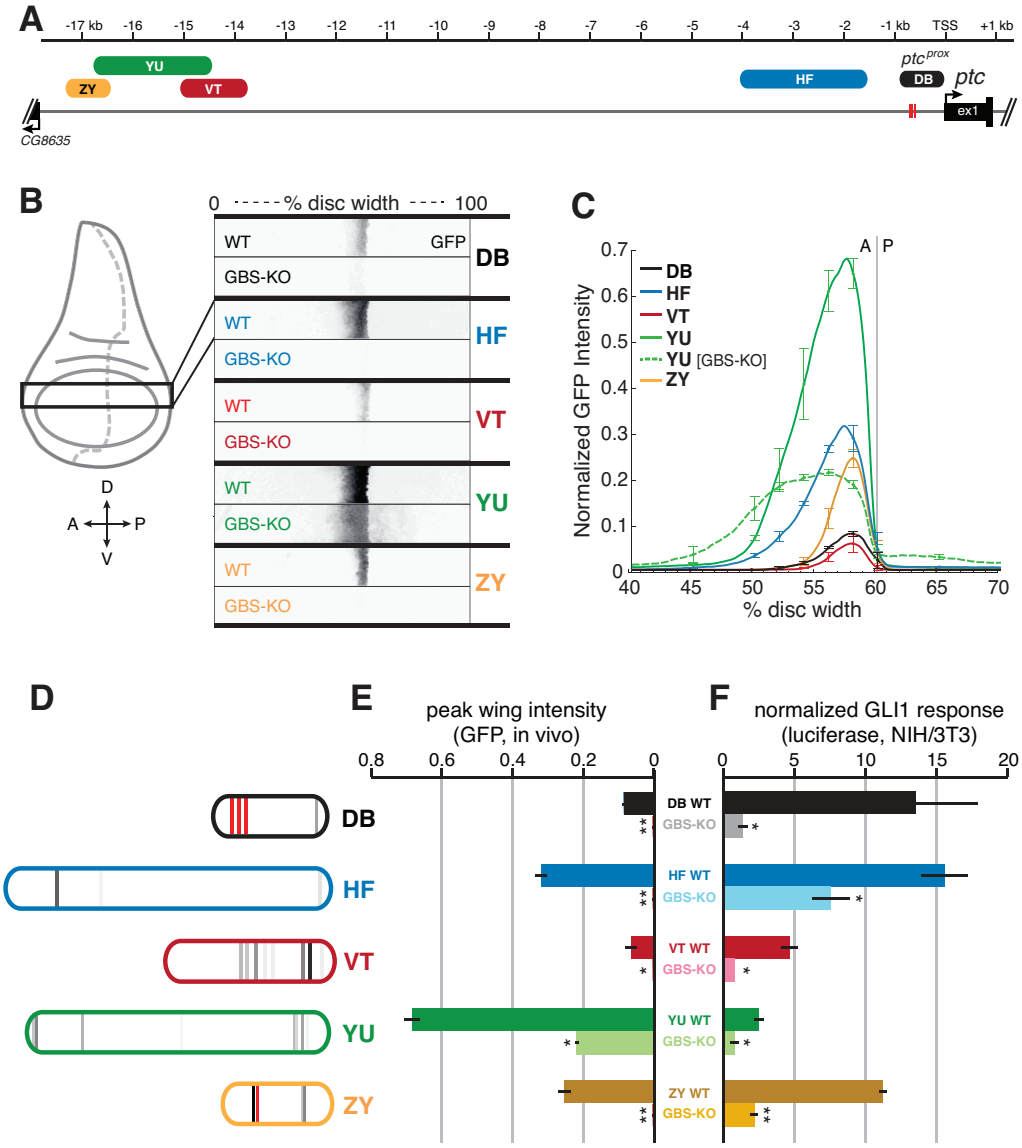


Figure 3.4 – Promiscuous wing stripes: multiple wing enhancers in the *ptc* locus.

(A) The *ptc* locus showing the location of wing enhancers displayed as colored cigars. (B) On the left, cartoon of a wing imaginal disc highlighting the dorsal pouch region used for GFP quantification. On the right, confocal images of dorsal wing pouch segments of wing discs showing GFP expression in black. (C) Normalized GFP fluorescence data collected from wing discs. Error bars indicate 1 s.d. (D) Diagrams of the DB, HF, VT, YU and ZY enhancers showing GBSs as vertical lines. (E) Peak wing intensity values for each wild-type and GBS mutant enhancer. (F) Luciferase assays using each enhancer in the presence of GLI1 and upon mutation of the GBSs. A, anterior; P, posterior, D, dorsal; V, ventral.

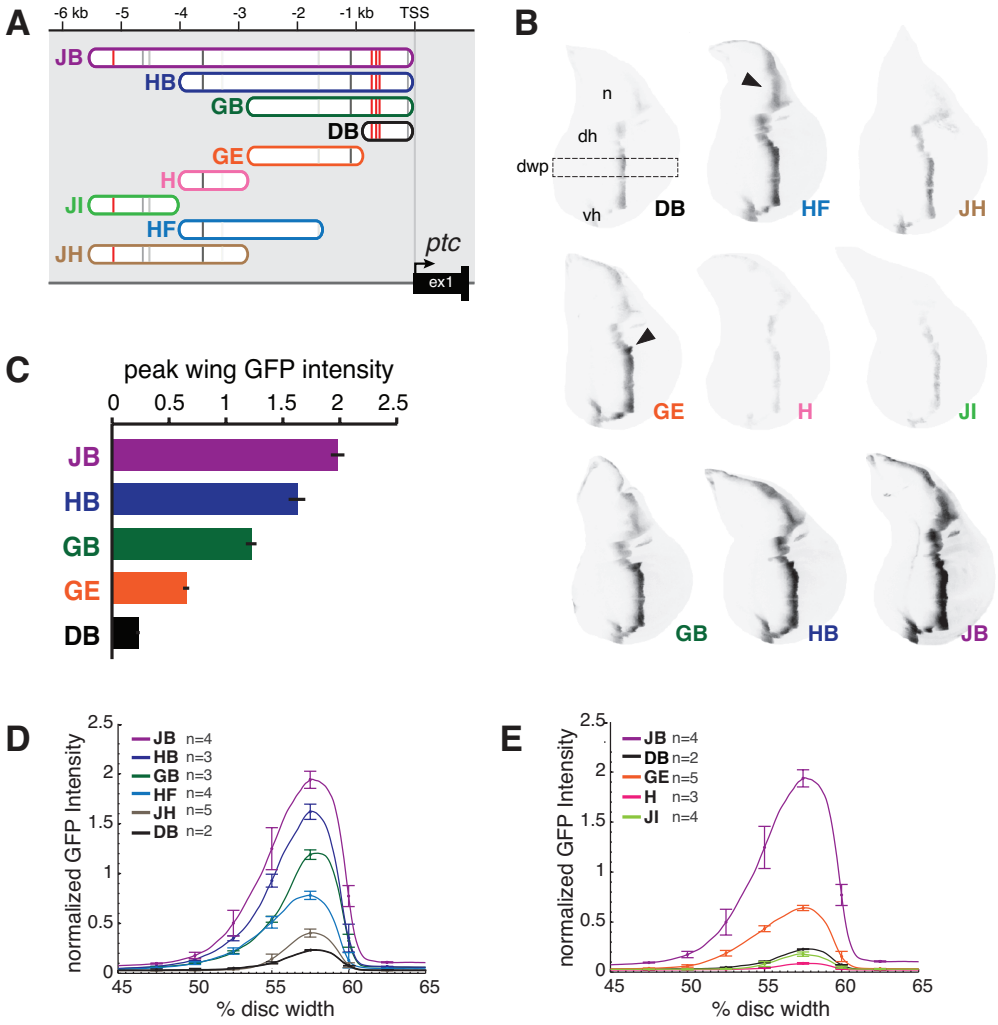


Figure 3.5 – Detailed analysis of the *ptc* proximal regulatory sequence.

(A) The *ptc* locus showing the location of wing enhancers displayed as colored cigars with predicted GBSs shown as vertical bars. (B) Representative confocal images of third-instar wing imaginal discs carrying the enhancer constructs shown in A driving GFP. Arrowheads indicate regions of the wing where expression is highly variable among the enhancers. (C) Peak wing intensity values. (D-E) Normalized GFP fluorescence data collected from wing discs. Error bars indicate 1 s.d. N, notum; dwp, dorsal wing pouch; dh, dorsal hinge; v, ventral hinge.

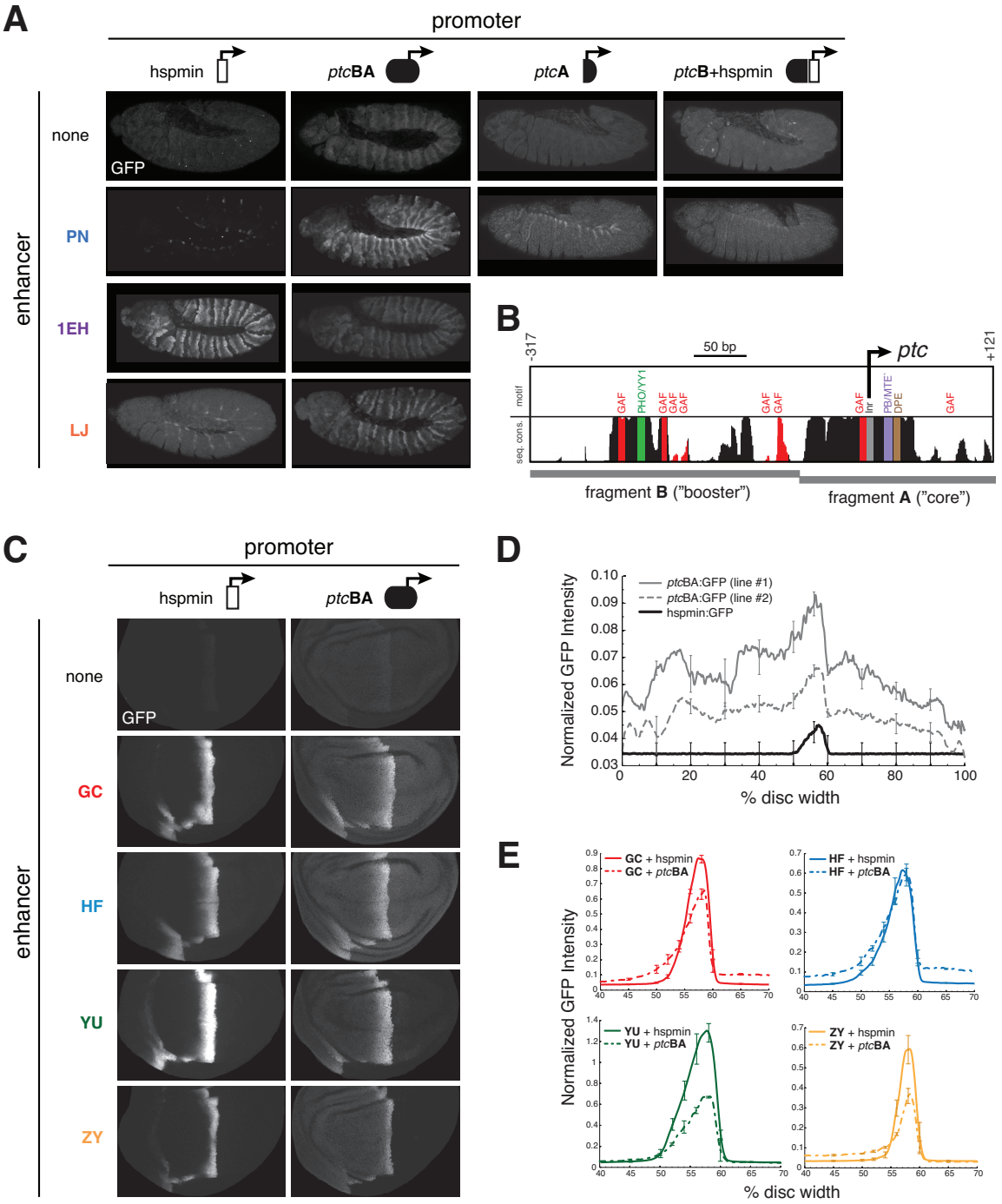


Figure 3.6 – Enhancer-promoter specificity in the *ptc* locus.

(A) Transgenic stage 11-12 embryos containing different promoters (y-axis, top) with PN, 1EH, LJ or no enhancer driving GFP, gray is GFP. (B) Diagram of the *ptc* promoter region. (C) Confocal images of third-instar wing imaginal discs containing the hspmin promoter (left column) or the *ptc*BA promoter (right column) with GC, HF, YU, ZY or no enhancer driving GFP. (D) Normalized GFP fluorescence data collected from wing discs containing the *ptc*BA or the hspmin promoters alone. Error bars indicate 1 s.d. (E) Normalized GFP fluorescence data collected from wing discs. Error bars indicate 1 s.d.

Table 3.1 – PCR primers used to amplify the *ptc* enhancers.

Primer	Sequence
BA_F	CACCCCGCGGAACAAACTTGGTATTTTC
BA_R	GTAGGATCCGTTCTGTGATATCTATCTTGT
GC_F	CACCGGCGCGCCCTATGGGCCAATGACAAATG
GC_R	GTACCGCGGGGGTGATTCAGGAGTTTTTCC
HF_F	CACCGGCGCGCCTCCCACTTCATAACCCTC
HF_R	GTACCGCGGCGCTCTCTCTTTCGGGGAGAA
JH_F	CACCGGCGCGCCTACGTACTCTTATTACTCCACTC
JH_R	GTACCGCGGGCTATTGCATTTGTCATTGGC
LJ_F	CACCGGCGCGCCCTACTTGGTTTGATAAAT
LK_R	GTACCGCGGACGGTGTGTGTGAACCCAACTAATTG
LJ_R	GTACCGCGGATAAGTACAGTGCTGGTCATA
ML_F	CACCGGCGCGCCCGAGGCGAGATGGCTTCG
ML_R	GTACCGCGGGCAGCGACTGCTGAGCCAGTC
NM_F	CACCGGCGCGCCTACAATATCTATTATCTA
NM_R	GTACCGCGGAGTTCCCATTCAGCTTTGACA
PN_F	CACCGGCGCGCCCGCCAAAATGAAAATTATTACAAAG
PN_R	GTACCGCGGCGTCCCTCTGTCGTCGCTGGG
RP_F	CACCGGCGCGCCAGCATTAACAGCCGAAGC
RP_R	GTACCGCGGACCGAGCCGTACAAATATAAC
VR_F	CACCGGCGCGCCACAGACGGGGCTACACTGAG
VR_R	GTACCGCGGATTCGGAGACATTCGCAGAC
YU_F	CACCGGCGCGCCGCCCTGTCGTCTTTGTCTTC
YU5'_R	GTACCGCGGGTATTAATAGTGGGAGCTCTG
YUb_R	GTACCGCGGAAATTTGCATAATAATAAGTC
YUb_F	CACCGGCGCGCCTACAATTTTGGAAAAGG
YU3'_F	CACCGGCGCGCCTCGCTGCGTGTTGCTGTG
YU_R	GTACCGCGGAACCATCAAACCCACGAAAC
ZY_F	CACCGGCGCGCCCCGGATCGACCTAGGTAAGG
ZY_R	GTACCGCGGTTCTTCTGGGTGTGTGCTTG
DL2_F	CACCGGCGCGCCCCGCATACCCTATGATATGACTC
DL2_R	GTACCGCGGGACAGACCCCGAAGACTGAG
DB_F	CACCGGCGCGCCATGCATGCGCAGCCTGCCAC
DB_R	GTACCGCGGCGCGCGCCGCTGTTTCTC
1AC_F	CACCGGCGCGCCGGTGAGTGCCCAACTACAG
1AC_R	GTACCGCGGGTTGGGCGCCATATGTTTAC
DL3_F	CACCGGCGCGCCTCCGTCCAGCAGTCGAAAG
DL3_R	GTACCGCGGCCCATTCGATATACCCTCAAG
1CE_F	CACCGGCGCGCCTCGGTTGTAACGTGCTTTTG
1CE_R	GTACCGCGGTGCAGATGGCAGATCAAGTC
1GH_F	CACCGGCGCGCCGAAGTGCTTAACAAGTTAAC
1EH_R	GTACCGCGGCACGACAACCAATGAGATCG
DL5_F	CACCGGCGCGCCGGTAATAAATGCGGCAGACG
1HI_F	CACCGGCGCGCCGATTGATTGATGCGTGATGC
1HI_R	GTACCGCGGCTGAAAAATGCAACAAAATA

3.7 References

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.

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

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

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 (Cambridge, England) *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 *phi*C31 integrases. Proc Natl Acad Sci U S A *104*, 3312–3317.

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*, 418–431.

Butler, J.E.F., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & Development *16*, 2583–2592.

Campos-Ortega, J.A., and Hartenstein, V. (1985). The Embryonic Development of Drosophila Melanogaster (Springer).

Capdevila, J., Estrada, M.P., Sánchez-Herrero, E., and Guerrero, I. (1994). The *Drosophila* segment polarity gene *patched* interacts with *decapentaplegic* in wing development. The EMBO Journal *13*, 71–82.

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.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science.

Domínguez, M., and Hafen, E. (1997). Hedgehog directly controls initiation and propagation of retinal differentiation in the Drosophila eye. Genes & Development *11*, 3254–3264.

Estella, C., Voutev, R., and Mann, R.S. (2012). A dynamic network of morphogens and transcription factors patterns the fly leg. Curr. Top. Dev. Biol. *98*, 173–198.

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.

Gradilla, A.-C., and Guerrero, I. (2013). Hedgehog on the move: a precise spatial control of Hedgehog dispersion shapes the gradient. Curr. Opin. Genet. Dev.

Guo, Y.-Y., Zhang, J.-Y., Li, X.-F., Luo, H.-Y., Chen, F., and Li, T.-J. (2013). PTCH1 gene mutations in Keratocystic odontogenic tumors: a study of 43 Chinese patients and a systematic review. PLoS ONE *8*, e77305.

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.

Herz, H.M., Hu, D., and Shilatifard, A. (2014). Enhancer Malfunction in Cancer. Molecular Cell.

Hidalgo, A., and Ingham, P. (1990). Cell patterning in the Drosophila segment: spatial regulation of the segment polarity gene patched. Development (Cambridge, England) *110*, 291–301.

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.

Hursh, D.A., Padgett, R.W., and Gelbart, W.M. (1993). Cross regulation of decapentaplegic and Ultrabithorax transcription in the embryonic visceral mesoderm of Drosophila. Development (Cambridge, England).

Ingham, P.W. (1993). Localized hedgehog activity controls spatial limits of wingless transcription in the Drosophila embryo. *366*, 560–562.

Johnson, R.L., Grenier, J.K., and Scott, M.P. (1995). patched overexpression alters wing disc size and pattern: transcriptional and post-transcriptional effects on hedgehog targets. Development (Cambridge, England) *121*, 4161–4170.

Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter - Kadonaga - 2011 - Wiley Interdisciplinary Reviews: Developmental Biology - Wiley Online Library. Wiley Interdiscip Rev Dev Biol *1*, 40–51.

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

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. *28*, 3290–3300.

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

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. Genes & Development *18*, 1606–1617.

Liu, Z., Xu, J., He, J., Zheng, Y., Li, H., Lu, Y., Qian, J., Lin, P., Weber, D.M., Yang, J., et al. (2014). A critical role of autocrine sonic hedgehog signaling in human CD138+ myeloma cell survival and drug resistance. Blood.

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

Marigo, V., and Tabin, C.J. (1996). Regulation of patched by sonic hedgehog in the developing neural tube. Proc Natl Acad Sci U S A *93*, 9346–9351.

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 (Cambridge, England) *139*, 2663–2669.

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

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 (Cambridge, England) *127*, 2999–3007.

Nakano, Y., Guerrero, I., Hidalgo, A., Taylor, A., Whittle, J.R.S., 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.

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

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.

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. Science Signaling *4*, ra38–ra38.

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. Philosophical Transactions of the Royal Society B: Biological Sciences *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.

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 (Cambridge, England) *140*, 4490–4498.

Sebo, Z.L., Lee, H.B., Peng, Y., and Guo, Y. (2014). A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. Fly *8*, 52–57.

Smith, E., and Shilatifard, A. (2014). Enhancer biology and enhanceropathies. Nat. Struct. Mol. Biol. *21*, 210–219.

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

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

Thomas, M.J., and Seto, E. (1999). Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *236*, 197–208.

Tzelepi, V., Karlou, M., Wen, S., Hoang, A., Logothetis, C., Troncoso, P., and Efstathiou, E. (2011). Expression of hedgehog pathway components in prostate carcinoma microenvironment: shifting the balance towards autocrine signalling. Histopathology *58*, 1037–1047.

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 (Cambridge, England) *134*, 1977–1989.

Wang, Y.V., Tang, H., and Gilmour, D.S. (2005). Identification in vivo of different rate-limiting steps associated with transcriptional activators in the presence and absence of a GAGA element. *25*, 3543–3552.

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.

Yant, S.R., Zhu, W., Millinoff, D., Slightom, J.L., Goodman, M., and Gumucio, D.L. (1995). High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster. *23*, 4353–4362.

Yáñez-Cuna, J.O., Kvon, E.Z., and Stark, A. (2013). Deciphering the transcriptional cis-regulatory code. Trends Genet. *29*, 11–22.

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

CHAPTER 4

Multivalent Gli Motifs Integrate Hedgehog and Non-Hedgehog Transcriptional Inputs Via Shared Binding Sites

4.1 Abstract

Hedgehog (Hh) signaling gradients control transcription of key target genes during development. The transcription factor Cubitus interruptus (Ci) binds to enhancers of Hh target genes, activating transcription in response to Hh signaling, and repressing gene expression in the absence of the signal. The goal of this study was to characterize the transcriptional output of different Ci binding site (CiBS) variants in order to better understand how Hh-regulated enhancers integrate signaling inputs *in vivo*. Taking a minimalistic approach using synthetic Hh-regulated enhancers with defined inputs, we have uncovered the promiscuous nature of some CiBSs *in vivo*. We observed, to our surprise, that consensus CiBSs strongly repressed transcription in cells where Ci is not expressed. These Ci-independent inputs also repressed transcription of endogenous Hh-regulated enhancers that contain specific Ci site variants. We identified several transcription factors (TFs) that could potentially bind Ci sites *in vivo*, but further experiments are needed to test if these are the culprits of the Ci-independent repression that we have observed. Multivalent Ci sites may have evolved to prevent the ectopic expression of specific Hh target genes. On the other hand, it is possible that other Ci site variants might be under selective

pressure to avoid usurper binding. Together, these hypotheses may account for the widespread use of CiBS variants in Hh-regulated enhancers across species.

4.2 Introduction

The combinatorial binding of transcription factors (TFs) to regulatory sequences in the genome controls unique patterns of gene expression to establish cell fates and body plans (Levine, 2010). These regulatory elements, also known as enhancers, integrate TF inputs to activate or repress the transcription of target genes (Spitz and Furlong, 2012). TFs are sometimes under the control of major signaling pathways like Hedgehog (Hh) (Barolo and Posakony, 2002). The TF Cubitus interruptus (Ci), which is a member of the Gli family of TFs, is modified by Hh signaling (Briscoe and Thérond, 2013). In the absence of the Hh signal, the activation domain of Ci is protealytically cleaved, and Ci becomes a transcriptional repressor that binds short DNA motifs to repress transcription. Upon signaling activation, full-length activator Ci binds the same sequences to promote transcription (Ingham et al., 2011). It is not fully understood how enhancers integrate these competing Hh-regulated Ci inputs to produce discrete transcriptional outputs. However, in some instances, the affinity and number of Ci binding sites (CiBSs) at a given enhancer have been shown to be key to determine whether or not transcription proceeds (Oosterveen et al., 2012; Parker et al., 2011).

Previous findings from our lab, have shown that the *dpp*D enhancer which is regulated by Hh in the developing wing requires three low-affinity non-

consensus CiBSs to respond to Hh signaling in a broader domain (Müller and Basler, 2000; Parker et al., 2011; Ramos and Barolo, 2013). When the affinity of those sites was improved by changing the sequence of the CiBSs to look like the optimal Ci motif, the expression of the upgraded enhancer was more restricted. From these data, we proposed that high Ci-occupancy combined with a supposedly more cooperative form of Ci repressor (Ci^{Rep}) were responsible for restricting the transcriptional outputs of the high-affinity enhancer to the wrong wing domain (these data are discussed in more detail in Chapter 2) (Parker et al., 2011; Ramos and Barolo, 2013). In other words, consensus high-affinity sites seem to be better at repression, than the non-consensus sites in *dpp*D.

The characterization of Hh/Ci-regulated enhancers like *dpp*D has been extremely important in understanding how these sequences integrate the Hh signal. Recently, a lot of effort has been focused in decoding the rules behind specific transcriptional responses to Hh signaling (Oosterveen et al., 2012; 2013; Parker et al., 2011; Peterson et al., 2012; Ramos and Barolo, 2013; Winklmayr et al., 2010). Learning these rules might be useful to predict expression patterns from Hh/Ci-regulated enhancers and to identify new target genes from genomic sequences. However, endogenous enhancers integrate many unknown Hh-independent inputs, and this complicates the analysis of the contribution of individual CiBSs.

Here, we used synthetic enhancers to examine distinct CiBS variants while controlling, as much as possible, the unknown inputs that bind to these

minimal regulatory sequences. Using this approach we uncovered that some predicted CiBSs, including the consensus sequence, are hijacked by unknown TFs *in vivo,* and we identified several candidates that bind selectively to these sites *in vitro.* These and other yet-to-be identified TFs may refine the expression of Hh target genes. In addition, these findings provide an alternative explanation to our previous studies of the *dpp*D (Chaper 2) (Parker et al., 2011). It is possible that consensus sites are better at repression because they integrate additional repressive Ci-independent inputs, which are unable to bind the non-consensus sites in *dpp*D. This contrasts with our original model that binding cooperativity between Ci^{Rep} was the cause of more restricted expression patterns with higher-affinity Ci sites (Parker et al., 2011; Ramos and Barolo, 2013).

4.3 Results

4.3.1 CiBSs integrate repressive inputs in Ci-negative cells

Synthetic enhancers containing six TF binding sites, three Grainyhead (Grh) sites and three CiBSs, were generated to study how specific CiBS variants interpret the Hh gradient (Figure 4.1A). As shown in Chapter 2, we have used this approach in the past to test transcriptional outputs of isolated CiBSs from known Hh-target enhancers (Parker et al., 2011; Ramos and Barolo, 2013). Isolated CiBSs alone are insufficient to activate transcription; hence, we added Grh binding sites to our constructs as additional activator inputs. Grh sites are sufficient to activate gene transcription in the wing in the absence of other inputs

(Furriols and Bray, 2001; Ramos and Barolo, 2013). Grh expression was used as a baseline to detect activating and repressive inputs from Ci sites, which can then be measured as changes in gene expression in Grh+Ci reporters, relative to a Grh-alone reporter (Ramos and Barolo, 2013).

We generated three versions of synthetic enhancers with three Grh binding sites upstream of three high-affinity CiBSs (Ci #1), three low-affinity CiBSs (Ci #37, #172, #1512), and three mutant Ci sites (KO) (Figure 4.1B-D). All of these transgenic constructs drive Hh/Ci-regulated stripes of different width and strength, with the exception of the 3xGrh-only construct which drives basal levels of expression throughout the wing disc (Figure 4.1B). We quantified the normalized GFP fluorescence data from these synthetic reporters as described in Chapter 2 (Ramos and Barolo, 2013). Surprisingly, we observed that the 3xGrh-3xCi#1-GFP construct is unable to drive expression in posterior cells where 3xGrh-3xCiKO-GFP is weakly active in response to Grh inputs (Figure 4.1E). This region of the wing imaginal disc goes from about 60-100% disc width. On the other hand, the synthetic enhancer containing the lower-affinity sites was as active as the Grh-only control construct. These results show that 3xGrh-3xCi#1-GFP is repressing transcription in posterior cells, whereas 3xGrh-Ci#37,#172,#1512-GFP is not. This result is surprising because Ci is not expressed in posterior cells. Our observations suggest that other factors hijack CiBSs to repress transcription in the absence of Ci.

4.3.2 **Selective binding of Ci-independent inputs to CiBSs**

We originally developed these synthetic Hh-regulated enhancers to better understand how the *ptc* and *dpp*D enhancers integrated the Hh gradient, and for that reason we kept the flanking sequence of the 9-mers the same as the flanking sequence in those enhancers (Alexandre et al., 1996; Müller and Basler, 2000). To have a cleaner system with less sequence variables, we re-designed the synthetic constructs so that all the 9-mers contain identical flanking sequence. We chose the flanking sequence from the first CiBS in the *ptc* enhancer because this site has the best *in vitro* binding affinity (Parker et al., 2011).

The new synthetic versions (Figure 4.1G-I) contained the same configuration of sites as the enhancers in Figure 4.1B-D. These enhancers produce similar transcriptional outputs as the original versions. 3xGrh-3xCiKO-GFP drives expression ubiquitously, while 3xGrh-3xCi#1-GFP and 3xGrh-3xCi#37 drive stripes of expression along the anterior/posterior (A/P) boundary (Figure 4.1G-I). Interestingly, CiBS #37 drives expression at similar levels as CiBS #1, although expression occurs in more anteriorly located cells. More importantly, 3xGrh-3xCi#1-GFP drives posterior repression, whereas 3xGrh-3xCi#37-GFPdoes not (Figure 4.1J-K). Together, these data pointed for the first time to the integration of repressive inputs other than Ci via consensus high-affinity CiBSs *in vivo*.

4.3.3 **Real enhancers integrate repressive inputs through CiBSs**

To determine if the posterior repression observed in the synthetic enhancers was real, and not an artifact of our minimal system, we tested the ability of isolated CiBSs from the *ptc* and *dpp*D enhancers to repress a *bona fide* wing enhancer. We chose the *vestigial* quadrant enhancer (*vg*QE) as a baseline because it drives strong expression in both anterior and posterior cells in the wing pouch, and it is not directly regulated by Hh signaling (Kim et al., 1996). We placed three CiBSs (as in the synthetic enhancers) downstream of *vg*QE (Figure 4.2A) to make three versions of this enhancer.

First, to keep the distance between *vg*QE and the promoter, and to test that the flanking sequence of the CiBSs did not affect the endogenous pattern of *vg*QE, we tested a version of *vg*QE with three mutated CiBSs (Figure 4.2B). *vg*QE-CiKO-GFP drives expression in a pattern that is indistinguishable form wild-type *vg*QE (data not shown). *vg*QE-3xCi#1-GFP drives transcription along the A/P boundary and repression in anterior cells (Figure 4.2C). This response is characteristic of Hh/Ci-regulated enhancers. Strikingly, *vg*QE-3xCi#1-GFP also represses expression in posterior cells that lack Ci (Figure 4.2C, white arrow). Lastly, we tested the effect of low-affinity sites (Ci#37, 172,1512). These non-consensus sites drive expected expression patterns in anterior cells and no posterior repression (Figure 4.2D). These data show that consensus CiBSs integrate repressive inputs in posterior cells, and these inputs are sufficient to repress the activity of a *bona fide* wing enhancer.

4.3.4 **Ubiquitous Ci-independent inputs bind CiBSs in the wing**

To determine if the repression seen in posterior cells was Ci-mediated, we knocked down the expression of Ci in the dorsal part of the wing imaginal disc

using the *ap*-Gal4 driver combined with a UAS-Ci RNAi responder. The expression of the *vg*QE reporters was examined, and we used the ventral part, which had wild-type levels of Ci, as an internal control to detect changes in gene expression (Figure 4.2E-G). *vg*QE-CiKO-GFP drives a different expression pattern when Ci is knocked down (Figure 4.2E). *vg*QE has been shown to be directly activated by *dpp* signaling in the wing (Certel et al., 2000), and since *dpp* is a target of Hh signaling in this tissue, the loss of vgQE-3xCiKO-GFP activation is most likely an indirect effect of the Ci knockdown.

If Ci mediates posterior repression, then knocking down Ci expression should result in de-activation in cells at the A/P boundary and de-repression in both anterior and posterior cells. *vg*QE-3xCi#1-GFP drives almost no expression across the wing when Ci is knocked down (Figure 4.2F). This result suggests that in the absence of Ci, other TFs bind consensus CiBSs and repress transcription throughout the wing disc. *vg*QE-Ci#37,#172, #1512-GFP drives expected expression patterns in cells with lower levels of Ci which is characterized by de-repression in more anterior cells, de-activation in cells abutting the A/P boundary, and no change in gene transcription posterior cells (Figure 4.2G). Together, our data opens up the possibility that TFs other than Ci occupy consensus CiBSs to repress and maybe sharpen the expression patterns of target genes.

4.3.5 **Real Hh-regulated enhancers contain CiBSs that integrate Ci-independent inputs to avoid ectopic expression**

Up to this point, we have shown that synthetic enhancers and endogenous non-Hh-regulated enhancers integrate Ci-independent inputs when isolated binding sites are added, but can real Hh/Ci-regulated enhancers integrate these inputs as well? We found that two newly identified enhancers in the *ptc* locus, YU and ZY, are de-repressed in posterior cells when all the predicted CiBSs are

mutated. YU is de-repressed in anterior and posterior cells of the wing disc when all the predicted sites were destroyed (Chapter 3) and ZY was de-repressed in the distal region of the leg imaginal disc (Figure 4.3B-C). Interestingly, YU only contains non-consensus sites, whereas ZY has three non-consensus (Ci #2, #107, #43) and one consensus CiBS. A unique feature of these enhancers is that they contain binding sites for tissue-specific activators that bind to sequences outside of the predicted CiBSs. Together, these results show that predicted CiBSs in real Hh-regulated enhancers are hijacked by other factors *in vivo* through both consensus and non-consensus sites. This activity may happen to prevent the ectopic expression of these enhancers.

To better understand these "usurper" inputs, we made additional mutations to ZY and examined the expression of these constructs in the leg imaginal disc, where Hh signaling establishes a gradient similar to the one found in the wing (Figure 4.3D-G). We mutated the first pair of predicted evolutionarily conserved CiBSs which contains a Ci site #2 and a #1 site in close proximity (Figure 4.3H, sites highlighted in yellow and red), and then we mutated the second pair of sites which contains a poorly conserved Ci site #107 and a #43 (Figure 4.3H, sites highlighted in blue and green). To test the contribution of the last two CiBSs in ZY, we mutated the two sites with better-predicted affinity. ZY-[#2,#1CiKO]-GFP drives very strong repression and decent activation in anterior cells, but weaker repression in posterior cells (Figure 4.3D). The final expression pattern of this construct hardly resembles a stripe. On the other hand, ZY-[#107,#43CiKO]-GFP is sufficient to make a stripe-like pattern by integrating repressive inputs in anterior and posterior cells (Figure 4.3E). Together, these results show that ZY requires all the predicted CiBSs to make the final wing stripe both proximally and distally, and that other inputs beside Ci regulate the expression of this enhancer in the leg via CiBSs to refine the Hh response.

To determine which of the two best CiBSs in ZY is responsible for posterior repression, we made versions of ZY that contained only Ci site #2 or the optimal site #1 (Figure 4.3F-G). Site # 2 (ZY-[#1,#107,#43CiKO]-GFP) represses expression in posterior distal cells, but surprisingly does not completely repress in anterior cells (Figure 4.3F). Site #1 (ZY-[#2,#107,#43CiKO]-GFP) represses expression in both leg compartments (Figure 4.3G). However, site #2 drives better posterior repression than site #1, yet complete repression is achieved when both sites are present (Figure 4.3E). Together, these results show that predicted CiBSs other than the consensus motif integrate inputs from other TFs to produce sharp expression patterns in response to Hh signaling in.

4.3.6 **Potential CiBS usurpers**

To find TF candidates that might recognize predicted CiBSs, we used the TOMTOM motif analysis tool (Bailey et al., 2009; Zhu et al., 2011). We identified several zinc finger DNA binding proteins that belong to the Gli/Zic, EGR and KLF/SP families (Enuameh et al., 2013). These factors are Lameduck (Lmd), Sugarbabe (Sug), Odd paired (Opa), Klumpfuss (Klu), Stripe (Sr), Buttonhead (Btd) and *Drosophila* Sp1 (DSp1) (Figure 4.4A). Out of the seven candidates, Opa, Klu, Sr, Btd and DSp1 are known to be expressed in larval tissues (Ghazi et al., 2003; Klein and Campos-Ortega, 1997; Lee et al., 2007; Mann and Carroll, 2002). Lmd and Sug are expressed in the embryonic mesoderm, but their expression in larval tissues is unknown (Duan et al., 2001; Zinke et al., 2002). We found the predicted binding motifs of these candidates in Fly Factor Survey (Zhu et al., 2011), a database that contains the DNA binding specificities of many *Drosophila* TFs, and we aligned these logos to the binding motif of Ci (Figure

4.4A). This alignment highlights position 5 of the Ci 9-mer, because most of the candidates prefer to bind a C instead of an A (Figure 4.4A). Interestingly, this position is not contacted by Ci/Gli directly (Pavletich and Pabo, 1993). Although we found several promising candidates, we cannot rule out the possibility that other yet to be characterized TFs may recognize CiBSs.

To determine the sequence similarity of the zinc finger domains (ZFDs) of these factors, we aligned these residues using ClustalW (Figure 4.4B) (Larkin et al., 2007). As shown by the crystal structure of a Gli-DNA complex, fingers 4 and 5 make extensive base contacts within the 9-mer (Figure 4.4A, shown as arrowheads), while fingers 2 and 3 mainly contact the phosphate backbone (Figure 4.4A, indicated as black circles). Finger 1 mainly establishes protein–protein contacts with finger 2 (Pavletich and Pabo, 1993). Finger 5 contacts bases 1-4 in the 9-mer, while finger 4 contacts bases 6-9 and the first base in the downstream flanking sequence (Pavletich and Pabo, 1993). Fingers 2, 3 and 4 are alignable in all the proteins, but only TFs belonging to the Gli/Zic families have alignable fingers 1 and 5. This observation is in agreement with the logo alignment in Figure 4.4A because most of the binding motifs align with bases 4-9 of the Ci 9-mer. These observations show that partial overlap of the binding motifs of the candidates may account for usurper binding, and also suggest that the 3' end of the Ci motif provides specificity.

4.3.7 **Usurper candidates selectively bind CiBSs *in vitro***

The zinc finger regions of Lmd, Sug, Opa, Klu and Ci were expressed *in vitro* and the products were tested for binding using six different oligonucleotides (Table 4.1) that included different Ci sites (Figure 4.4C-H). We were not able to make the ZFDs of Sr, Btd and DSp1. First, in addition to Ci, Lmd, Opa and Klu

bind to an oligonucleotide containing the consensus CiBS (lanes 1). Sug is the only candidate that does not bind to the consensus sequence (Figure 4.4E, lane 1). Because of our observation that most of the candidates prefer a C instead of an A at position 5 (Figure 4.4A), we tested Ci site #5 (lanes 2). All the ZFDs bind to that oligonucleotide except for Opa. In lanes 3 and 4, oligonucleotides with a G or a T instead of an A were tested. Only Ci, Opa and Klu recognize a Ci site with a G at position 5 (lanes 3), and Ci, Lmd, Opa and Klu bind to a site with a T at that position (lanes 4). Next we tested Ci #37, which does not mediate repression independently of Ci *in vivo* (Figure 4.2G). None of the candidates bind to this oligonucleotide, only Ci does (lanes 5). This result was encouraging as we were looking for TFs that did not bind to site #37. Finally we tested an oligonucleotide that contains the consensus site for Opa and Sug, which happens to be a Ci #146, and has a C at position 5 and a G at position 9. All of the ZFDs bind this site except for Klu (lanes 6). In summary, all of the candidates except for Sug bind to consensus Ci sites, but not to site #37 *in vitro* (see Table 4.2 for a summary of these data). These results show that most of the usurper candidates bind selectively to Ci sites, and implicate their potential competition with Ci for binding to these CiBSs *in vivo*.

4.3.8 **CiBS #5 drives more expression than the consensus motif *in vivo***

We characterized the relationship between motif variant #5 (GACC<u>c</u>CCCA) and gene expression using the previously described synthetic enhancer approach. We tested Ci site #5 because it was bound by many of the usurper candidates *in vitro*, and because this position varied in the logo alignment in Figure 4A. This site should have lower affinity for Ci than the optimal motif (Hallikas et al., 2006). The transcriptional output driven by CiBS #5 was

compared to the output of the consensus site and KO (Figure 4.5A). As seen previously 3xGrh-3xCiKO-GFP drives ubiquitous expression in the wing, and 3xGrh-3xCi#1-GFP drives a stripe across the tissue (Figure 4.5B-C). Strikingly, 3xGrh-3xCi#5-GFP drives stronger expression than the synthetic enhancer carrying the optimal motifs. The difference between these constructs is only three base pairs. The stripe in Ci #5 is about 1.5 fold more active than Ci #1, but both CiBSs repress transcription at the same position in anterior cells (~46% disc width) (Figure 4.5E). In the posterior compartment, only Ci #1 drives repression, while Ci #5 behaves as Ci #37 and shows no repression (Figure 4.1F and K). These interesting results show for the first time that sub-optimal sites beat consensus high-affinity CiBSs in making a wing stripe. These results also show that the unidentified repressor does not bind site #5 (nor #37 as shown before). Another possibility is that an unknown activator binds site #5, and based on the *in vitro* data, this is a very likely possibility. Our results highlight the relevance of position 5 of the 9-mer, even if it is not directly contacted by the zinc fingers (Pavletich and Pabo, 1993).

We looked at the expression of these enhancers in the embryonic ectoderm to determine if the counterintuitive output of these sites in the wing was tissue-specific. The expression pattern of 3xCiKO-GFP in the embryonic epidermis recapitulates the expression pattern of Grh at this developmental stage (Figure 4.5G) (Hammonds et al., 2013). 3xGrh-3xCi#1-GFP drives expression on either side of the Hh/En-positive cells (Figure 4.5H), and interestingly, this pattern recapitulates the endogenous pattern of *ptc* (Chapter 3). As in the wing, 3xGrh-3xCi#5-GFP drives more expression than 3xGrh-3xCi#1-GFP (Figure 4.5I). However, the pattern driven by Ci #5 includes expression in Hh-positive cells where Hh target genes are not normally expressed. Taken together, these results show that sites with worse predicted affinity activate expression in Hh/Ci-

regulated patterns more strongly than consensus Ci sites, and these #5 sites do not bind repressive usurper inputs in the wing.

4.3.9 **Rogue CiBSs: when the best is not good enough**

We created a logo with the top ten Ci sites with the highest predicted affinity (Hallikas et al., 2006). This logo shows that positions 1, 3, 4, 6 and 9 are the most important because they remain constant in all the sites (Figure 4.6A). Then we created another logo with the ten most frequent Ci motifs in experimentally validated enhancers (including the newly identified *ptc* enhancers described in Chapter 2). This logo shows that positions 1, 3, 4, 5 and 6 are constant (Figure 4.6B). By comparing these logos we noticed that the 5' end of the 9-mer shows the same binding specificity in both sets of Ci sites, while the 3' shows dramatic variation (Figure 4.6A versus B). These findings are interesting because bases at the 3' end of the Ci motif are contacted by zinc finger 4, and we showed that this domain is present among all the usurper candidates (Figure 4.4B). On the other hand, bases at the 5' end of the motif are contacted by zinc finger 5, which is present only in the Gli and Zic factors, are identical in both logos (Figure 4.6A versus B). Interestingly, we found that the most frequent CiBS in *Drosophila* enhancers is not the optimal motif instead it is Ci #9, GACCACaCA (124 times versus 78). Ci site #5 is found only 4 times in our dataset. Together, these data suggest that CiBSs that deviate at the 3' end of the motif like Ci #9 may be under selective pressure to avoid binding of other TFs even if these sites have lower Ci affinity. This strategy potentially explains the prevalence of specific non-consensus CiBSs in Hh-regulated enhancers across species (Ramos and Barolo, 2013; Winklmayr et al., 2010). During evolution, it is probably easier to select for CiBSs variants that produce a more accurate Hh response, than to

adapt Ci to bind a slightly different sequence, although in this case a combination of both modifications is attractive.

4.4 Discussion

The counterintuitive transcriptional behavior of Hh-regulated enhancers with different binding affinities for Ci, prompted the characterization and quantification of transcriptional outputs from CiBS variants to better understand how these enhancers integrate signaling inputs *in vivo*. The best approach to answer this question is to study CiBSs in isolation. Using synthetic Hh-regulated enhancers with defined inputs, we have uncovered the promiscuity of some CiBSs *in vivo*.

Multivalent TF binding sites are not a unique feature of Ci. Many years ago Mike Levine and colleagues showed two examples of *Drosophila* enhancers that contain overlapping binding sites that are recognized by different TFs: one is the *eve* stripe 2 enhancer and the other is the *rhomboid* NEE enhancer (Ip et al., 1992; Small et al., 1991). In both cases, competition for binding occurs between activators and repressors. This regulatory logic is probably easier to identify because the transcriptional outputs of the enhancers are very distinct in the absence of either TF. However, if both TFs have the same transcriptional function, then the identification of multivalent sites becomes more challenging. Examples of TFs that compete for multivalent sites in mammals include Myc and Mad which are transcriptional antagonists that recognize identical binding motifs (James and Eisenman, 2002). A similar situation occurs between the TFs Blimp1, IRF-1 and IRF-2 during B-cell differentiation (Kuo and Calame, 2004). In *C. elegans*, the TF Ces-1 competes with bHLH factors to bind identical motifs (Reece-Hoyes et al., 2009). Competition for shared or partially overlapping

binding sites seems to be a common strategy in many species to produce tissue-specific responses. Competition between Ci^{Act}, Ci^{Rep} and usurper factors might be used to add more complexity to the Hh response.

In the past, we have grouped all non-consensus binding sites together when thinking about the transcriptional outputs of Hh-regulated enhancers. However, the findings presented in this chapter made us aware of the inherent complexity of each variant. Our data suggest that usurpers might not bind CiBSs that deviate at 3' end of the motif, particularly at positions 7 and 8. Mutations to these positions disrupt the string of Cs at positions 6, 7, and 8, which are preferred by usurpers as shown by *in vitro* binding data (Enuameh et al., 2013). Avoidance of CiBSs with "too many" Cs might explain why Ci site #5 is not frequently found in enhancers, although it drives more expression than Ci site #1. It is possible that Ci #5 integrates usurper activators. These hypotheses will be tested *in vivo* and *in vitro*, using a combination of *Drosophila* genetics, synthetic reporter transgenes and competitive EMSAs.

Finally, our results showing that Ci-independent inputs bind consensus CiBSs, but not sites from the *dpp*D enhancer to repress transcription in posterior cells (Figure 4.4), challenges previous claims by the Basler Lab that two types of Hh target genes exist: those that respond to Ci activation and repression like *dpp*; and those that only respond to Ci activation like *ptc* and *engrailed* (Méthot and Basler, 1999). These conclusions were posited after *dpp* was ectopically expressed in Ci mutant clones in posterior cells of the wing while *ptc* was not expressed (Méthot and Basler, 1999). Our observations suggest an alternative explanation where *ptc* is regulated by usurper factors that suppress its ectopic activation in posterior cells whereas *dpp* is not regulated by these factors.

We have shown that Ci-independent inputs regulate transcription through CiBSs. However, further sequence-function analyses are necessary to fully understand how these multivalent sites integrate these inputs in different contexts.

4.5 Materials and methods

DNA cloning

Synthetic enhancers were generated by assembly PCR (Swanson et al., 2010), and were sub-cloned into the pENTR/D-TOPO vector (Life Technologies). Using the Gateway system, the synthetic constructs were finally cloned into the pHPdest-eGFP vector (Boy et al., 2010). The ZY enhancer was amplified by standard PCR (Roche Expand High Fidelity PCR System) from genomic DNA, and cloned as the synthetic enhancers. Targeted GBS mutations were created by overlap extension PCR (Swanson et al., 2010).

Drosophila transgenesis

Site-directed transformation by embryo injection was performed as previously described (Bischof et al., 2007). Reporter transgenes were integrated into a *phi*C31 landing site at genomic position 86FB.

Drosophila stocks

The following stocks were used to knockdown the expression of Ci in the wing: ap-Gal4 [y^1 w^{1118}; P{GawB}ap^{md544}/CyO] (kindly provided by Scott Pletcher) and UAS-Ci RNAi [y^1 v^1; P{TRiP.JF01715}attP2] (BDSC Stock # 28984).

Immunohistochemistry and microscopy

Embryos were fixed and stained using standard methods as previously described (White et al., 2012). Embryos were staged as described (Campos-Ortega and Hartenstein, 1985). Third-instar wing imaginal discs were dissected and fixed as described (Parker et al., 2011). Confocal images were captured on an Olympus FluoView 500 Laser Scanning Confocal Microscope mounted on an Olympus IX-71 inverted microscope. Samples to be directly compared were fixed, prepared, and imaged under identical confocal microscopy conditions and settings. The primary antibodies used included rabbit anti-EGFP (Invitrogen), diluted 1:100, and mouse anti-En (Developmental Studies Hybridoma Bank), diluted 1:50. AlexaFluor488 and AlexaFluo568 conjugates with secondary antibodies from Invitrogen were used at 1:2000 dilutions.

Quantitation of transgenic reporter expression data

Wing confocal images were collected and quantified as previously described (Parker et al., 2011). The Matlab program *Icarus* (E. Ortiz-Soto, A.I.R., and S.B., manuscript in preparation) was used to process and plot wing imaginal disc fluorescence data. Each experiment was performed at least two times, and fluorescence was measured from at least two wings per construct.

Electromobility shift assays

Electromobility shift assays were performed as described (Parker et al., 2011). The zinc finger domains of Ci, Klu, Lmd, Sug, and Opa were made *in vitro*

using the TNT® SP6 High-Yield Wheat Germ Protein Expression System (Promega, Madison, WI.) according to the manufacturer's instructions. As a template, we used a PCR amplified fragment of the proteins that contains the DNA binding motifs. The primers used to generate the template introduce a SP6 promoter and Kozak sequence upstream, and poly-T sequence downstream of the protein. Oligonucleotides corresponding to six Ci site variants (see Table 4.1 for oligonucleotide sequences) were end-labeled with ATP, γ-^{32}P using T4 Polynucleotide Kinase, annealed, and used as probes in EMSA assays. 1.6nM radiolabeled probe were mixed and incubated on ice for 15 minutes with the ZFDs under the following conditions: 10mM Tris-HCl, pH7.5, 50mM NaCl, 10mM DTT, 1mM EDTA, 27.5 µg/ml salmon sperm DNA, 100 µg/ml poly(dIDC). Reactions were analyzed via electrophoresis on a 4% non-denaturing polyacrylamide gel in 0.5x TBE buffer.

Ci binding site prediction and ranking

A mono-nucleotide distribution matrix for Ci binding sites, derived from competitive DNA binding assays (Hallikas et al., 2006) was downloaded from the Genomatix Software Suite (www.genomatix.de; Genomatix, Germany). Matrix similarity scores (Quandt et al., 1995) were calculated using data from the first nine nucleotide positions of the Ci matrix, which contain the majority of the information content. Ci site rankings are determined by sorting all possible 9-mers in order of matrix similarity score, such that the optimal motif (GACCACCCA), with a score of 100, has a rank of 1. Nonamers with a lower matrix score than their reverse-complement sequences, such as TGGGTGGTC, are removed from the ranking so that each high-scoring site is included only once. Gli binding sites

were identified *in silico* by screening the *ptc* locus for defined motifs using Genepalette (Rebeiz and Posakony, 2004).

TOMTOM motif discovery tool

To identify other transcription factors that had similar DNA binding motifs as Ci, we entered the Ci consensus motif in the TOMTOM motif comparison tool and select the FlyFactorSurvey database (Bailey et al., 2009; Zhu et al., 2011).

Amino acid sequence alignment

Amino acid sequences were aligned using Clustal W (Larkin et al., 2007). Zinc finger amino acid sequences were obtained from previously published alignments and from Uniprot (Brown et al., 2005; Klein and Campos-Ortega, 1997; Sen et al., 2010; Shimeld; UniProt Consortium, 2010; Wimmer et al., 1996).

Logo generation

Logos (Schneider and Stephens, 1990) describing the binding preferences of different sets of CiBSs were generated with WebLogo (Crooks et al., 2004).

4.6 Acknowledgements

Thanks to Elliott Ortiz-Soto for creating Icarus, and special thanks to Victoria Blake, Rafael Rivera and David Lorberbaum for research support. Thanks to past members of the Barolo lab, Niki, Christina and Dave, for guidance and mentorship during the design and construction of the synthetic enhancers. We thank the staff at the Microscopy and Image Analysis Laboratory (MIL) at the

University of Michigan Medical School for their assistance obtaining confocal images. We thank the TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing transgenic RNAi fly stocks. This work was supported by the Cellular and Molecular Biology Training Grant (NIH T32-GM007315) and a Center for Organogenesis predoctoral fellowship (NIH T32-HD007505) to A.I.R. and by NIH grant GM076509 and NSF grant MCB-1157800 to S.B.

The text and figures presented in this chapter are part of the manuscript: Multivalent Gli Motifs Integrate Hedgehog and Non-Hedgehog Transcriptional Inputs Via Shared Binding Sites. Ramos, AI and Barolo, S. (Manuscript in preparation).

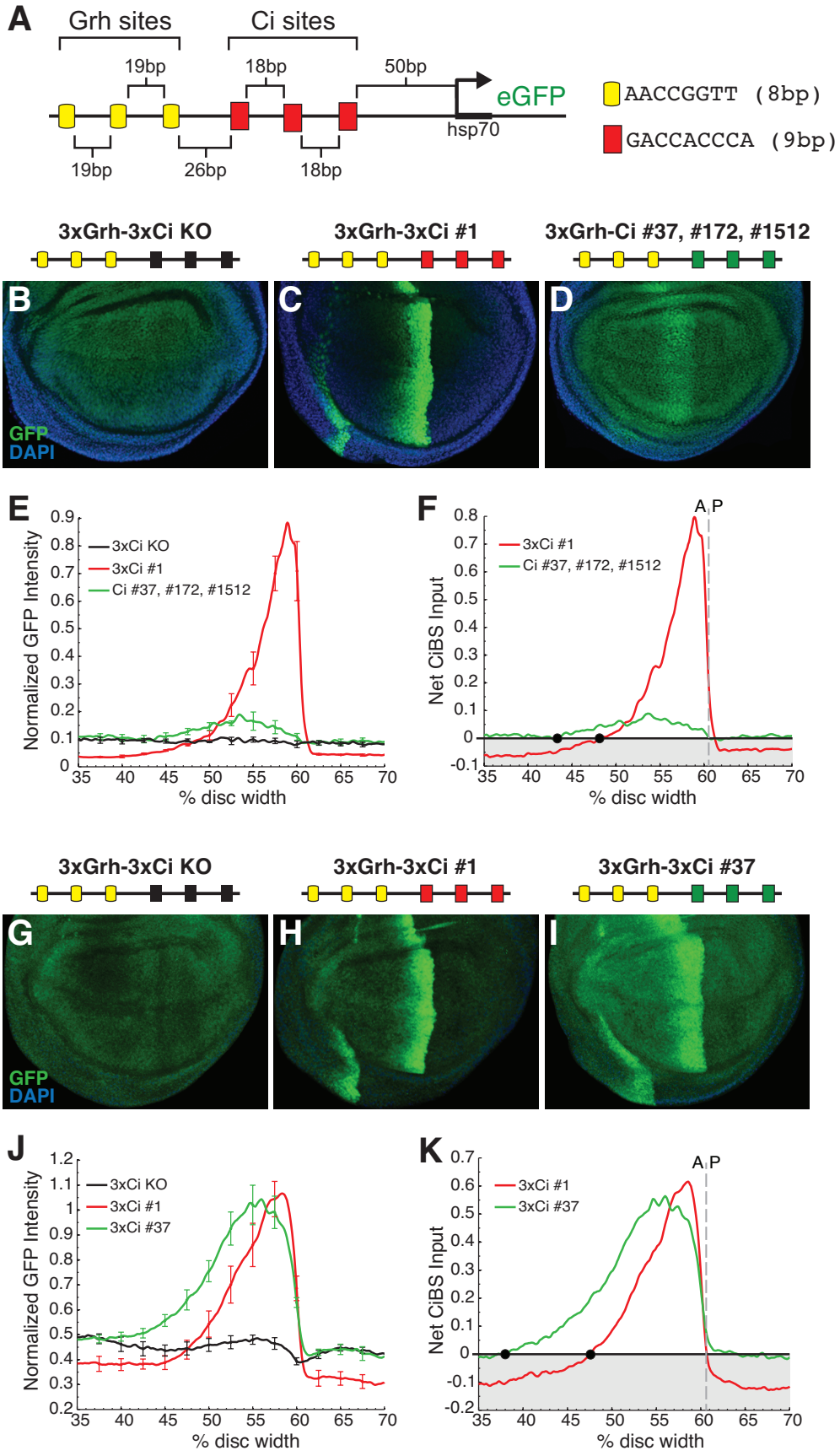

A
Grh sites
Ci sites
19bp
18bp
50bp
eGFP
hsp70
19bp
26bp
18bp
AACCGGTT (8bp)
GACCACCCA (9bp)
3xGrh-3xCi KO
3xGrh-3xCi #1
3xGrh-Ci #37, #172, #1512
B
C
D
GFP
DAPI
E
Normalized GFP Intensity
3xCi KO
3xCi #1
Ci #37, #172, #1512
% disc width
F
Net CiBS Input
A P
3xCi #1
Ci #37, #172, #1512
% disc width
3xGrh-3xCi KO
3xGrh-3xCi #1
3xGrh-3xCi #37
G
H
I
GFP
DAPI
J
Normalized GFP Intensity
3xCi KO
3xCi #1
3xCi #37
% disc width
K
Net CiBS Input
A P
3xCi #1
3xCi #37
% disc width

Figure 4.1 – Synthetic enhancers recapitulate endogenous patterns of gene expression in anterior cells, and reveal transcriptional repression in posterior cells via Ci sites.

(A) Diagram of the composition and organization of the synthetic enhancers with Grainyhead (Grh) sites and Ci sites indicated. (B-D) Representative confocal images of the pouch regions of wing imaginal discs from transgenic third-instar larvae are shown. The Ci sites used in each construct are indicated by ranking number (KO, `GACaAaaCA`; #1, `GACCACCCA`; #37, `GgCCACCtA`; #172, `GACCgCCCg`; #1512, `tACCtCCCc`). (E) Normalized GFP fluorescence data collected from the dorsal section of the wing pouch. Error bars indicate 1 s.d. (F) Net effect of Ci sites (calculated as the normalized transgene expression of 3xGrh-3xCi #1 or 3xGrh-3xCi #37,#172,#1512 minus normalized 3xGrh-3xCi KO expression). (G-I) Representative confocal images of the pouch regions of wing imaginal discs from transgenic third-instar larvae are shown. (E) Normalized GFP fluorescence data collected from the dorsal section of the wing pouch. Error bars indicate 1 s.d. (F) Net effect of Ci sites (calculated as the normalized transgene expression of 3xGrh-3xCi #1 or 3xGrh-3xCi #37 minus normalized 3xGrh-3xCi KO expression). Black circles indicate the positions on the A/P axis at which the synthetic reporter constructs switch from activation to repression. A, anterior; P, posterior.

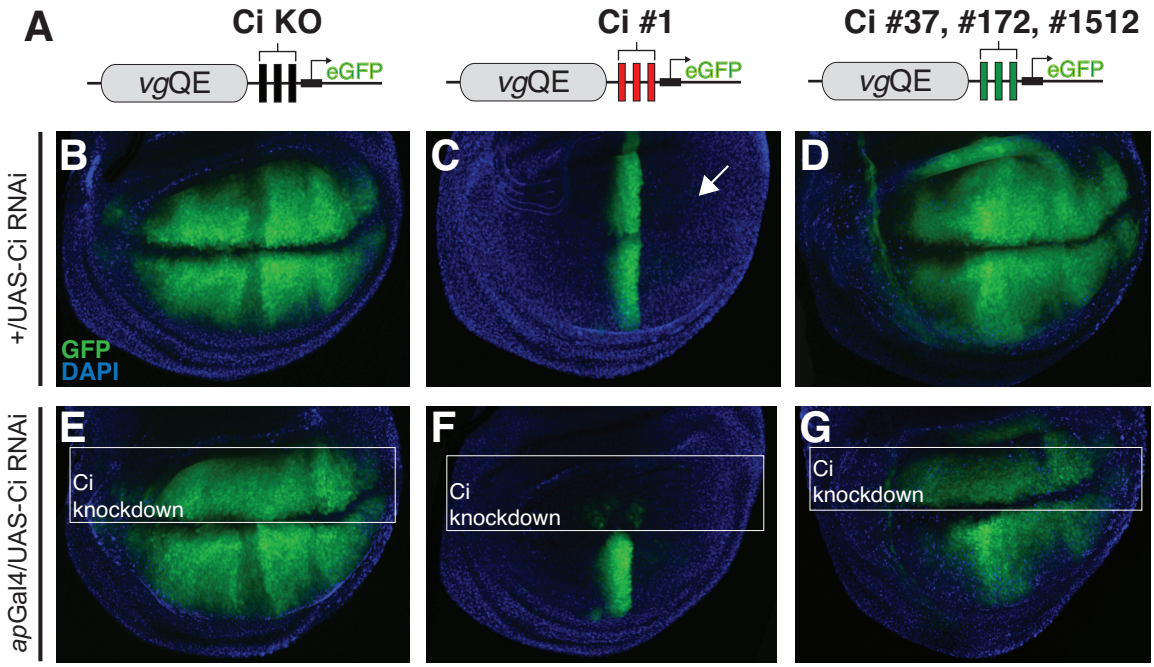


Figure 4.2 – Ci-independent repression of gene expression via optimal Ci sites.

(A) Diagram of the composition and organization of the synthetic enhancers containing the *vestigial* quadrant enhancer (*vg*QE) (Kim et al., 1996) with diverse Ci sites. (B-D) Representative confocal images of the pouch regions of wing imaginal discs from transgenic third-instar larvae are shown. The white arrow indicates posterior repression. The Ci sites used in each construct are indicated by ranking number (KO, `GACaAaaCA`; #1, `GACCACCCA`; #37, `GgCCACCtA`; #172, `GACCgCCCg`; #1512, `tACCtCCCc`). (E-G) Representative confocal images of wing imaginal discs where Ci expression was reduced in the dorsal region of the wing (indicated with white rectangles). The ventral region was used as a wild-type internal control to compare changes in reporter gene expression.

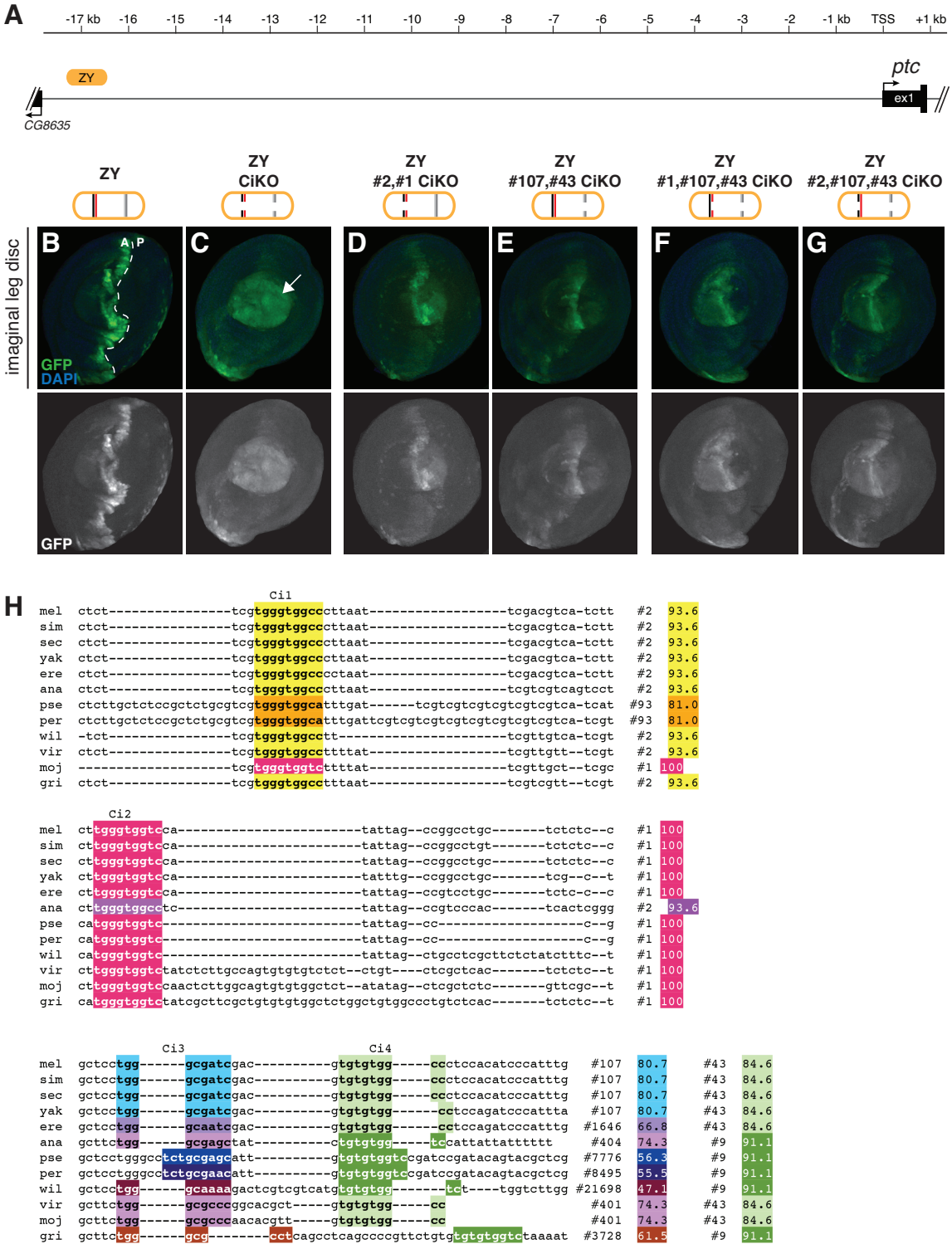


Figure 4.3 – Real Hh/Ci-regulated enhancers are posteriorly repressed via predicted Ci sites.

(A) Diagram of the *ptc* locus (as described in Chapter 3) showing the location of the ZY enhancer. (B-G) Representative confocal images of leg imaginal discs from transgenic third-instar larvae carrying modified versions of ZY are shown. The Ci sites modified in each construct are indicated by ranking number (KO, GACaAaaCA; #1, GACCACCCA; #2, GgCCACCCA; #107, GAtCgCCCA; #43, GgCCACaCA). (H)Twelve-species *Drosophila* sequence alignments are shown for selected regions of the ZY enhancer. Selected 9-mers are shaded, and Ci matrix similarity scores for those motifs are shown to the right. Sequences are from the following *Drosophila* species, from top to bottom: *D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, D. persimilis, D. willistoni, D. virilis, D. mojavensis*, and *D. grimshawi*. Dashes indicate gaps. A, anterior; P, posterior.

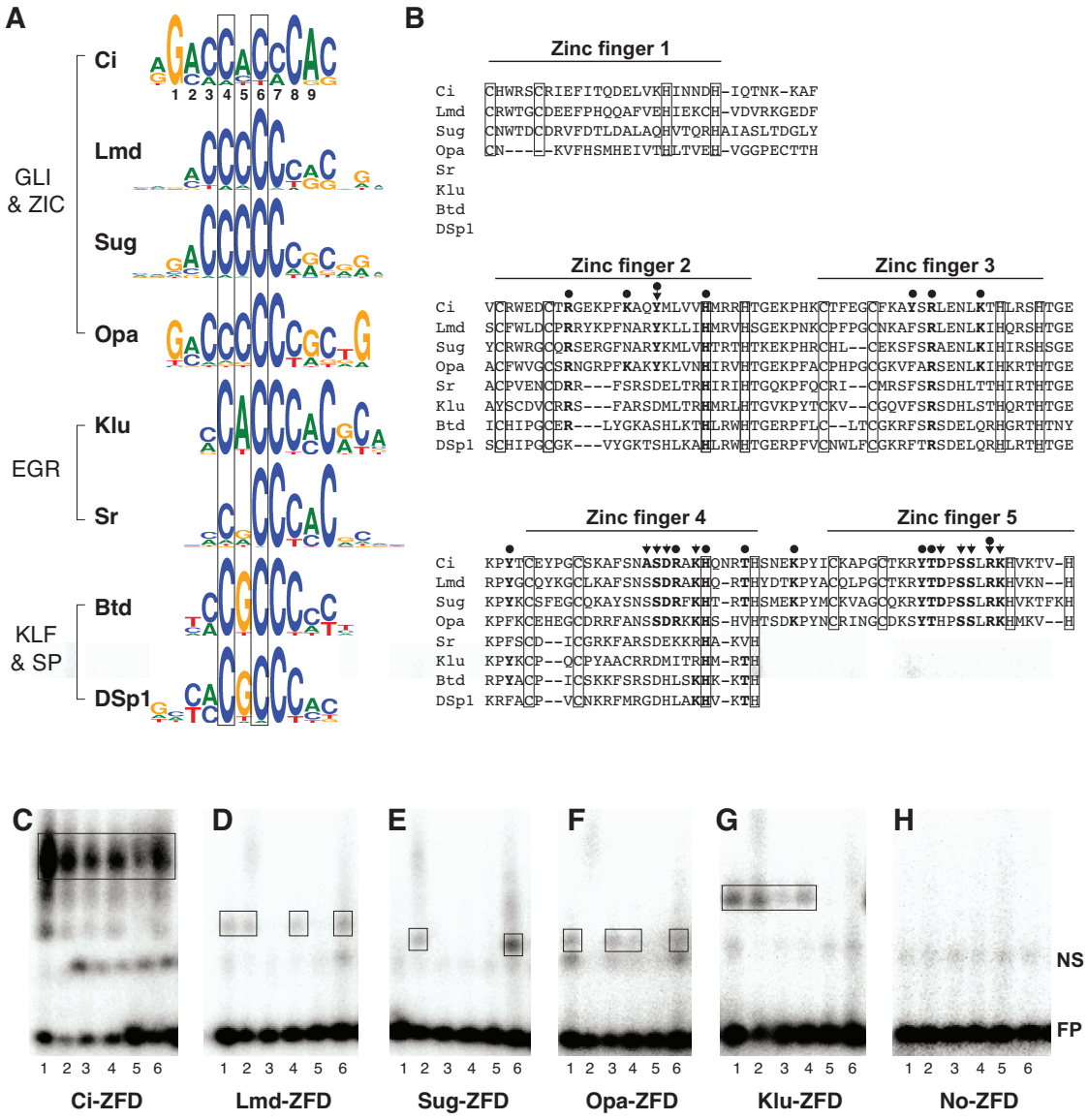


Figure 4.4 – TFs other than Ci bind to predicted CiBSs *in vitro*.

(A) Sequence logos comparing the binding preferences between Ci and the "usurper" candidates (Bailey et al., 2009; Zhu et al., 2011). (B) Amino acid sequence alignment of the zinc finger domains of the candidates and Ci. Residues that make DNA contacts are shown with arrowhead, and residues that make contacts phosphate contacts are shown with black circles (Pavletich and Pabo, 1993). (C-H) EMSAs using radioactively labeled probes (Table 4.1) incubated with the *in vitro* transcribed/translated zinc finger domains of Ci and several of the usurper candidates. Central boxes show specific ZFD-DNA complexes. NS, non-specific shift; FP, free probe.

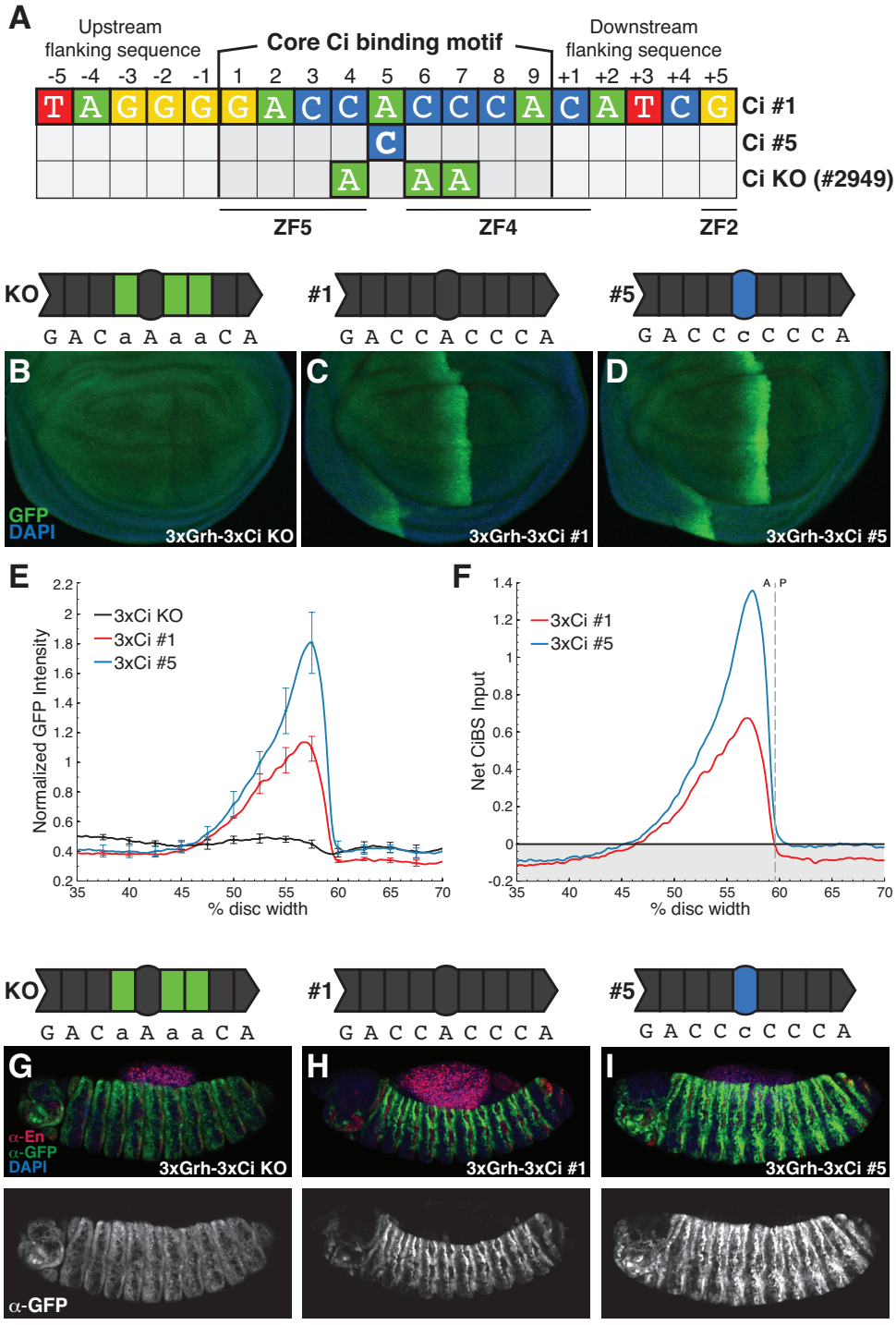


Figure 4.5 – CiBS variant #5 is better than the optimal motif *in vivo*.

(A) Sequence of the Ci sites used in synthetic reporter transgenes, including upstream and downstream flanking sequences. Divergent nucleotides are highlighted in colored boxes. (B-D) Representative confocal images of the pouch regions of wing imaginal discs from transgenic third-instar larvae are shown. The Ci sites used in each construct are indicated on top of the images. (E) Normalized GFP fluorescence data collected from the dorsal section of the wing pouch. Error bars indicate 1 s.d. (F) Net effect of Ci sites (calculated as the normalized transgene expression of 3xGrh-3xCi #1 or 3xGrh-3xCi #5 minus normalized 3xGrh-3xCi KO expression). (G-I) Confocal images of stage 14 transgenic embryos carrying 3xGrh-3xCi KO, 3xGrh-3xCi #1, and 3xGrh-3xCi #5 driving GFP. The Ci sites used in each construct are indicated on top of the images. A, anterior; P, posterior.

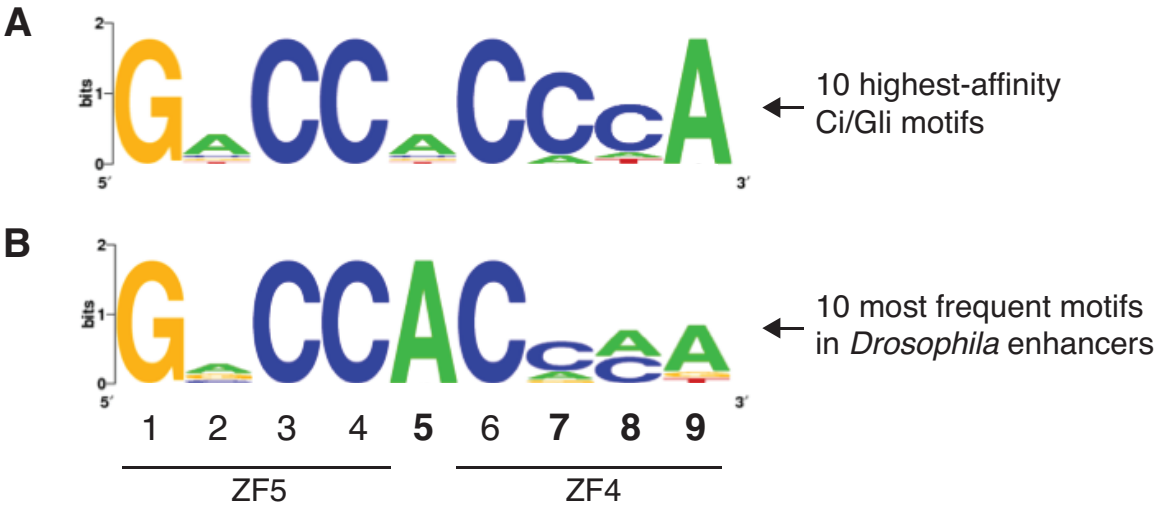


Figure 4.6 – Sequence binding preferences vary between Ci sites with highest predicted affinity and the most frequent motifs found in *Drosophila* enhancers.

(A) Logo showing the sequence binding preference of the top 10 highest-affinity Ci/Gli motifs (Hallikas et al., 2006). (B) Logo showing the sequence binding preference of the 10 most frequent sites in experimentally validated *Drosophila* enhancers. Positions in bold highlight the main differences between the logos. The zinc fingers that contact different parts of the 9-mer are shown.

Table 4.1 – Probes used to test binding preferences *in vitro.*

Lane	Oligonucleotides	Comments
1	GCGCCTAGGG**GACCACCCA**CATCGGACGC	Ci #1, used in synthetics
2	GCGCCTAGGG**GACCcCCCA**CATCGGACGC	Ci #5, optimal Lmd motif
3	GCGCCTAGGG**GACCgCCCA**CATCGGACGC	Ci #7
4	GCGCCTAGGG**GACCtCCCA**CATCGGACGC	Ci #4
5	GCGCCTAGGG**GgCCACCtA**CATCGGACGC	Ci #37, used in synthetics
6	GCGCCTAGGG**GACCcCCCg**CATCGGACGC	Ci #146, optimal Sug and Opa motif

Table 4.2 – Summary of *in vitro* binding preferences.

		Ci	Lmd	Sug	Opa	Klu
Ci #1	GACCACCCA	+	+	-	+	+
Ci #5	GACCcCCCA	+	+	+	-	+
Ci #7	GACCgCCCA	+	-	-	+	+
Ci #4	GACCtCCCA	+	+	-	+	+
Ci #37	GgCCACCtA	+	-	-	-	-
Ci #146	GACCcCCgA	+	+	+	+	-

+ This indicates binding was detected.

- This indicates that no binding was detected.

4.7 References

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.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202–W208.

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

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific *phi*C31 integrases. Proc Natl Acad Sci U S A *104*, 3312–3317.

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*, 418–431.

Brown, J.L., Grau, D.J., DeVido, S.K., and Kassis, J.A. (2005). An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. Nucleic Acids Res *33*, 5181–5189.

Campos-Ortega, J.A., and Hartenstein, V. (1985). The Embryonic Development of Drosophila Melanogaster (Springer).

Certel, K., Hudson, A., Carroll, S.B., and Johnson, W.A. (2000). Restricted patterning of vestigial expression in Drosophila wing imaginal discs requires synergistic activation by both Mad and the drifter POU domain transcription factor. Development (Cambridge, England) *127*, 3173–3183.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. Genome ….

Duan, H., Skeath, J.B., and Nguyen, H.T. (2001). Drosophila Lame duck, a novel member of the Gli superfamily, acts as a key regulator of myogenesis by controlling fusion-competent myoblast development. Development (Cambridge, England).

Enuameh, M.S., Asriyan, Y., Richards, A., Christensen, R.G., Hall, V.L., Kazemian, M., Zhu, C., Pham, H., Cheng, Q., Blatti, C., et al. (2013). Global analysis of Drosophila Cys_2-His_2 zinc finger proteins reveals a multitude of novel recognition motifs and binding determinants. Genome Res *23*, 928–940.

Furriols, M., and Bray, S. (2001). A model Notch response element detects Suppressor of Hairless-dependent molecular switch. Curr. Biol. *11*, 60–64.

Ghazi, A., Paul, L., and VijayRaghavan, K. (2003). Prepattern genes and signaling molecules regulate stripe expression to specify Drosophila flight muscle attachment sites. Mechanisms of Development *120*, 519–528.

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 Biol. *14*, R140.

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

Ip, Y.T., Park, R.E., Kosman, D., Bier, E., and Levine, M. (1992). The dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the Drosophila embryo. Genes & Development *6*, 1728–1739.

James, L., and Eisenman, R.N. (2002). Myc and Mad bHLHZ domains possess identical DNA-binding specificities but only partially overlapping functions in vivo. Proc Natl Acad Sci U S A *99*, 10429–10434.

Kim, J., Sebring, A., Esch, J.J., Kraus, M.E., Vorwerk, K., Magee, J., and Carroll, S.B. (1996). Integration of positional signals and regulation of wing formation and identity by *Drosophila vestigial* gene. Nature *382*, 133–138.

Klein, T., and Campos-Ortega, J.A. (1997). klumpfuss, a Drosophila gene encoding a member of the EGR family of transcription factors, is involved in bristle and leg development. Development (Cambridge, England) *124*, 3123–3134.

Kuo, T.C., and Calame, K.L. (2004). B lymphocyte-induced maturation protein (Blimp)-1, IFN regulatory factor (IRF)-1, and IRF-2 can bind to the same regulatory sites. J. Immunol. *173*, 5556–5563.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A.,

McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Lee, H., Stultz, B.G., and Hursh, D.A. (2007). The Zic family member, odd-paired, regulates the Drosophila BMP, decapentaplegic, during adult head development. Development (Cambridge, England) *134*, 1301–1310.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763.

Mann, R.S., and Carroll, S.B. (2002). Molecular mechanisms of selector gene function and evolution. Curr. Opin. Genet. Dev. *12*, 592–600.

Méthot, N., and Basler, K. (1999). Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. Cell *96*, 819–831.

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 (Cambridge, England) *127*, 2999–3007.

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.

Oosterveen, T., Kurdija, S., Ensterö, M., Uhde, C.W., Bergsland, M., Sandberg, M., Sandberg, R., Muhr, J., and Ericson, J. (2013). SoxB1-driven transcriptional network underlies neural-specific interpretation of morphogen signals. Pnas *110*, 7330–7335.

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. Science Signaling *4*, ra38–ra38.

Pavletich, N., and Pabo, C. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. Science *261*, 1701–1707.

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.

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. Philosophical Transactions of the Royal Society B: Biological Sciences *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.

Reece-Hoyes, J.S., Deplancke, B., Barrasa, M.I., Hatzold, J., Smit, R.B., Arda, H.E., Pope, P.A., Gaudet, J., Conradt, B., and Walhout, A.J.M. (2009). The C. elegans Snail homolog CES-1 can activate gene expression in vivo and share targets with bHLH transcription factors. Nucleic Acids Res *37*, 3689–3698.

Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Res *18*, 6097–6100.

Sen, A., Stultz, B.G., Lee, H., and Hursh, D.A. (2010). Odd paired transcriptional activation of decapentaplegic in the Drosophila eye/antennal disc is cell autonomous but indirect. Dev. Biol. *343*, 167–177.

Shimeld, S.M. C2H2 zinc finger genes of the Gli, Zic, KLF, SP, Wilms' tumour, Huckebein, Snail, Ovo, Spalt, Odd, Blimp-1, Fez and related gene families from Branchiostoma floridae. Dev. Genes Evol. *218*, 639–649.

Small, S., Kraut, R., Hoey, T., Warrior, R., and Levine, M. (1991). Transcriptional regulation of a pair-rule stripe in Drosophila. Genes & ….

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

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

UniProt Consortium (2010). The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res *38*, D142–D148.

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.

Wimmer, E.A., Frommer, G., Purnell, B.A., and Jäckle, H. (1996). buttonhead and D-Spl: a novel Drosophila gene pair. Mechanisms of Development *59*, 53–62.

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.

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–D117.

Zinke, I., Schütz, C.S., Katzenberger, J.D., Bauer, M., and Pankratz, M.J. (2002). Nutrient control of gene expression in Drosophila: microarray analysis of starvation and sugar-dependent response. The EMBO Journal *21*, 6162–6173.

CHAPTER 5

Conclusions

5.1 Summary of findings

What are the mechanisms of enhancer regulation by Ci/Gli across the Hh gradient? The objective of this thesis was to answer this question. The findings discussed in Chapter 2 indicate that non-consensus Ci binding sites are essential to respond to Hh signaling in different contexts. In fact, most of the Hh target enhancers in *Drosophila* and mouse are regulated by degenerate Ci/Gli binding sites (Ramos and Barolo, 2013; Winklmayr et al., 2010). Mutational analysis of enhancers containing non-consensus Ci binding sites revealed that these regulatory elements require sites with predicted lower affinity to respond optimally to Hh in the larval wing and in the embryo. Affinity upgrades of these enhancers by swapping non-consensus sites with consensus sites were detrimental to enhancer function in various tissues. We found that in the wing, upgraded enhancers drove expression in more restricted patterns, and in the embryo the overall expression was diminished (Parker et al., 2011; Ramos and Barolo, 2013; White et al., 2012). We proposed that maybe the repressor form of Ci is more cooperative than its activator counterpart, hence the observed restricted expression patterns when the occupancy of these enhancers increases by improving the affinity for Ci (Parker et al., 2011; Ramos and Barolo, 2013; White et al., 2012). In addition, the evolutionary conservation of many of these non-

consensus Ci binding sites suggests that many of these sequences require lower Ci/Gli inputs to respond optimally to Hh signaling.

In Chapter 3, we identified many novel Hh-regulated tissue-specific enhancers in the *ptc* locus. This detailed *cis*-regulatory analysis revealed that the universal response of *ptc* to Hh signaling is achieved through multiple cell type-specific enhancers that mostly contain non-consensus sites. These findings provide more examples of Hh target enhancers that require non-consensus sites for optimal expression in response to Hh signaling. Since we found that most of the *ptc* locus encodes regulatory information, this regions could be labeled as a "super-enhancer" or "stretch enhancer" (Hnisz et al., 2013; Parker et al., 2013). These newly and somewhat controversial types of enhancers are mostly based on chromatin signatures that span large genomic areas. However, our analysis is based on functional reporter data suggests that the complex regulatory information in the *ptc* locus is simply a collection of many individual enhancers. It will be interesting to compare different chromatin signatures across different cell-types in the *ptc* locus to determine whether or not this locus looks like a "super-enhancer", but in reality it just may be many enhancers acting together to control and refine the expression pattern of *ptc* throughout development. Furthermore, our findings of separable tissue-specific enhancers in the fly *ptc* locus are similar to recent findings in the mammalian *patched1* (*ptch1*) locus, where distinct tissue-specific enhancers containing functional non-consensus Gli sites were identified (Lopez-Rios et al., 2014; Vokes et al., 2007). Since these genomic regions seem to use similar regulatory strategies, it is tempting to speculate that the transcriptional regulation of *ptc* is conserved from flies to humans.

The analysis of endogenous enhancers provided very useful in our quest to understand how Ci sites integrate signals from the Hh signaling pathway to regulate gene transcription; however, the additional inputs that regulate these

sequences are mostly unknown. In Chapter 4, to avoid confounding effects from these unidentified inputs, we studied isolated Ci sites in synthetic Hh-regulated enhancers with defined inputs. With this system, we discovered that some Ci binding sites are promiscuous *in vivo*. Initially, we found that in the developing wing the consensus sequence (Ci #1) integrates repressive inputs in addition to Ci, whereas a different site (Ci #37) is immune to these inputs. We found several examples of endogenous Hh-regulated enhancers that are regulated in a similar way. We used synthetic enhancers to simplify our *in vivo* analysis of Ci binding sites. Yet, our minimal approach revealed that these sites are far from being decoded. Even with our current limited data, our findings open up a pandora's box of regulatory complexities for these Ci binding sites. Furthermore, we identified several TFs that can potentially recognize Ci sites *in vivo*. We tested the binding preferences of these candidates *in vitro*, and found that most of these factors bind the consensus Ci site, but not site #37. These results are encouraging, but additional experiments are required to really understand these usurper inputs, and to determine exactly which Ci variants behave as multivalent sites.

5.2 Implications

One of the contributions of my dissertation work is that some Ci binding sites seem to integrate additional transcriptional inputs, besides Ci, in Hh responsive tissues *in vivo*. Based on these findings, I propose that we should re-evaluate previous conclusions about Hh target enhancers containing different Ci binding sites. Previously, as described in Chapter 2, we proposed that the affinity and number of Ci sites explained the expression patterns driven by the wild-type and upgraded *dpp*D enhancer (Parker et al., 2011; Ramos and Barolo, 2013;

White et al., 2012). Now, in light of our recent findings, a simple competition model between activator Ci, repressor Ci, and a broadly expressed transcriptional repressor also explains our observations. For example, wild-type *dpp*D drives expression more broadly because it only integrates inputs from Ci; whereas the upgraded enhancer drives expression only in cells with the highest levels of Hh because it integrates broad repressive inputs in addition to Ci (Figure 5.1). In this model, the sequence of the Ci sites rather than the predicted affinity dictates the final expression pattern.

Our results suggest that some Ci sites may be under complex selective pressures to balance inputs from activator Ci, repressor Ci, and other transcription factors with overlapping binding preferences. This binding site selection might explain why most Hh target enhancers contain non-consensus sites. Furthermore competition for shared or partially overlapping Ci binding sites might provide tissue-specificity to enhancers that are regulated by Hh signaling.

5.3 Future directions

What are the mechanisms of gene regulation by Ci/Gli? How are cell type-specific responses determined? This dissertation aimed to answer these questions, yet our findings suggest that we uncovered just the tip of the iceberg.

First, we identified many novel Hh-regulated enhancers in the ptc locus that respond to Hh signaling via non-consensus sites. These findings in addition to nine, out of ten, previously identified Hh target enhancers in Drosophila suggest a key role for non-consensus Ci sites during development. However, we still lack any in vivo evidence to determine whether or not Ci recognizes these sequences differently *in vivo*. Since there is not a good Ci antibody to perform ChiP experiments, maybe a tagged Ci fusion protein could be overexpressed in

the wing or in embryos, and antibodies that recognize the tag can be used to perform ChIP experiments in order to measure *in vivo* binding of Ci to distinct sites.

We also identified Hh/Ci-independent inputs into some Ci binding sites. In the near future, the unknown factors that bind to these sites could be identified using a yeast-one-hybrid approach or maybe using affinity immunoprecipitation from nuclear lysates. Once these TFs are identified, we can address whether or not they regulate Hh target genes across multiple tissues using genetic approaches. Ideally, in these experiments we will generate cells that are mutant for these factors and we will look at the expression levels of key target genes such as *ptc* or *dpp.* Alternatively, we could use RNAi to knockdown the expression of these factors in different tissues to determine if they affect Hh target gene expression.

Up to this point we have only tested four Ci site variants (#1, #5, #37 and #2949), but we have to measure the expression of many more isolated Ci sites *in vivo* to determine which ones are multivalent. To do this, we can generate more synthetic enhancers to test isolated variants from endogenous enhancers (Figure 5.2). We can also do a sequential analysis where we mutate each position in the 9-mer to the least likely base as shown in Figure 5.3. This experiment may provide clues as to whether parts of the 9-mer are preferentially bound by usurpers. The expression of these synthetic enhancers can be scored across many tissues to further study how Ci regulates tissue-specific responses.

These synthetic enhancer experiments must be complemented with analysis of endogenous Hh-regulated enhancers. For that, we have the largest set of Hh/Ci-regulated enhancers in *Drosophila* (as discussed in Chapter 3). To really understand the rules that control the transcription of target genes, we have to identify additional inputs into sequences other than the predicted Ci binding

sites. More importantly, from a locus-wide perspective, we have to determine whether these *ptc* enhancers are functionally relevant *in vivo*. To address this critical question, we can use the CRISPR/Cas system to delete or mutate these enhancers from the endogenous locus (Cong et al., 2013; Sebo et al., 2014)

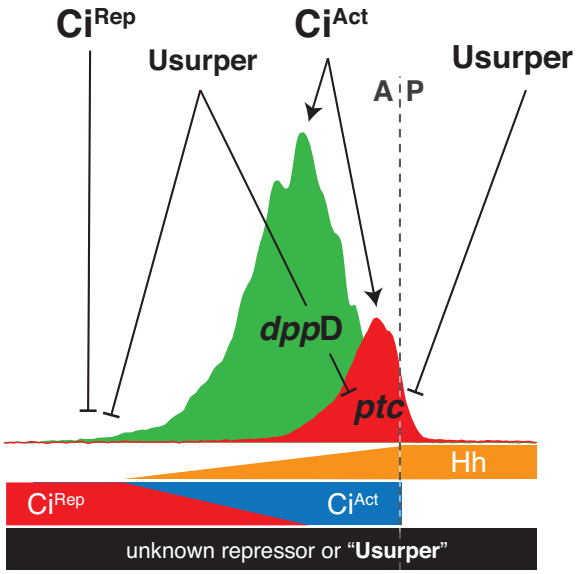


Figure 5.1 – Revised working model that explains the expression patterns driven by the *ptc* and *dppD* enhancers.

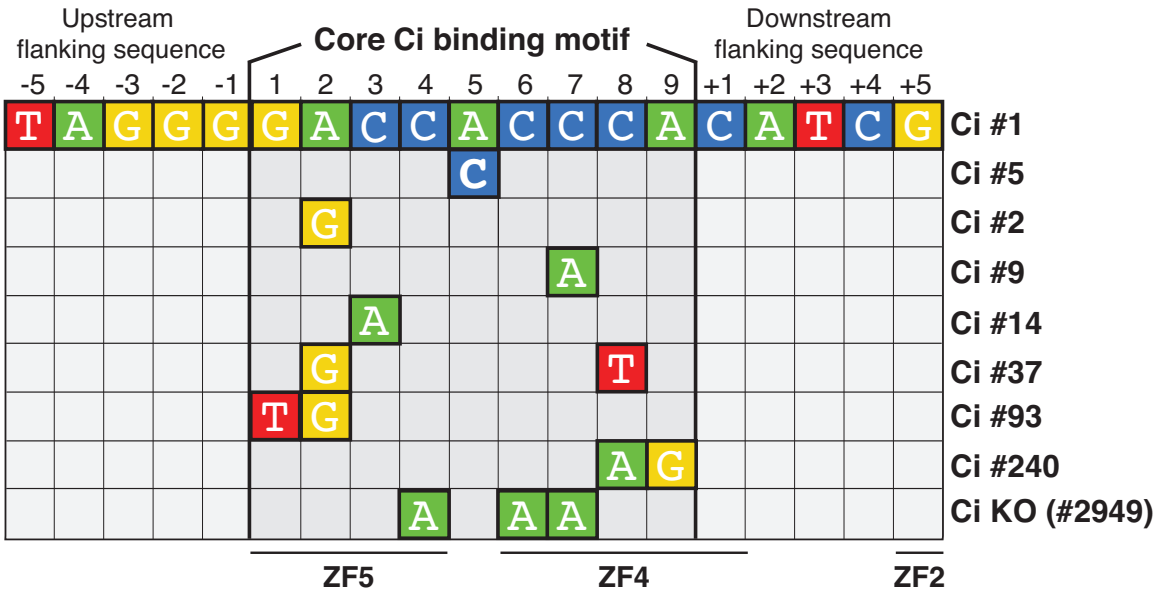


Figure 5.2 – Diagram depicting the Ci site variants that will be tested *in vivo* to understand the role of individual sites in known Hh-regulated enhancers.

Ci Sites #1, #5 and #37 were analyzed in Chapter 4. Ci #2 is present in a *ptc* enhancer (ZY); Ci #9 is present in several ptc enhancers; Ci #14 is present in the mammalian FoxA2 enhancer (Sasaki et al., 1997); Ci #93 is present in *ptc* enhancers; Ci #240 is present in *ptc*, *dpp*HO (Hepker et al., 1999), Nkx6.9 enhancers (Oosterveen et al., 2012). The zinc fingers that make direct DNA contacts are indicated below the cartoon.

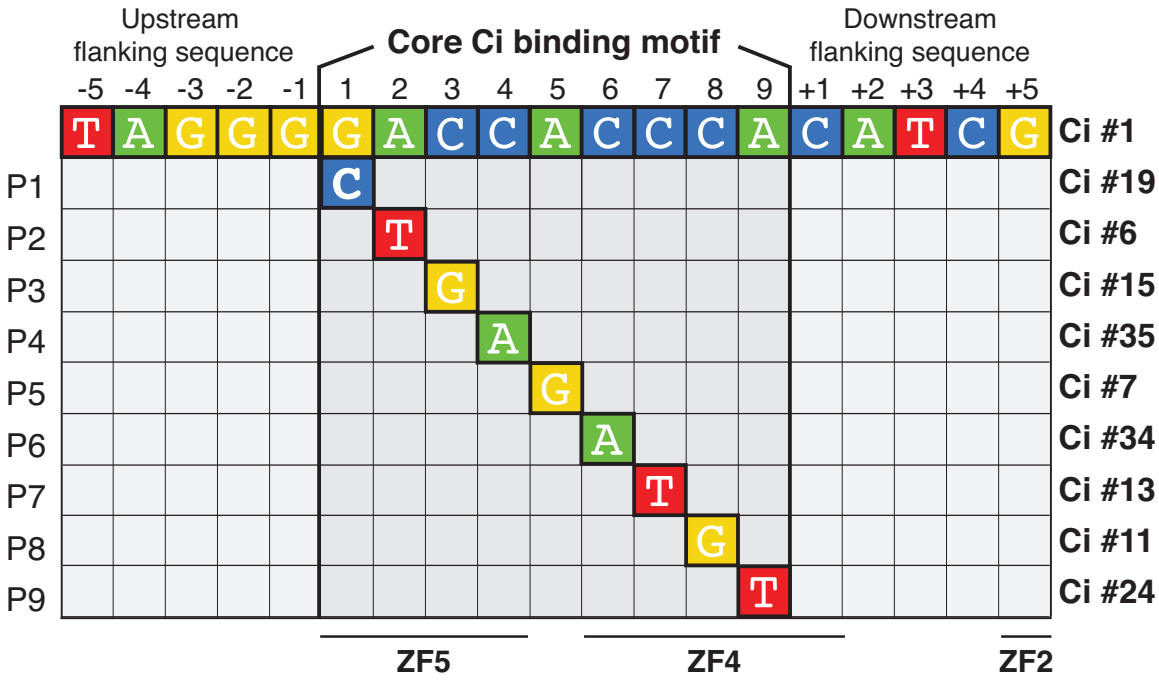


Figure 5.3 – Diagram depicting the Ci site variants that will be tested *in vivo* to study the contribution of different positions in the 9-mer.

The predicted affinity of the sites is indicated on the right, and the zinc fingers that make direct DNA contacts are indicated below the cartoon. Each position will be changed to the least probable base at that position as predicted by the PWM (Hallikas et al., 2006).

5.4 References

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science.

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.

Hepker, J., Blackman, R.K., and Holmgren, R. (1999). Cubitus interruptus is necessary but not sufficient for direct activation of a wing-specific *decapentaplegic* enhancer. Development (Cambridge, England) *126*, 3669–3677.

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.

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

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.

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. Science Signaling *4*, ra38–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., Black, B.L., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Pnas *110*, 17921–17926.

Ramos, A.I., and Barolo, S. (2013). Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. Philosophical Transactions of the Royal Society B: Biological Sciences *368*, 20130018.

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 (Cambridge, England) *124*, 1313–1322.

Sebo, Z.L., Lee, H.B., Peng, Y., and Guo, Y. (2014). A simplified and efficient

germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. Fly *8*, 52–57.

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 (Cambridge, England) *134*, 1977–1989.

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.