Long-range enhancer-promoter interactions: The *Sine oculis* story

by
Amy R. Strom

Thesis advisor: Scott Barolo
Research mentor: Nicole C. Evans
Submitted on April 2, 2012

A senior thesis submitted in partial fulfillment
of the requirements for the degree of
Honors Bachelor of Science
(Cellular and Molecular Biology)
in The University of Michigan
2012

Table of Contents

Chapter 1:

Introduction

Differential gene expression and *cis*-regulation

The driving question behind the field of developmental biology is how a single-celled zygote divides and develops to determine the complex tissue structure that exists in an adult organism. Almost every multicellular organism has many cell types that are extremely diverse, but each contains the same DNA. How does this occur? Part of the answer is *differential gene expression*, meaning that each cell expresses, or produces, a unique combination of proteins from only a subset of the genes that exist in its DNA.

More than 98% of the human genome is noncoding DNA, meaning it does not contain information that produces protein (Elgar 2008). Mice have an even greater proportion of noncoding to coding DNA, while bacteria and yeast occupy the opposite end of the spectrum, with minimal amounts of noncoding DNA (Ning et al 2001, Chinwalla et al 2002). Of note, *Drosophila*, the organism discussed in this paper, lie somewhere in between these two extremes.

These noncoding regions have been found to perform a diverse array of essential functions. For example, repetitive sequences called telomeres exist at the ends of linear chromosomes to preserve DNA stability and prevent degradation, which causes aging and is associated with diseases like cancer (Aubert and Lansdorp 2008). Additionally, transposons, elements thought to be derived from viral sequences, are regions capable of copying themselves and inserting the copied sequence into another site in the genome, called a locus (Rubin and Spradling 1982). Importantly, some of

these noncoding sequences have been shown to be responsible for regulating the expression of proteins produced by coding regions, and thus are critical drivers of differential gene expression.

The noncoding sequences that regulate the initial transcription of coding regions into mRNA are called *cis*-regulatory elements, or enhancers, and are responsible for integrating complex cellular signals in order to direct precise expression of genes. They are able to drive accurate levels of tissue-specific expression at exact times during development and throughout the life of the organism (Banerji 1981, Latchman 1997).

This precise regulation is absolutely required for the proper development and health of the organism. While it is true that many diseases result from mutations in coding regions of the genome that result in nonfunctional protein products, mutations in *cis*-regulatory sequences also cause disease. For example, mutations in the *FOXC2* regulatory region have been implicated in cardiac disorders in humans (Kleinjan and van Heyningen 2005). Additionally, understanding the molecular mechanism of general enhancer function is essential to the design of “custom” enhancers proposed for use in gene therapies and artificial tissue development (Corbo 2008).

Enhancers achieve precise spaciotemporal regulation ability by interacting with DNA-binding proteins called transcription factors. Among other things, these proteins are capable of interacting with 1) the DNA strand and 2) RNA polymerase or other protein cofactors in order to stimulate transcription. Generally, multiple proteins act in combination to stabilize RNA polymerase at the target gene promoter, where it begins transcription (Latchman 1997). Only a specific, and unique, subset of transcription factors are expressed in each cell type, and enhancers contain binding sites for a

particular set of transcription factors. Therefore, a target gene driven by an enhancer is only expressed when the set of transcription factors present in the cell match the ones required to activate expression through the enhancer.

Transcription factors can also act to repress gene transcription. Repressors may prevent activators from binding to an enhancer by competing for an overlapping binding site, by crowding out activating proteins or other by other methods (Xu et al 2000). Transcription factors may also be converted from activator to repressor forms and vice versa, often by proteolytic cleavage (Barolo 2002). Consequently, regulation of gene expression is determined by the combinatorial pattern of activators and repressors that act on an enhancer (Barolo 2002).

Gene expression is influenced by not only by DNA sequence, but also by the three-dimensional structure of its environment, called chromatin. Chromatin exists in at least two states: 1) euchromatin, or open chromatin, that is loosely coiled and easily accessed for protein binding, and 2) heterochromatin, or closed chromatin, that is tightly packed and difficult for proteins to access (Carey 2007). Generally, coding sequences that are actively being transcribed exist in euchromatin and coding sequences not actively being transcribed are packed into heterochromatin.

Enhancers can regulate transcription not only by binding to a specific set of transcription factors, but also by changing the activation state of chromatin. This alteration of chromatin state is enacted through at least two types of chromatin modifiers: ATP-dependent nucleosome remodeling complexes like the SWI/SNF family, and covalent histone modifiers like histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Carey 2007). The studies presented in this paper will focus on

transcription factors that directly interact with the enhancer DNA sequence, which in turn may or may not recruit chromatin-modifying cofactors.

While enhancers are necessary for almost every gene to function, the molecular mechanism of their structure and function remains largely unknown. Important questions that still exist include: What are the basic necessary components of an enhancer? Are there rules that can be defined and then generalized? Are there universal mechanisms of enhancer function?

Models of general enhancer function

At least two theoretical models exist to explain general enhancer function and will be addressed in this paper. One, called Information Display, suggests that the overall 'strength' of the enhancer, the number of activator binding sites versus repressor binding sites, is what determines the level of gene transcription (Figure 1A, Arnosti et al 2005). This model would suggest that the identity of the activators or repressors does not matter as much as the quantity of each. To use a language analogy, consider words to be transcription factor binding sites (Arnosti et al 2005). In this model, the grammar of the enhancer sentence would not matter, as long as there are the correct words. This model predicts that transcription factor binding site position is flexible and that there may be redundancy in function of certain enhancer sequences (Arnosti et al 2005). Redundancy is commonly observed in enhancer elements in the form of multiple binding sites for the same individual or a combination of transcription factor(s).

Experimental evidence supporting the Information Display model has shown that the same gene expression levels can be produced by multiple configurations of binding

sites. For example, one enhancer in *Drosophia*, called the *eve* stripe 2 enhancer, can tolerate a variety of changes in activator placement and identity, and still be able to produce the same levels of gene expression (Arnosti et al 2005). Still, there are limits to how much the spacing or arrangement of transcription factor binding sites can change before gene expression level changes (Arnosti et al 2005).

A second model, called the Enhanceosome model, suggests that successful transcription only arises from very specific combination of transcription factors in a certain order (Figure 1B, Arnosti et al 2005, Carey 1998). In other words, enhancer word identity and grammar are essential for gene expression. Studies of the *runt* enhancer in *Drosophila* show that it is not only the combination and order of transcription factor binding sites, but also the spacing between them contribute to proper enhancer function (Klingler 1996).

The two models presented above are not mutually exclusive, and in most known cases, experimental data points to a combination of the two (Orphanides 2002, Klingler 1996, Carter 2002, Arnosti 2005). The method of enhancer function is a pertinent question in the field of transcriptional regulation. Elucidating the molecular mechanisms of gene transcription could have major implications on developmental, evolutionary and medical studies in the future.

Long-range enhancers and models of their function

Most known enhancers are located within 1 kb upstream of their target gene, but can be found upstream, downstream, or in an intron (break in coding sequence) of a gene. Intriguingly, an enhancer can be located far from the gene it regulates, even

many genes away. For example, the *Sonic Hedgehog* limb enhancer in mice and humans is located one megabase away from the promoter it regulates, in an intron of a different gene (Lettice et al 2003). Little is known about how an enhancer is able to find and interact with the gene it regulates when it is linearly far away in the genome.

Of course, cells exist in a three-dimensional world, and two-dimensional space does not always translate to the third dimension. We know that for proteins, secondary or 'linear' sequence determines much of the three-dimensional structure of the folded protein. Not so with DNA, which is always twisted into a double-helix, no matter the base sequence. The important structural components of DNA arise when it is packaged into more complex arrangements.

All the DNA in a single human cell would stretch out to be about 6 feet, if linear. The cell is able to hold this much DNA because it packages it tightly wrapped around proteins called histones. This packaging requires looping of the DNA first around the histones and then into a coiled 30-nm fiber. These levels of secondary and tertiary structure bring DNA into close physical proximity that would be considered very distant on the linear scale.

A few models of long-range enhancer function exist to explain how enhancers interact with specific promoters from great linear distances. The first model is looping (Figure 2A). The looping model suggests that transcription factors bind to the enhancer and bend the DNA in such a way that they contact RNA polymerase bound at the promoter to stabilize it and promote transcription (Li et al 2006). Recently developed techniques such as chromatin conformation capture have provided evidence for certain enhancers acting through the looping model. Specifically, long-range physical

interactions have been observed in the *TNF* and β-globin loci (Tolhuls et al, 2002; Tsytsykova et al, 2007). Additionally, some distal enhancers can physically interact with their promoters through promoter-proximal "tethering units" that facilitate looping (Calhoun 2002).

The second model is linking (Figure 2B), in which small 'linker' proteins bind at the enhancer and spread out toward the promoter (Li et al 2006). This is the only model in which the transcription factors bound at the enhancer do not directly mediate transcription at the promoter. Rather, the linking proteins facilitate the interaction. Relatively few examples of linking have been carefully studied. One example is the *cut* wing margin enhancer in *Drosophila*, which recruits the small protein, Chip, to link to its promoter (Morcillio et al 1996).

The third model is tracking (Figure 2C). The tracking model suggests that RNA polymerase binds along with transcription factors at the enhancer and that the complex tracks along the DNA until it reaches the promoter (Li et al 2006). Evidence for tracking stems from data suggesting that non-coding RNAs are transcribed between some enhancers and promoters, such as at the human ε- and β-globin loci, and *Drosophila bithorax* complex (Zhu et al. 2007, Gribnau et al, 2000; Bae et al, 2002). Furthermore, chromatin immune precipitation has revealed the presence of RNA polymerase II along DNA between several enhancers and promoters, such as the human β-globin and *prostate specific antigen* loci (Johnson et al 2001, Wang et al 2005).

The fourth model is facilitated tracking (Figure 2D). Facilitated tracking suggests that RNA polymerase binds at the enhancer like in the tracking model, but in this model the transcription factors stay bound to the enhancer sequence as the RNA polymerase

tracks to the promoter, creating a loop structure in the DNA (Li et al 2006). Evidence for facilitated tracking is difficult to separate from evidence for the tracking model discussed in Figure 2C.

Changes in subnuclear localization of enhancer elements in relation to their promoters could also be responsible for enhancer function. Additionally, recruitment of histone-modifiers could alter the local chromatin state and thus activate or repress target gene expression from a distance. Many theories exist to attempt to define the function of enhancers at a distance, but experimental evidence for the molecular mechanism and specific sequences by which these enhancers locate and interact with the correct promoter has yet to be shown.

Drosophila as a model organism for enhancer studies

The studies to be discussed in this paper have been performed in the model system *Drosophila melanogaster.* This small fruit fly is an ideal candidate for studies of transcriptional mechanisms for many reasons. *Drosophila* have complex long-range transcriptional regulation like mammals and other vertebrates. They have a short reproductive cycle (~2 weeks), which makes large numbers of experiments feasible, and are easily sustainable in a laboratory setting. Additionally, almost all *Drosophila* proteins are conserved to humans, making them an ideal system in which to study clinically relevant phenomena in a relatively simple context.

Notably, the field of *Drosophila* research has reliable and relatively inexpensive methods of genetic manipulation. Research in transcriptional regulation often requires the DNA manipulation *in vivo*. Often, manipulation of endogenous genes and

enhancers can have deleterious effects. Hence, reporter constructs are used to create gene expression data without deleterious mutations. Reporter constructs are DNA sequences constructed in the lab that usually consist of an enhancer sequence driving expression of a reporter gene under the control of a heterologous, minimal promoter like that of *Heat shock protein 70* (Hsp70). Most common reporter genes have visual phenotypes, like color production. For example, *Green Fluorescent Protein* and *LacZ* produce green and blue protein products, respectively.

These constructs must be incorporated into the fly genome in order to be expressed *in vivo.* Two main methods are utilized to incorporate foreign DNA constructs into the fly genome: p-element insertion and site-specific integration. P-element insertion integrates the construct randomly into the genome and site-specific integration occurs only at one distinct site. The *in vivo* data shown in this paper is a product of p-element insertion unless otherwise noted.

The *sparkling* enhancer

The *sparkling* (*spa*) mutation in *Drosophila melanogaster* was characterized in 1997 by Fu and Noll at Cold Spring Harbor Laboratories. The mutation caused a 'sparkling'-like phenotype in the eye due to a severe defect in development that caused an eye deformation, due to lack of cone cells and primary pigment cells (Fu & Noll, 1997). Interestingly, when mapped, this mutation did not fit into any known protein-coding sequence.

The mutation was identified as a deletion of a 1.58 kb region of the *dPax-2* gene, spanning part of exons 3 and 4, including the fourth intron. This intron holds an

enhancer for expression of *dPax-2* in cone cells of eye imaginal discs (Fu & Noll 1997). Imaginal discs are groups of cells set aside in the developing larva that will become adult structures after morphogenesis (Sullivan, Ashburner, and Hawley 2000). *dPax-2* is the *Drosophila* homolog of the PAX2 gene in mammals. PAX2 is important for eye development and cell-type specification as well as for patterning the central and peripheral nervous systems (Fu & Noll 1997, Fu et al 1998). The minimal sequence required for cone cell expression was subsequently defined by Flores et al in 2000. This 362 base pair (bp) sequence is referred to as the *sparkling* minimal enhancer, or *spa*.

dPax-2 expression is known to be regulated during eye development in cone cells through the sparkling enhancer by the transcription factor Lozenge, as well as effectors of two well-known signaling pathways (Flores et al 2000). The Notch pathway regulates Supressor of Hairless (Su(H)), and EGFR signaling regulates the Ets factors Pointed P2 and Yan (Figure 3A, Brunner et al 1994, Flores et al 2000). *spa* drives cone cell specific expression of the GFP reporter gene in 3rd instar larval eye discs (Figure 3B) from distances of 121 or 846 base pairs upstream of the promoter (Figure 3C).

Yet, the known inputs of *spa* are not sufficient for enhancer activity, because a construct in which the known binding sites are saved but all other sequence is mutated is not able to drive expression of the reporter gene (Figure 4A, synth). *spa* was chosen as a prime candidate for long-range enhancer study in the Barolo lab because it was well-characterized, but additional inputs must exist, as the known inputs were insufficient to drive reporter gene expression (Johnson et al 2008).

Chapter 2:

Close-range enhancer function of *sparkling*

Experimental Design and Preliminary Data

In order to determine whether the Information Display or Enhanceosome model of enhancer function (Figure 1) best describes *spa* function, activator and repressor inputs to the s*pa* enhancer had to be determined. As stated above, the known regulators of *spa,* Lz, Ets, and Su(H), were not able to drive cone cell-specific expression (Figure 4A). Therefore, the previously uncharacterized sequence in between the known transcription factor binding sites was investigated. The uncharacterized sequences were named regions one through six, with the known binding sites acting as region boundaries (Figure 4A).

In order to characterize the function of each region, a mutational analysis of the enhancer was performed in which each region was mutated or deleted while keeping all others intact. Constructs were tested for ability to drive expression of GFP on a +/- scale where – meant no expression and +++ indicated wild-type expression, with increasing levels of expression represented by corresponding number of + symbols. Each construct was tested at a distance of 846 bp upstream of the transcriptional start site. This was not an unfair test of spa activity because *spa* endogenously must drive transcription from about 7 kb downstream of the promoter (Fu & Noll 1997). These studies were supplemented by testing the same constructs at a promoter-proximal distance of -121 bp. The resulting comparison of the same construct at two distances

can lead to powerful conclusions about long-range transcription that will be discussed in a later section.

Regions found to be necessary for enhancer function at both -846 bp and -121 bp were dubbed 'activators' and regions that repress enhancer function were dubbed 'repressors.' Overall, Regions 1, 4, 5 and 6 each contribute as activators to cone cell expression of GFP. Mutation or deletion of Region 2 or 3 separately do not significantly affect enhancer function. Of note, mutation of Region 5 leads to ectopic expression of GFP in photoreceptor cells, which likely indicates presence of a repressor of photoreceptor expression present in Region 5 (Figure 4C, Swanson 2010).

Additionally, Region 1 had a significant and unique property—it is necessary for enhancer function when the enhancer is distant from the promoter, but is completely dispensable when the enhancer is proximal to the promoter. To date, this is the only enhancer subelement that contributes to long-range capabilities of an enhancer without affecting pattern, and to this end has been named the Remote Control Element (RCE). The RCE and its involvement in long-range enhancer function will be discussed at length in a later section. A pictorial summary of the findings of this region characterization can be found in Figures 4A and 4C.

Identification of Regions 1, 4, 5 and 6 as essential contributors to *spa* function was significant because it showed that not only were the known transcription factor binding sites insufficient to drive expression of the reporter gene, but in fact almost all of the enhancer sequence was necessary for successful gene transcription. In an attempt to identify smaller sequences within each region that were responsible for contributing to enhancer function, each Region was split into three parts (1^a, 1^b, 1^c, etc.) and

mutated individually (1^{muta}, 1^{mutb}, 1^{mutc}, etc.). Surprisingly, every subregion of both Region 1 and Region 4 was necessary for wild-type levels of enhancer function (Swanson et al 2010, summarized in Figure 4B). These results further solidified the essential nature of almost the entire enhancer sequence for successful reporter gene transcription.

Results on the close-range function of *sparkling*

To address the Information Display versus Enhanceosome models, constructs were created with the *order* or *identity* of regions changed, but the ratio of activators to repressors kept the same. In other words, in a set of "region-swap" experiments, the enhancer sentence was made of the same words in a different *order*, and in a set of "identity" experiments, words were replaced with synonyms. These two sets of experiments not only investigated the differences between the two models of enhancer function but also separated enhancer grammar questions from region identity questions.

The ability of these rearranged constructs to drive reporter gene expression would suggest that regions within the enhancer are modular and flexible, which would point to the Information Display model. Region swap constructs unable to drive reporter gene expression would indicate that enhancer grammar is important for proper function. If constructs with region identity altered were unable to drive transcription, region identity would be considered important for enhancer function as well. Meticulous stipulations on enhancer grammar and word identity would point toward the Enhanceosome model.

From the promoter-proximal position, Region 1 is unnecessary for wild type enhancer function (as stated above, Figure 5A, X23456). Also, Region 4 can be duplicated and moved to the 5' end of the enhancer in place of Region 1 (423456) with no detrimental effects on GFP expression (Figure 5A). More interestingly, Regions 4 and 1 are able to be swapped and still maintain enhancer function from close range (423156). However, Region 1 in two copies is not sufficient for general enhancer function (123156), suggesting a unique role for Region 4.

The inability of 423X56 to drive GFP expression indicates the context-specific action of Region 4, because the X23456 enhancer has the same components in a different order and is able to drive transcription (Figure 5A). This lead us to hypothesize that some undefined input is necessary from the region in the middle of the enhancer, because every construct with mutated sequence there was nonfunctional (123X56, 423X56 in Figure 5A and all constructs in Figure 5C, presented below).

Results from an upcoming experiment (X23156) will help determine conclusions about region placement and spacing. If X23156 cannot drive reporter gene expression while 423156 can, the necessary and unique function of Region 4 would be enforced. However, if X23156 is able to drive transcription, some very different conclusions could be drawn. The ability of X23156 to drive transcription would suggest that Region 4 does not, in fact, have a necessary and unique input to the enhancer function. Additionally, it could suggest that two copies of Region 1 (123156) creates a transcriptional repressor, while only one copy of Region 1 placed in the middle of the enhancer acts as an activator of transcription from close-range.

Because both the X23456 and 423456 constructs were able to drive wild type expression of GFP, we investigated the contribution of the 5' copy of Region 4 in the 423456 construct by creating the a, b and c subregion mutations from earlier experiments ($4^{muta}23456$, $4^{mutb}23456$, $4^{mutc}23456$). Remember that mutation of any subregion of Region 4 in context of the wild-type enhancer at a distance resulted in a nonfunctional enhancer (Figure 4B). Mutation of subregions a or c within the 5' copy of Region 4 ($4^{muta}23456$, $4^{mutc}23456$) when the enhancer is close-range did not affect GFP expression levels. This suggests that the 5' copy of Region 4 is not necessary for close-range enhancer function, specifically when there is another copy in the central region.

The $4^{mutb}23456$ enhancer construct has been cloned and injected into flies multiple times, but no transformants have been recovered from the injections. Low efficiency of P-element insertion can be caused by many factors. New amplification, purification and reinjection of this construct will hopefully allow for creation of transgenics.

In a related set of experiments, the same subregions of the 5' copy of Region 4 were mutated, this time in the context of a mutated internal copy of Region 4 ($4^{muta}23X56$, $4^{mutb}23X56$, $4^{mutc}23X56$). Each of these constructs is nonfunctional. This is not surprising, as the 423X56 construct was also unable to drive GFP expression. These results, taken in full, indicate that an intact Region 4 is necessary for close-range enhancer function, and that some undefined element of the enhancer (shared by Regions 1 and 4) must be present in the central region in order to activate reporter gene transcription.

Chapter 3:

Long-range enhancer-promoter interactions

and the potential role of *Sine oculis*

Long-range enhancer experimental design and preliminary data

Region 1 was identified as the Remote Control Element of the *sparkling* enhancer because of its unique function in regulating only the long-range abilities of the enhancer (Swanson et al 2010). Characterizing the capabilities and functions of the RCE has been the graduate focus of my mentor, Nicole C. Evans.

The RCE is flexible in its location and copy number, or multiplicity. The RCE can be moved to the 3' end of the enhancer (234561), or separated from the rest of the enhancer and moved up to 1 kb upstream of the rest of the enhancer while retaining the ability to work with the rest of the enhancer to drive expression of GFP (Evans, unpublished). Multiple copies of the RCE do not greatly affect function, demonstrated by wild type levels of gene expression driven by a construct with one copy of the RCE on either end of the enhancer (1234561, Evans, unpublished).

Three experiments discussed in an earlier section are relevant here as well. Mutation of any subregion of the RCE ($1^{muta}23456$, $1^{mutb}23456$, $1^{mutc}23456$) when the enhancer is at -846 bp shows that each subelement is necessary for wild type levels of expression (Figure 4B). Likely, multiple transcription factors essential for activating transcription bind to the RCE, which make each subregion essential for enhancer activity.

The region-swap constructs used in the above discussion to test which model of general enhancer function model best describes *spa* close-range activity were also tested in the distal enhancer context, to attempt to differentiate general functional necessity from long-range functional necessity. Significantly, the construct with Region 4 in the place of the RCE (423456) was able to drive expression at a distance (Figure 6A). This compensation ability was significant and unique because a construct with Region 4 swapped into the position of Region 5 (123446) was not able to drive expression at a distance (Figure 6A). We hypothesized that there was some similarity in sequence or function between Region 4 and the RCE that caused this ability to drive transcription of the enhancer from a distance.

Another proposed test of the uniqueness of the long-range compensation ability of Region 4 is creation of two more region swap constructs, in which one of the other two known activating regions of the *spa* enhancer were swapped into the place of the RCE. These swap constructs (523456, 623456) have thus far been tested only at -121 bp and were able to drive reporter gene expression (Figure 6A). These results are unsurprising because Region 1 at the 5' end of the enhancer is unnecessary to drive expression of GFP from the -121 bp distance. The implications of the results of the 523456 and 623456 constructs at -846 bp will be discussed below.

Results on the long-range function of *spa*

To identify the sequence within Region 4 responsible for the RCE compensation abilities, we tested constructs with the same three subregion mutations within the 5' copy of Region 4 ($4^{muta}23456$, $4^{mutb}23456$, $4^{mutc}23456$) as the experiments discussed

earlier (Figure 5B), this time at 846 bp upstream of the promoter. Recall that subregions 4^a and 4^c were dispensable from close-range, and that we are working towards gathering data on subregion 4^b at -121. If any of these three constructs were able to drive expression of GFP at a distance, the mutated sequence could be considered dispensable for long-range function. Conversely, any mutated sequence that prevented long-range enhancer function could be considered necessary for long-distance transcriptional activation.

Loss of subregion 4^a ($4^{muta}23456$) in this context did not affect long-range enhancer function, while loss of either subregion 4^b or 4^c individually ($4^{mutb}23456$, $4^{mutc}23456$) prevented reporter gene transcription (Figure 6B). Thus, the sequence from Regions 4^b and 4^c are both required for the long-distance transcriptional activation mechanism.

These data lead us to do a DNA sequence comparison of Region 4 and the RCE to identify common transcription factor binding motifs. Sequence and conservation analysis (data not shown) of the RCE and Region 4 reveal predicted protein binding sites in both for 1) proteins containing a Homeodomain binding motif (H) and 2) *Sine oculis* (*So*) (Figure 7A).

Homeodomain binding proteins are transcription factors that bind a certain short sequence in DNA (ATTA or TAAT). They tend to function in large complexes in combinations that include other, more specific transcription factors (Latchman 1997).

We are especially interested in *So* as a possible regulator of *spa* because it is known to be implemented in eye development (Serikaku and O'Tousa 1994). *So* is a Six family transcription factor required to specify eye primordium, and *So* null mutants

have an eyeless phenotype (Figure 7B, adapted from Cheyette et al 1994). *So* interacts with *Drosophila Pax6* homologs *eyeless* and *twin of eyeless*, and plays a critical role in cone cell and photoreceptor development (Pauli et al 2005). However, it is unclear how *So* could mediate long-range enhancer-promoter interactions based on current literature (Serikaku and O'Tousa, 1994; Pauli et al, 2005). We decided to further investigate the possibility of *So* as a contributing factor to the long-range function of the *spa* enhancer.

To test the ability of So to bind to the RCE and Region 4, we performed a competition Electromobility Shift Assay (EMSA). EMSAs are typically used to determine binding ability of a candidate protein to a radiolabeled DNA probe. The probe is incubated with the candidate protein and run on a nondenaturing polyacrylamide gel. If the protein binds the labeled DNA, it will run at a higher molecular weight on the gel than the free probe, which is not bound by protein. Binding specificity can be tested by competing for protein binding with an unlabeled ("cold") DNA competitor, usually in great excess to the concentration of the original probe. If the competitor successfully competes off protein binding to the labeled probe, it also binds the candidate protein and appears as no shift on the gel (Kothinti et al 2011).

We used a [^{32}P]-radiolabeled DNA probe that included all of Region 4 and 10 bp of flanking sequence on either side to shift *in vitro* transcribed and translated *So* protein (Figure 7C). We competed for So binding with RCE wt and RCE *So* mutant probes. As a positive control for protein-DNA interaction, we used a sequence found in the regulatory sequence of the *Hedgehog* gene that is known to be bound by *So* (HhSo).

The Region 4 probe does shift *So*, and the RCE wt probe is able to compete for So binding, while the RCE So mutant probe is unable to compete (Figure 7C). Therefore, *So* does bind specifically to both Region 4 and the RCE.

To determine which subregion of Region 4 binds *So*, we also competed with Region 4 probes that had the 4^{muta}, 4^{mutb}, and 4^{mutc} subregion mutations. The 4^{mutb} probe was not able to compete for *So* binding, indicating that *So* likely binds to wt region 4^{b}. This binding assay suggests that *So* interacts with Region 4^{b} *in vitro* and, when taken with above results, that Region 4^{b} is essential to long-range function of the 423456 enhancer construct, suggesting that *So* could be involved in facilitating the long-range function of the *spa* enhancer.

Chapter 4:

Future Directions

Specificity-determining compensation experiments

In light of the 423456 enhancer construct's ability to reproduce wild type function at a distance by mediating long-range enhancer-promoter interaction, we have created two additional region-swap constructs. These constructs utilize the other two known activating regions of the *spa* enhancer (5, 6) to attempt to compensate for RCE function (523456, 623456). As presented above, both of these constructs are able to drive expression of GFP from the promoter-proximal 121 bp. The results of these constructs at -846 could inform us about compensation abilities of the other activating regions of the enhancer, and again address the Information Display and Enhanceosome models of enhancer function.

Region 5 contains a homeodomain binding site as well as a repressor of photoreceptor expression (Swanson et al 2010). This region has been shown to be unable to compensate for Region 4 at -846 bp (123556, Figure 6A). Therefore, we do not expect that it would be able to compensate for the RCE in the 523456 construct at -846 bp. The 623456 construct serves as a comparison for the 523456 and 423456 constructs. If neither 523456 nor 623456 show GFP expression, we will consider the ability of Region 4 and the RCE to mediate long-range activity specific to those elements.

Sine oculis in vivo supplemental experiments

Because *So* binds both Region 4 and the RCE in vitro, it is a good candidate for contributing to the long-range function of the *spa* enhancer in vivo. We designed two approaches to confirm *So*'s role in regulating *spa* activity *in vivo*: 1) mutation of the *So* binding sites in the construct enhancer sequence and 2) knock-down of *So* protein in flies using RNAi.

To create enhancer constructs with mutated *So* binding sites, we used the mutations shown by Pauli et al (2005) to abrogate *So* DNA binding. We will test two constructs with mutated *So* binding sites, one targeting the sites from the RCE and Region 4 ($1^{mut\mathit{So}}234^{mut\mathit{So}}56$) and one targeting the site from the 5' copy of Region 4 in the 423456 context ($4^{mut\mathit{So}}23456$). These constructs will be tested at both -121 and -846 distances to determine the effect of loss of *So* binding on long-range and general enhancer functions.

Because *So* is necessary for much of eye development, onset of RNAi knockdown in order to deplete levels of *So* has to occur after larval eye development through cone cell specification. This will be accomplished using a tissue-specific, temperature-induced Gal4/UAS system. This system will express interfering RNA specifically in the eye, after shift into permissive temperature. We have tested several Gal4 drivers and have determined that Glass Multiple Repeat (GMR)-Gal4 met cell-type specificity and expression timing knockdown qualifications the best.

Chapter 5:

Conclusions and Discussion

Conclusions pertaining to general enhancer function

Through the experimental design described in this paper, we were able to separate functions of long-range versus general enhancer mechanisms, Information Display versus Enhanceosome models of enhancer function, even differences between enhancer “word choice” versus “grammar.”

Enhancer grammar, or order of regions, was found to be important in some contexts and dispensable in others. The location of Region 1, the RCE, is particularly flexible, as this region can function to mediate long-range enhancer-promoter interaction from either 3’ or 5’ end of the enhancer. Conversely, a space in the center of the enhancer, which normally contains Region 4, was found to be an important location for some input of enhancer function, because any construct that had mutated sequence in that position was incapable of driving GFP expression (Figure 5).

Swapping Region 4, 5 or 6 into the place of Region 1 (423456, 523456, 623456) in close-range enhancers had minimal effect on transcriptional ability (Figure 6). This was unsurprising given that Region 1 was dispensable at the -121 distance (X23456). Accordingly, the addition of a region known to activate cone-cell expression into the space of the dispensable Region 1 did not have had major effects.

Interestingly, two copies of Region 4 (423456) could drive wild type levels of GFP while two copies of Region 1 in the same enhancer positions (123156) could not activate expression of GFP, even from close-range (Figure 6). This indicates a unique

contribution of Region 4 to the overall enhancer function. Furthermore, each subregion of both Regions 1 and 4 were necessary for enhancer function from a distance, indicating that almost all the sequence contained in the enhancer contributes somehow to overall enhancer function.

Taken together, these data suggest an actual model of enhancer function somewhere along the continuum in between the Information Display and Enhanceosome models. Certain regions are able to undergo significant changes in enhancer position, while others are only able to function from a specific location within the enhancer. Moreover, some regions that are functionally activators are semi-redundant: Region 4 can compensate for 1, but 1 cannot compensate for 4. Still, others that are functionally activators are not redundant: Regions 4 and 5 cannot compensate for each other.

Overall, we have learned that each region and subregion of the *spa* enhancer is mechanistically complex, and that neither the Enhanceosome model nor Information Display model describe a complete model of the enhancer. A complete description of all inputs into a single enhancer is necessary before any definite conclusions can be made about general enhancer function. Additionally, many enhancer mechanisms must be described before "universal rules," if they exist, could be described and extrapolated.

The potential role of *Sine oculis* in long-range transcription

The results presented here implicate a known transcription factor, *Sine oculis*, in a novel regulatory function. The ability to facilitate transcription at a distance in relation to any specific transcription factor had not been extensively studied. Most previous

experimental methods had not specifically separated involvement in general enhancer function versus involvement in long-range activity.

We demonstrated herein that enhancer constructs lacking an *So* binding site at their 5' ends are unable to drive transcription of GFP from -846 bp ($1^{muta}23456$, $4^{mutb}23456$). Moreover, EMSAs show that *So* binds to both Region 4b and the Region 1 *in vitro*. Constructs with targeted mutation of *So* binding sites and knockdown of *So* protein *in vivo* will supplement this data and more definitively implicate *So* in *spa* long-range enhancer-promoter interaction.

Our implication of *Sine oculis* in distal enhancer-promoter interactions could take one of four or more forms (Figure 2): 1) DNA looping; 2) linking of enhancer to promoter through recruitment of linker proteins; 3) transcription machinery tracking; or 4) facilitated tracking (Li et al 2006). The methods described here to identify and characterize the interaction between *spa* and *So* do not distinguish the model of *So-spa* long-range activity, or if *So* is required definitively for this function; rather they simply implicate So in the complex regulatory mechanism of *spa* enhancer control over expression of its target gene.

Implications on future research

The study presented in this paper is valuable for its in-depth analysis of the close- and long-range functions and capabilities of one model enhancer. Even though our understanding of transcriptional mechanisms will remain incomplete until at least one enhancer is described in full, rarely is such a comprehensive study performed. Transcriptional regulation in *cis* is far from being understood, despite obvious

importance in development and disease. Misregulation of gene transcription can lead to severe developmental defects, lifelong disease, even cancer (Kleinjan and van Heyningen 2005). Moving toward understanding of the basic molecular mechanisms behind enhancer-promoter interactions will both elucidate essential developmental pathways, and allow for subsequent studies of transcriptional misregulation in disease.

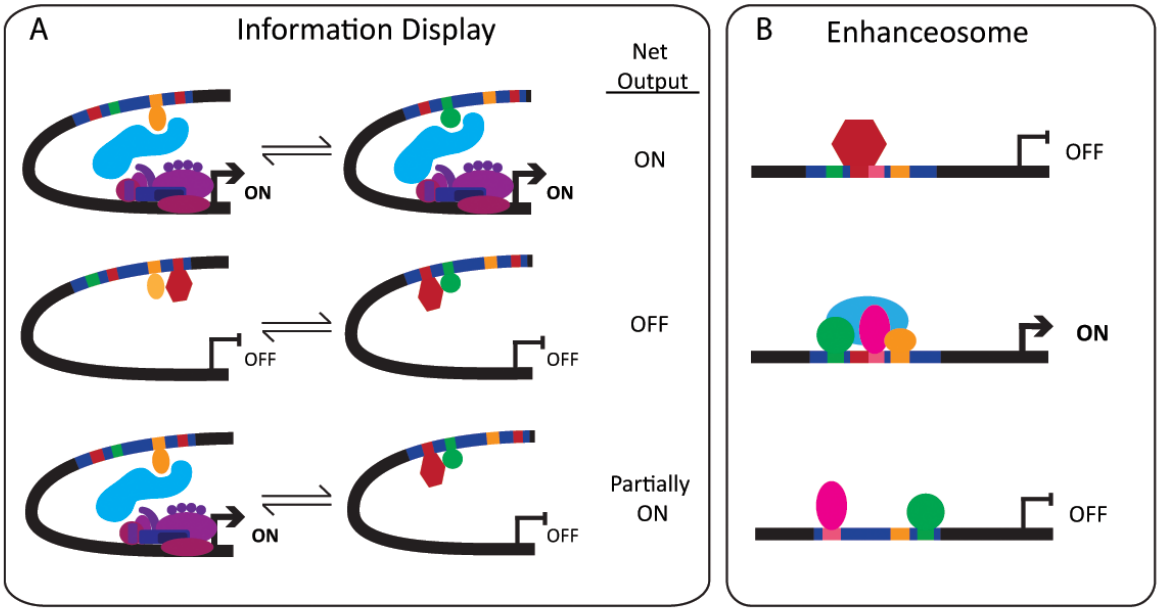
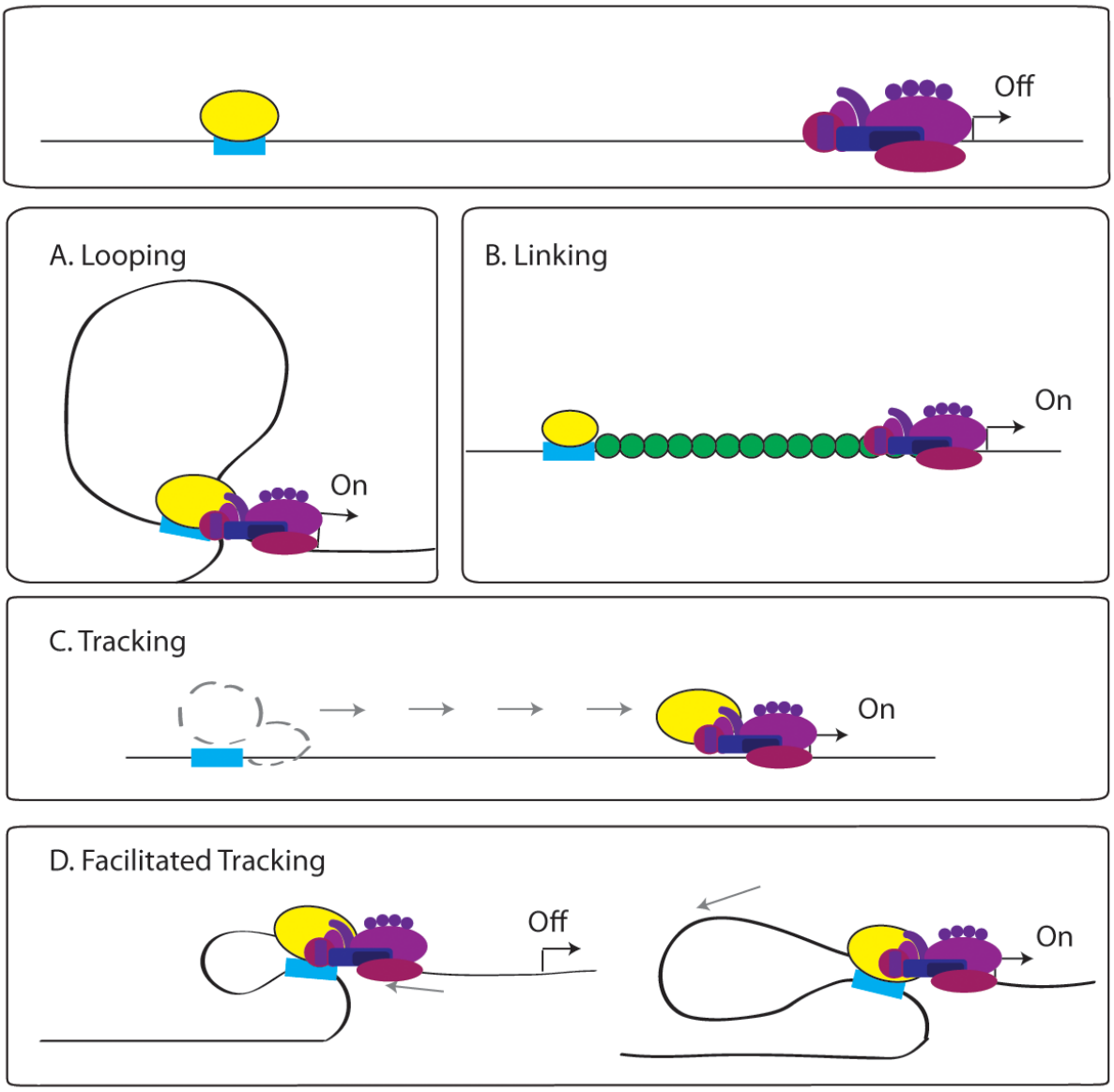


Figure 1: Models of general enhancer function
A. In the Information Display Model, when activator transcription factors (green) are bound, the target gene is transcribed. When repressor transcription factors (red) are bound, target gene transcription is blocked. Transcription factors that are not strong activators or repressors allow transcription as long as there is not a repressor bound. **B.** In the enhanceosome model, repressor transcription factors still block transcription of the target gene. Only a certain combination of transcription factors will allow gene transcription. Even without a repressor blocking transcription, the target gene will not be transcribed if any of the necessary proteins are missing or if the order of binding sites is changed.



adapted from Li et al 2006

Figure 2: Proposed models of long-range enhancer activity
A. A transcription factor bound to the enhancer (light blue) directly contacts the basal transcription machinery bound at the promoter (arrow), creating a loop. **B.** Linking proteins bind at the enhancer, then spread toward the promoter, and facilitate transcription. **C.** A transcription factor and the basal transcription machinery bind together at the enhancer, then track along the DNA strand until reaching the promoter. **D.** A transcription factor and the basal transcription machinery bind together at the enhancer, then the transcription factor stays bound to the DNA and the basal transcription machinery while the basal transcription machinery tracks to the promoter. The enhancers pictured here are all upstream of the depicted promoter, but these models can be used to explain facilitation of long-range enhancer-promoter interactions when the enhancer is downstream of the promoter.

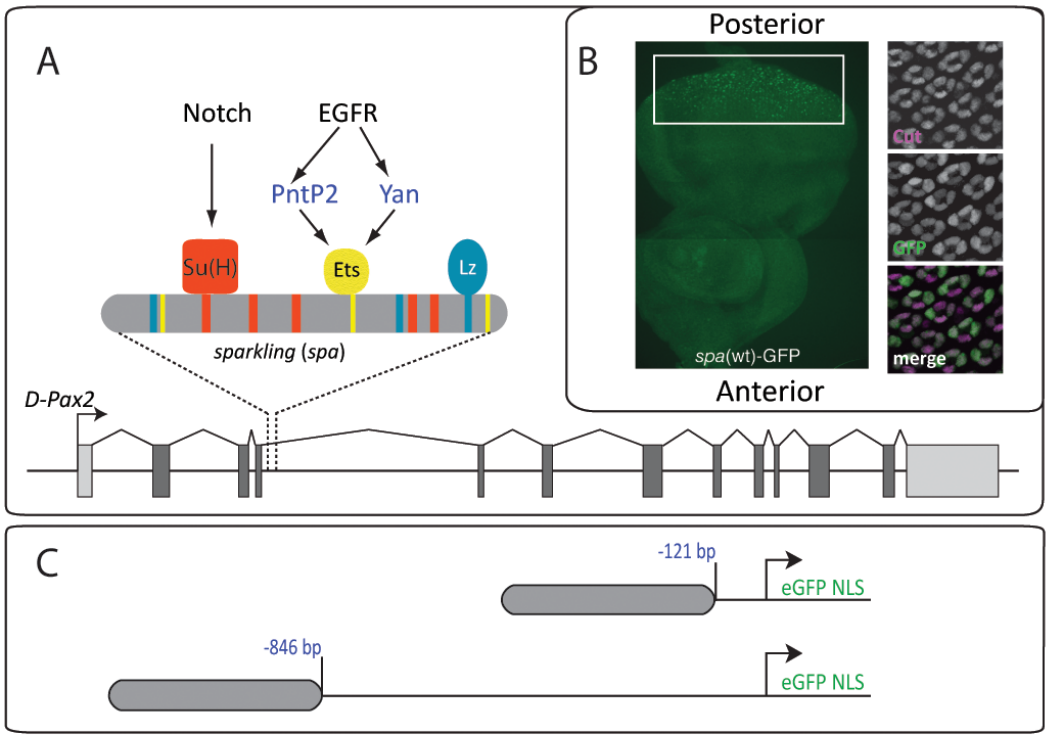


Figure 3: *sparkling* is a cone-cell specific enhancer of the *D-Pax2* gene
A. The *sparkling* enhancer is endogenously found in the fourth intron of the *D-Pax2* gene. It has twelve binding sites for the three transcription factors Suppressor of Hairless (Su(H)), Ets, and Lozenge (Lz). Su(H) integrates signals from the Notch pathway. Ets factors Pointed P2 (PntP2) and Yan integrate signals from the EGFR/MAPK pathway.
B. Left: 3rd instar larval eye imaginal disc expressing GFP driven by *spa*. White box indicates posterior area of the retina, where GFP-expressing cone cells exist. Right: 24-hour pupal eye disc immunostained for GFP and Cut, a marker of cone cells.
C. Top: "Proximal" enhancer construct design. *spa* drives expression of enhanced GFP tagged with a nuclear localization signal from a distance of 121 base pairs upstream. Bottom: "Distal" enhancer construct design. *spa* drives expression of the reporter gene from 846 base pairs upstream.

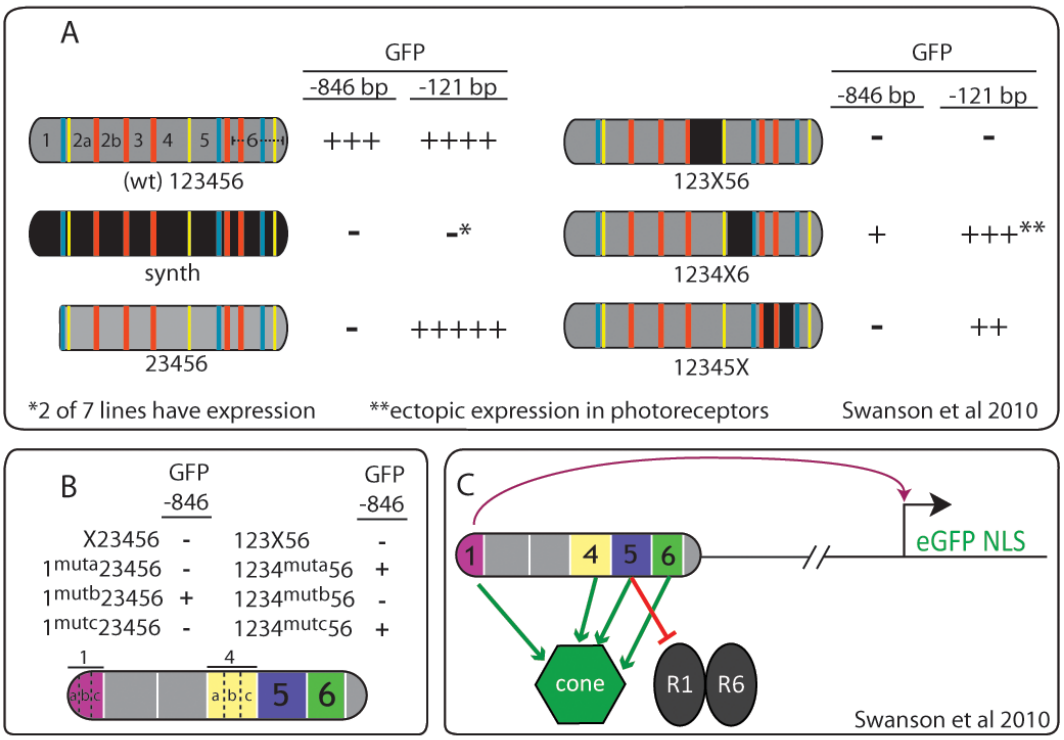


Figure 4: Preliminary Data

A. Wild type (wt) *spa* drives expression of GFP in cone cells from proximal (-121) and distal (-846) locations. The known transcription factor binding sites, in their native spacing (synth), are not sufficient to recreate this expression. Region 1 is dispensable when the enhancer is at a distance from the promoter. Regons 4, 5 and 6 are necessary for wt levels of cone cell specific expression of *spa* at both distances. Deletion of Region 5 at -121 bp can cause ectopic expression of *spa* in photoreceptors. **B.** Summary of subregion mutation constructs of spa driving GFP at promoter distal (-846 bp) position. "X" notation indicates mutation of entire region. Superscript muta, mutb, or mutc indicate mutation of a single subregion. Each subregion of Region 1 and of Region 4 are necessary in their native spacing for general enhancer function at -846 bp. **C.** Model of spa activity. Regions 1, 4, 5 and 6 each contribute individually to cone-cell activation. Region 5 actively represses reporter gene expression in photoreceptors R1 and R6. Region 1 contributes to long-range capabilities of *spa*.

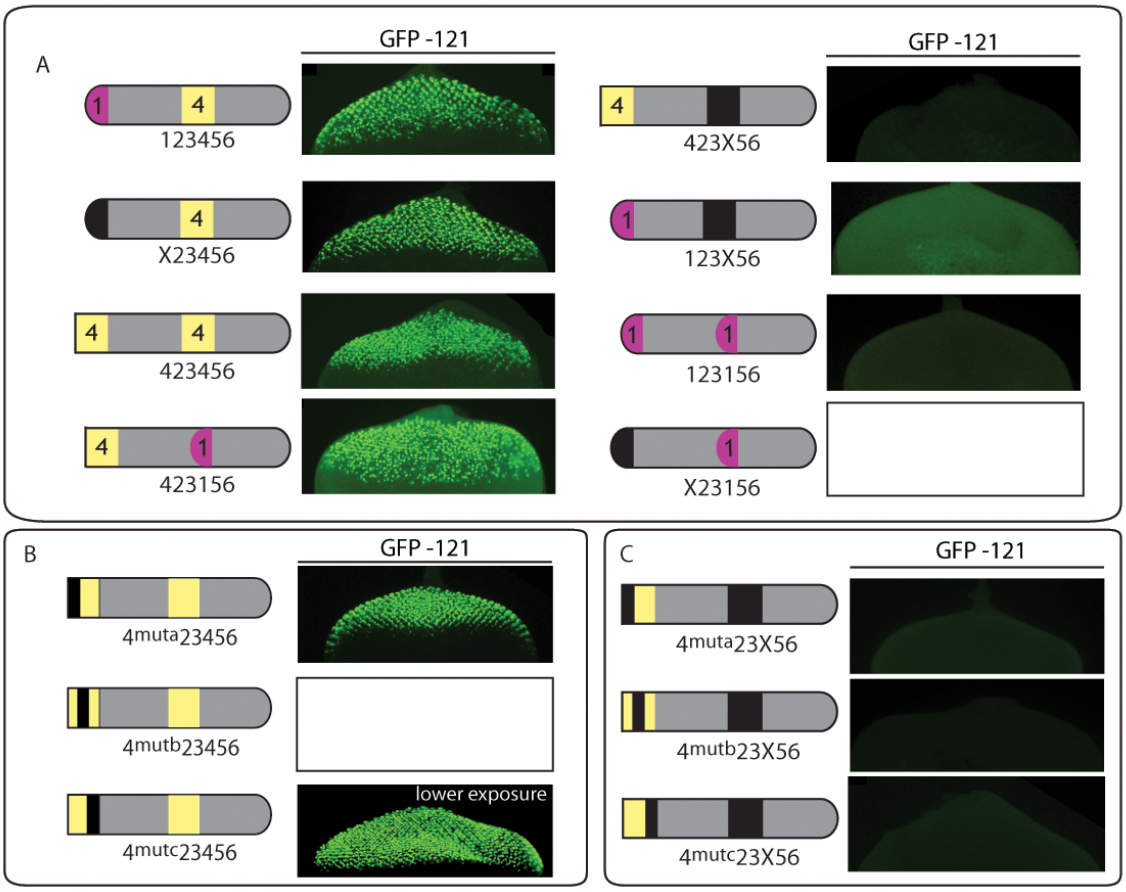


Figure 5: On the close-range function of *spa*

A. From close-range, some constructs with rearrangements of and deletions from the enhancer are able to drive wild-type levels of expression of GFP (X23456, 423456, 423156). Other arrangements and mutations prohibit GFP expression (423X56, 123X56, 123156). White rectangle indicates data is yet to be collected. **B.** Smaller deletions within the 3' copy of Region 4 have no detrimental effects when the enhancer is in the proximal position. $4^{mutc}23456$ lines express GFP at such high levels that data was collected at a lower exposure time to avoid overexposure. White rectangle indicates data is yet to be collected. **C.** Smaller deletions within the 3' copy of Region 4 cannot rescue expression of constructs with mutated sequence in place of Region 4.

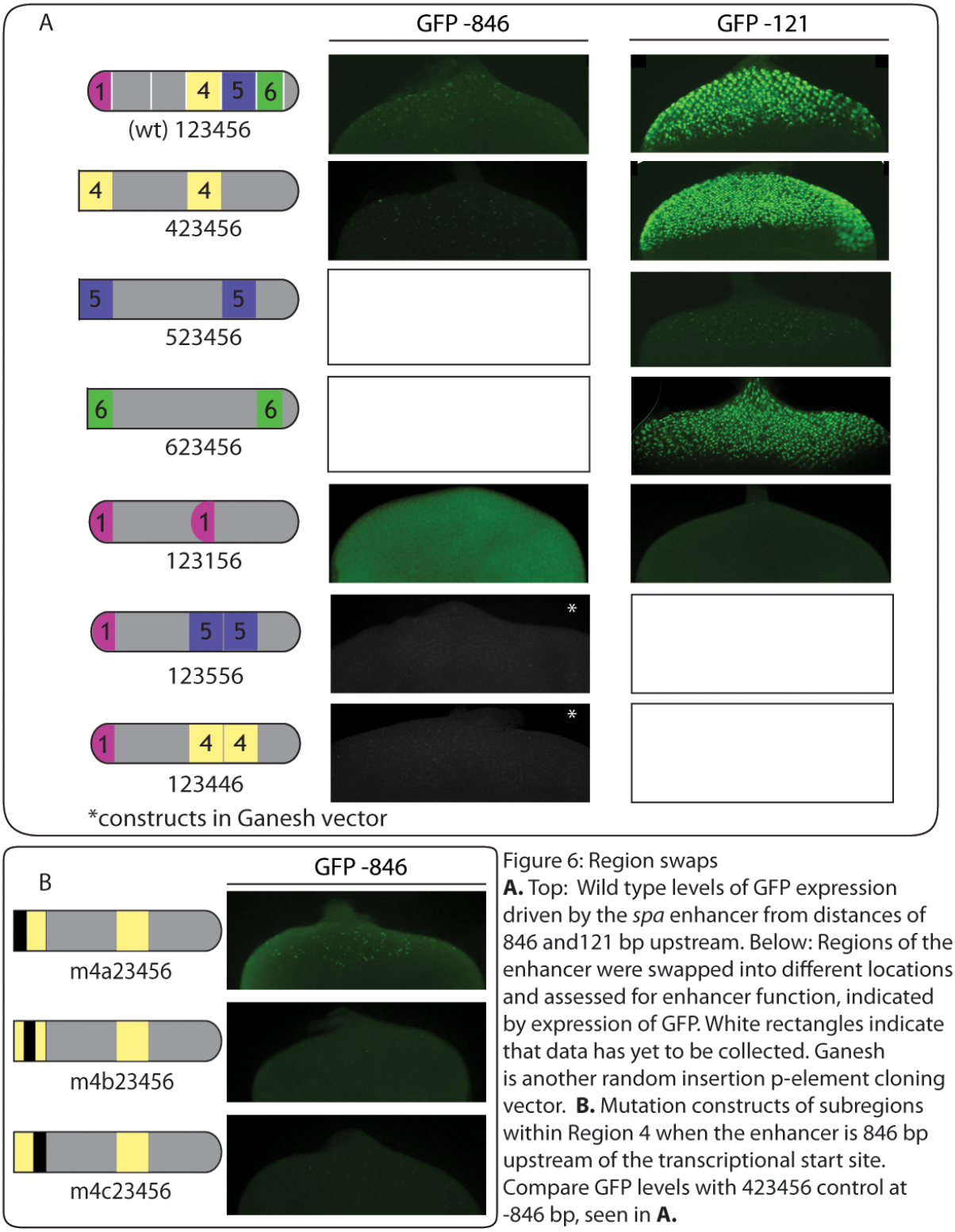


Figure 6: Region swaps
A. Top: Wild type levels of GFP expression driven by the *spa* enhancer from distances of 846 and121 bp upstream. Below: Regions of the enhancer were swapped into different locations and assessed for enhancer function, indicated by expression of GFP. White rectangles indicate that data has yet to be collected. Ganesh is another random insertion p-element cloning vector. **B.** Mutation constructs of subregions within Region 4 when the enhancer is 846 bp upstream of the transcriptional start site. Compare GFP levels with 423456 control at -846 bp, seen in **A.**

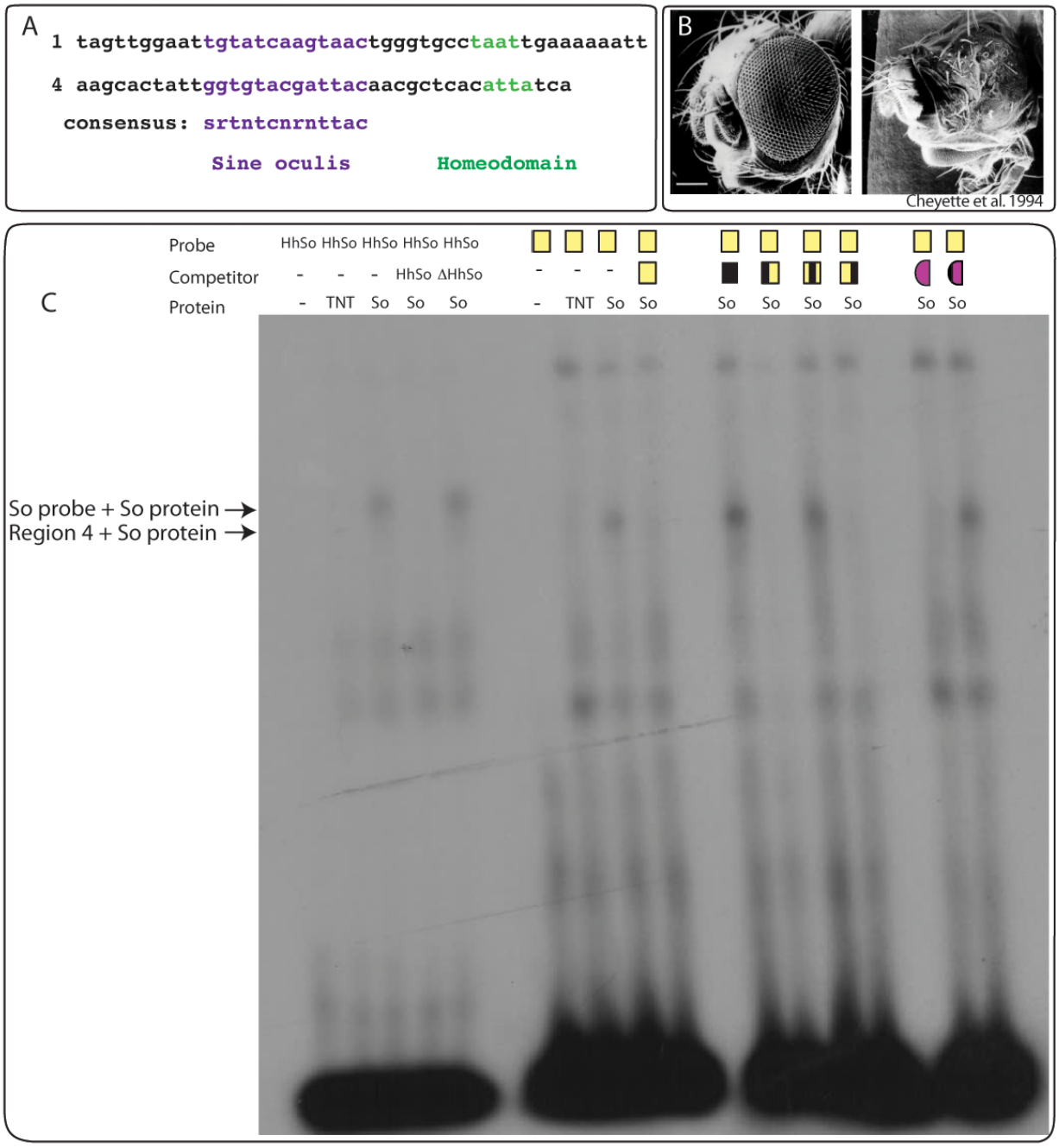


Figure 7: Sine oculis binds Regions 1 and 4

A. Sequence comparison of Regions 1 and 4. Both contain a binding site for Homeodomain protein(green) and the transcripion factor *Sine oculis* (purple). **B.** Left: Scanning electron micrograph of wild-type *Drosophila* eye. Right: Scanning electron micrograph of *Sine oculis* null *Drosophila* eye (Cheyette 1994). **C.** EMSA showing Regions 1 and 4 shifting *in vitro* transcribed and translated *So* protein. TNT is kit used to transcribe and translate *So* protein *in vitro*. HhSo is a DNA probe known to bind *So*, here used as positive control for *So* binding to probe. ΔHhSo indicates the HhSo DNA probe with minimal mutation to abrogate protein binding. Yellow box indicates full Region 4 probe. Yellow box with black stripe indicates Region 4^{muta}, 4^{mutb} or 4^{mutc} probes. Pink semicircle indicates Region 1 (RCE) probe. Pink semicircle with black stripe indicates RCE^{mutSo}.

Methods

Detailed protocols can be found online at http://umich.edu/Barolo.protocols

Generation of constructs

Enhancer constructs with various mutations were created by Polymerase Chain Reaction amplification (Dillon and Rosen 1990). Mutations were induced into enhancer constructs with DNA oligos encoding the mutations. Constructs requiring mutations in the center of the enhancer were created as two fragments and then annealed together in a subsequent reaction known as sewing PCR. PCR mix (1-3 μl template DNA, 3 μl $MgCl_2$, 2.5 μl 10X NEB PCR buffer, 1 μl dNTPs (10 μM), 1 μl forward primer (10 μM), 1μl reverse primer (10 μM), 0.375 μl NEB PCR enzyme mix (10 μM), 15 μl H_2O) was incubated for our standard PCR protocol (95 °C for 2', 29 cycles of 95 °C for 30", 55 °C for 30", 72 °C for 1' 10", then 72 °C for 10').

PCR products and pHstinger cloning vector were digested with EcoRI High Fidelity and BamHI High Fidelity restriction endonucleases from NEB for 1 hour at 37 °C. Fragments were isolated and purified by gel electrophoresis and purified using QIAgen quick gel extraction kit.

Amplification of constructs

PCR product was introduced into the pHstinger cloning vector by ligation: Ligation mix (50 ng vector, 150 ng insert, 2 μl ligation buffer (10X), 1 μl T4 DNA Ligase (10 μM), H_2O to equal 20 μl total reaction) was incubated at 25 °C for 10 minutes. 4 μl ligation product was added to 20 μl NEB10β chemically competent *E. coli* cells and incubated on ice for 30'. Cells were heat shocked at 42 °C for 30" and returned to ice for 2', then allowed to

recover for 1 hour shaking at 37 °C. Circular plasmids are amplified in bacteria by transformation and subsequent growth of bacteria on Lauria Broth (LB) agar plates with selection antibiotic overnight (12-16 hours).

Verification of constructs

DNA was isolated from *E. coli* cultures using QIAgen's Qiaspin mini prep kit and sequenced at the University of Michigan DNA Sequencing Core. Sequence analysis was performed using the computer program Sequencher.

Preparation for Integration into fly genome

E. coli containing constructs of interest were grown in larger quantities and DNA was isolated using Promega's Plasmid midi purification kit. Constructs were combined with integration plasmid Turbo (Δ2-3), which contains DNA sequence for expression of an integrase enzyme, then ethanol precipitated and subsequently resuspended in Injection Buffer.

Creation of Transgenic Animals

White^{-} *Drosophila melanogaster* embryos, which have white eyes as adults, were prepared for injection by dissolving chorion (outer casing) with 100% bleach for 1', and decreasing osmotic pressure by dessication with DrieRite (time depends on ambient humidity). Prepared DNA was injected into posterior of embryos prior to germ cell formation. Embryos were allowed to mature for 24 hours, then transferred to food vials and allowed to develop to adulthood.

Homozygosing *D. mel* lines

Injected flies were crossed to *white*$^{-}$ mates and transgenic progeny (identified by functional *white* gene introduced along with construct plasmid, visually selected by presence of red pigment in eyes) was backcrossed to *white*$^{-}$ to create heterozygous transgenics. Heterozygous transgenics were crossed to create homozygous transgenics. Homozygous transgenics were crossed for three generations to ensure homozygosity.

Collection of Data

Eye discs from third instar larvae were dissected in 1X PBS and fixed in 4% paraformaldehyde for 30 minutes. Fixative was removed by 3 washes in 1X PBS before mounting eye discs onto slides in ProLong Gold Antifade agent with DAPI. GFP expression was observed using Olympus 52X and Confocal microscopes at the University of Michigan Microscopy and Image Analysis Laboratory.

Electromobility Shift Assay

Sine oculis protein was created using an SP6 wheat germ *in vitro* transcription and translation kit (Promega). 3' phosphates were switched with [^{32}P] using T4 Phosphonucleotide Kinase (PNK) (NEB). Probes were incubated with protein and with or without 100X cold competitor for 15 minutes on ice, according to Kothinti et al 2011. DNA-protein interaction was determined by running mixture on a 4-8% acrylamide gel and visualized by exposure to x-ray sensitive film.

References

Arnosti, DN; and Kulkarni, MM. (2005). Transcriptional enhancers: intelligent enhanceosomes or flexible billboards? *J Cell Biochem.* **94,** 890-898.

Aubert, G; Lansdorp, PM. (2008). Telomeres and aging. *Physiol Rev.* **88**, 557-573.

Bae, E; Calhoun, VC; Levine, M; Lewis, EB; Drewell, RA. (2002). Characterization of the intergenic RNA profile at abdominal-A and abdominal-B in the *Drosophila* bithorax complex. *Proc Natl Acad Sci*. **99**, 16847–16852.

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

Barolo, S; and Posakony, JW. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control of developmental cell signaling. *Genes and Development* **16,** 1167-1181.

Brunner, D; Ducker, K; Oellers, N; Hafen, E; Scholz, H; and Klambt, C. (1994). The ETS domain protein Pointed-P2 is a target of MAP kinase in the Sevenless signal transduction pathway. *Nature* **370**, 386-389.

Calhoun, VC; and Levine, M. (2003). Long-range enhancer-promoter interactions in the *Scr-Antp* interval of the *Drosophila* Antennapedia complex. *PNAS* **100(17),** 9878-9883.

Carey, M. (1998). The enhanceosome and transcriptional synergy. *Cell* **92**, 5-8.

Carter, D; Chakalova, L; Osborne, CS; Dai, YF; and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nature Genetics* **32**, 623-626.

Cheyette, BNR; Green, PJ; Martin, K; Garren, H; Hartenstein, V; Zipursky, SL. (1994). The Drosophila *sine oculis* locus encodes a homeodomain-containing protein required for the development of the entire visual system. *Neuron.* **12,** 977-996.

Corbo, JC. (2008). The role of cis-regulatory elements in the design of gene therapy vectors for inherited blindness. *Expert Opin Biol Ther.* **8(5),** 599-608.

Dillon, PJ; and Rosen, CA. (1990). A rapid method for the construction of synthetic genes using the polymerase chain reaction. *BioTechniques* **9**, 298-300.

Elgar G, Vavouri T (July 2008). "Tuning in to the signals: noncoding sequence conservation in vertebrate genomes". *Trends Genet.* **24(7),** 344–52.

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

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

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

Gribnau, J; Diderich, K; Pruzina, S; Calzolari, R; Fraser P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol. Cell*. **5(2),** 377-386.

Johnson, KD; Christensen, HM; Zhao, B; Bresnick, EH. (2001). Distinct mechanisms control RNA polymerase II recruitment to a tissue-specific locus control region and a downstream reporter. *Mol Cell*. **8(2),** 465-71.

Kleinjan, DA; van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet.* **76**(1):8-32.

Klingler, M; Soong, J; Butler, B; Gergen, JP. (1996). Disperse versus compact elements for the regulation of *runt* stripes in *Drosophila*. *Developmental Biology.* **177**, 73-84.

Kothinti, R; Tabatabaj, NM; Petering, DH. (2011). Electromobility shift assay of zinc finger proteins: competition for Zn(2+) bound to Sp1 in protocols including EDTA. *J. Inorg Biochem.* **105(4),** 569-76.

Latchman, DS. (1997). Transcription Factors: An Overview. *J. Biochem. Cell Biol.* **29,** 1305-1312.

Lettice, LA; Heaney, SJH; Purdie, LA; Li, L; de Beer, P; Oostra, BA; Goode, D; Elgar, G; Hill, RE; de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Mol Genetics.* **12(14),** 1725-1735.

Li, B; Carey, M; Workman, JL. (2007). The role of chromatin during transcription. *Cell*. **128(4),** 707-719.

Li, Q; Barkess, G; Qian, H. (2006). Chromatin looping and the probability of transcription. *Trends in Genetics*. **22(4),** 197-202.

Morcillo, P; Rosen, C; Dorsett, D. (1996). Genes regulating the remote wing margin enhancer in the Drosophila cut locus. *Genetics*. **144(3),** 1143-54.

Chinwalla, AT et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature **420**, 520-562.

Ning et al (2001). Initial sequencing and analysis of the human genome. *Nature*, **409:** 860-921.

Orphanides, G; and Reinberg, D. (2002). A unified theory of gene expression. *Cell* **108**, 439-451.

Pauli, T; Seimiya, M; Blanco, J; and Gehring, WJ. (2005). Identification of functional *sine oculis* motifs in the autoregulatory elements of its own gene, in the *eyeless* enhancer and in the signalling gene *hedgehog*. *Development* **132**, 2771-2782.

Rubin, GM; and Spradling, AC. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348-353.

Serikaku, MA; and O'Tousa, JE. (1994). *Sine oculis* is a homeobox gene required for *Drosophila* visual system development. *Genetics* **138**, 1137-1150.

Sullivan, W; Ashburner, M; and Hawley, RS. (2000). *Drosophila* protocols. *Cold Spring Harbor Laboratory Press*; Cold Spring Harbor, New York.

Swanson, CI; Evans, NC; Barolo, S. (2010). Structural Rules and Complex Regulatory Circuitry Constrain Expression of a Notch- and EGFR-Regulated Eye Enhancer. *Dev Cell.* **18**(3):359-370.

Swanson, CI; Schwimmer, DB; Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr Biol.* **21(14),** 1186-1196.

Tolhuls, B; Palstra, R; Splinter, E; Grosveld, F; and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular Cell* **10**, 1453-1465.

Tsytsykova, AV; Rajsbaum, R; Falvo, JV; Ligeiro, F; Neely, SR; and Goldfeld, AE. (2007). Activation-dependent intrachromosomal interactions formed by the *TNF* gene promoter and two distal enhancers. *PNAS* **104(43)**, 16850-16855.

Wang, Q; Carroll, JS; Brown, M. (2005). Spatial and temporal recruit-ment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol Cell.* **19**, 631–642.

Wang, G; Balamotis, MA; Stevens, JL; Yamaguchi, Y; Handa, H; and Berk, AJ. (2005). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. *Molecular Cell* **17**, 683-694.

Zhu, X; Ling, J; Zhang, L; Pi, W; Wu, M; Tuan, D. (2007). *Nucleic Acids Research*. **35(16),** 5532-5554.